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

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

May 2024

**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. Households differ in the ease with which they can respond to events in terms of the a) provision of technology to remotely control devices in their home 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 5%. Having to take an action, one as small as pushing a button on an app versus not having to do so, results in a 5-fold difference in response. We find an additional "manual" decentralized program, one with the same incentives but without remote control technology installed, indistinguishable in their consumption reduction (5%) to the decentralized program with technology. These findings speak to the importance of reducing the effort and cognitive burden on consumers in markets where inattention is prevalent.

*We are grateful to our partner electric utility company for sponsoring and managing the field experiment on which this study is based. We thank Meredith Fowlie, Katrina Jessoe, Robert Metcalfe, and attendees at the 2022 Electricity Camp in the Rockies, 2022 Berkeley Energy Camp, 2022 Canadian Resource and Environmental Economics Annual conference, 2023 ASSA sessions, the University of Alberta, 2023 Berkeley POWER conference, 2023 NBER Spring Environment and Energy Economics Meeting, and 2023 AERE Summer 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.

†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

¶Program and Energy and Sustainable Development and 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 variable renewable energy generation and new sources of flexible demand (such as electric vehicles) will flip that prior adage upside down, with grid operators increasingly forecasting supply and dispatching demand. Yet, despite the growing importance and value of responsive demand, questions remain as to how best to elicit flexibility from typically inattentive and inelastic residential electricity consumers.

While economists have long advocated for dynamic pricing signals to incentivize demand response, this assumes a high level of awareness and understanding on the part of consumers. Consumers must be informed of both their price and usage in real-time, and, even if made aware of both, they must be capable of understanding how adjusting the various devices in their home translates to electricity reductions. For many consumers, this may be unrealistic (Schneider and Sunstein, 2017).

We consider an even more fundamental obstacle to demand response: the reward for adjusting consumption may simply not be worth the opportunity cost of a consumer's time (Becker, 1965). Demand response programs are often infrequent and though beneficial in aggregate to the system, and cumulatively to the consumer, are typically associated with relatively small financial rewards for individual events. Given the opportunity cost of exerting effort and paying attention to price signals, consumers may be reluctant to actively 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 use a field experiment to examine the potential for centrally-initiated demand response to overcome this aforementioned obstacle to unlock flexible electricity demand. Partnering with a large electric utility in Canada, we provide experimental evidence on the relative effectiveness of utility-initiated ("centralized") versus household-initiated ("decentralized") responses to electricity prices. Centralized responses involve an electric utility company or third party altering the consumption of in-home appliances on behalf of customers. Recent evidence has pointed to the potential for automation to facilitate electricity demand response (Harding and Lamarche, 2016; Bollinger and Hartmann, 2020; Blonz et al., 2021). However, such automation still relies on user-generated technology/device settings. Centralized demand response has the potential to further reduce the attention and effort that consumers devote to responding to prices by creating and initiating the settings for them.

We recruited approximately 1,650 participants to a demand response field experiment that ran for 18 months. Participating households were treated with random "peak event" offers, roughly one every 1 to 2 weeks, whereby they could earn money by reducing their electricity consumption during the 3-hour window of each event.[1] Households were assigned to different demand response programs that varied in the technology provided to assist in responding to price signals and whether consumption reductions were initiated by the household (decentralized) or by the utility (centralized). We examine how households respond to the events, and importantly, how that differs across the various programs.

The degree of effort and attention required to respond to events decreases by program. Households in the most basic demand response program, the **Manual program**, received financial rewards for consumption reductions during peak events but were not supplied with any enabling technology to minimize the effort required to do so. They needed to, as the name suggests, manually reduce consumption among their home appliances.

Households in the **Tech program** received the same incentives as the Manual program but were also equipped with app-enabled load controllers on devices our Utility partner installed in their home. The controllable devices include (i) baseboard thermostats, (ii) electric hot water heaters, and (iii) level 2 electric vehicle (EV) chargers. Notably, prior research that considers the role of automation in facilitating demand response has focused primarily on smart thermostats. The Tech program, like the Manual one, is part of our *decentralized* demand response—households in these programs need to actively respond to a peak event notification by taking an action. In the case of the Tech program, however, the effort required is less than the Manual program.

Finally, the **Central program** mirrors the Tech program in receiving incentives and having the same load control technology installed in participating households' homes but with one crucial difference: their load controllers' default response to a

[1] Rewards range from \$1 to \$6 Canadian dollars per event, depending on the level of consumption reduction achieved and whether the event is specified as a regular or high rewards event type. 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 CAD\$1.11 to CAD\$2.22 per kWh as compensation for consumption reductions for the average household in our sample. CAD\$1 equals approximately USD\$0.75 or EUR€0.68 as of December 2023.

peak event notification is to have the Utility automatically reduce their devices' consumption, i.e. *centralized* demand response.[2] Households in the Central program could override the automated consumption reductions during an event but, in contrast to the Tech program, they needed to actively make an effort to *not* reduce consumption by pushing a button on their app. Even if they exert no effort, or pay no attention to the event, households in this program will have their consumption reduced and can earn rewards.[3]

A novel feature of our experimental design is that each household receives their own unique randomized schedule of treatment events. Households not experiencing an event on a given hour act as a control for those experiencing events, and vice versa during other hours. This allows us to estimate average treatment effects of peak events on the sample of participating households within each demand response program through a straightforward difference-in-differences approach. The key identifying assumption with this approach is that experiencing an event one day does not impact or "carryover over" to behavior in like hours on future days (Bojinov et al., 2021). We test the robustness of this assumption by comparing the non-event day consumption of households in our programs to those in a never-treated set of households that met the eligibility requirements and whose household consumption data was tracked but were not placed in a demand response program. We find no evidence that the no-carryover assumption is violated.

Our main results are stark. We find participants in the Central program reduce consumption by an average of 26.3% during events as compared to only 4.8% for those in the Tech program. This difference speaks to the importance of minimizing the effort cost of taking action in settings where both inattention is rife and the incentives per event are relatively small. Having to take an active action to reduce consumption, even one as small as pushing a button on an app to respond to a demand response event, results in one-fifth of the effect as compared to the centralized program. Perhaps more surprisingly, we see no meaningful difference between the Tech program's performance (-4.8%) and that of the Manual program (-5.3%). We find that technology alone is not sufficient to overcome barriers to price responsiveness; making consumption reductions the passive response is imperative.

[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).

[3]In presenting this research, we described this as the "Weekend at Bernie's effect", alluding to the fact that even Bernie, the main character of the popular 1980s movie *Weekend at Bernie's*, who is dead, would see a demand reduction in this scenario, to emphasize the lack of effort required.

One possible trade-off for the superior performance of centralized consumption reductions is that households may be less inclined to accept the centrally-managed program than household-initiated (decentralized) alternatives. Despite our expectations, we find a relatively modest difference in the take-up rates of offers to participate in the Central and Tech programs (42% and 48%, respectively). This surprising result indicates that consumers were not disproportionately deterred by the idea of the electric utility managing the consumption of their devices.

Our results include an analysis of the effect of varying prices on responsiveness. We randomly assign event types to include regular and "high" rewards events, where for the latter the incentives for demand reductions are elevated. We find no significant evidence that higher prices motivate greater consumption reductions. This further suggests a story of responsiveness that relates to overcoming a burden of effort and attention, rather than a smooth response to price signals.

We leverage the richness of our data and our household-specific randomized event schedule to estimate household-specific treatment effects. This allows us to dig deeper into what is driving the varied performance amongst households in different programs. We find the distributions of household treatment effects reveal that average treatment effects of the Tech and Manual programs are driven by a few high-performing households. In contrast, household treatment effects for those in the Central program have a normal distribution with a mean less than zero (specifically, -24%). Therefore, households in the Tech and Manual programs had to be unusual in some sense to significantly reduce electricity consumption during peak events, while the average household in the Central program made significant, large reductions.

We provide suggestive evidence that differences in performance are driven by differences in attention and effort by analyzing data on participant interactions with the utility's App. Across all programs, App interaction is correlated with larger household-level treatment effects. We provide evidence that the "high achievers" in the Tech and Manual programs had to interact with their App very frequently (61% and 51% of days on average during the experiment), which suggests that their performance may have relied on attention devoted to their electricity consumption and effort in reducing it during events. In contrast, the high achievers in the Central program interacted with the app on average significantly less than those in the Tech and Manual programs (31% of days on average). Overall, we find the average household-level treatment effect when households do not interact with the App to be about 3% for the Tech and Manual programs and 24% for the Central program. When households *do* interact with their App, these numbers increase to about 8.5% for the Tech and Manual programs, and 27% for the Central program. This highlights the Central program's "headstart", whereby they achieve consumption reduction even in the absence of App interaction; the decentralized participants had to exert more effort to achieve similar results.

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 program with no load control technology nor automation is broadly inline with those observed in prior CPP studies.

Second, our paper contributes to a growing 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). Consistent with the latter, we find our Tech program performs no better than the Manual program in altering electricity consumption in response to events on average. 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 focusing 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 take an 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. This speaks to the importance of recognizing the

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

cost of effort in settings where inattention is high and individual event rewards are relatively small.

Our analysis proceeds as follows. We begin with a discussion of the barriers to demand responsiveness in Section ??, including a review of the literature and a simple conceptual framework of how the cost of effort and attention can inhibit responsiveness. Next, we describe our experimental design and data in Section 2. We start the analysis of our results with simple descriptive analysis in Section 3. This is followed by our formal estimation framework in Section 4 and estimation results in Section 5. In Section 6, we return to our focus on effort and attention using app interaction data as a proxy to show how treatment effects differ depending on whether households are attentive to the event or not. Section 7 concludes.

## 2 Experimental Design and Data

### 2.1 Overview

We partnered with a large regulated Canadian electric utility (hereafter referred to as the "Utility") to create three "demand response programs" that vary in terms of the enabling technology provided to customers as well as household versus utility-initiated electricity demand changes on specific devices. We are primarily interested in how customers in each program reduce or shift their electricity consumption in response to "peak events", times during which the Utility asks consumers to reduce consumption and rewards them financially for doing so. This is a key parameter of interest for utility companies working to reduce electricity demand during specific hours to meet the needs of a rapidly changing electricity grid.

A novel feature of our study is the fact that we randomized the timing of peak events at the household level. This allows us to estimate the average treatment effect of a household being sent an event notification (with an offer to compensate households for reducing consumption during event hours), by demand response program. We additionally leverage the randomization of these events (as well as the richness of our data) to estimate household-specific treatment effects. These allow us to look at the distribution of effects within each program to better understand what is driving average program-specific results.

### 2.2 Treatment Events

Customers in each of the demand response programs received notifications of peak events through an electricity consumption management phone App offered by the Utility. Because events were randomized in time for each household, they are not correlated with other drivers of household electricity consumption. Peak events had the possibility of occurring at one of two time periods: morning (7am to 10am) or evening (5pm to 8pm). The schedule and timing (morning or evening) of events were unknown ex-ante to the customer, each receiving a unique, randomized schedule of events over the course of the experiment. Consequently, households could not predict the day or event time when they would receive a peak event. Households received event notifications 21 and 2 hours before the event that included an offer for households to receive financial rewards for reducing electricity consumption during the peak event period, relative to their household-specific baseline.[5]

"Event types" were also randomized and were one of two pricing levels: "normal" and "high", with rewards increasing in the latter for large reductions. High peak events were only possible during evening periods. During normal events, households could receive \$1 for a 10% reduction, \$2 for a 30% reduction, or \$3 for a 50% reduction. During high peak events, households could receive \$1 for 10%, \$3 for 30%, or \$6 for 50% reductions. By randomizing the pricing levels, we are able to estimate the effect of greater price incentives on household consumption behavior. The incentive amounts translate to payments ranging from approximately \$1.11 to \$2.22 per kWh of electricity reduced, for the average household.[6] These incentives are in the range of wholesale price caps that are used to limit electricity scarcity pricing in a number of jurisdictions in North America.[7]

Events randomly occurred on weekdays, excluding holidays. Households typically

[5] Baselines were calculated based on a household's average consumption during the relevant event time window over the last five weekdays prior to the event. Customers did not know how the baseline was calculated to avoid the potential for customers to "game" their baseline by over-consuming during event time periods on non-event days. 21 hours was selected as the longest notification period to avoid consumers having knowledge of a pending event while their baseline was still being set.

[6] The average household consumes 1.8 kWh in each hour in our sample. A 10%, 30%, and 50% reduction translates to a 0.54, 1.62, and 2.7 kWh reduction over the three-hour event, respectively. Consequently, for a 50% reduction during a normal peak event, we compensated households $\frac{\$3}{2.7} = \$1.11$ per kWh. For a 50% reduction during a high peak event, compensation was $\frac{\$6}{2.7} = \$2.22$ per kWh. The other percent reductions lie between these two cases.

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

experienced two "normal" and one "high" event per month. This schedule was altered in the summer months of July and August when the likelihood of peak events is lower in Canada. During these months, households experienced no "high peak" events. Events started on February 22, 2022 and continued until June 30, 2023.

Event notifications provided information on the time of the event and the financial incentives for the different demand reduction levels. Once consumers received the 21-hour notifications, they could also 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 were displayed in the App at a two- to three-day lag. The App also gave households a summary of their total rewards to date. See Appendix B.2 for details on each group's in-App experience.

### 2.3 Demand Response Programs

Our experimental sample consists of all households that accepted participation in one of three demand response programs, as well as never-treated households whose consumption we followed but were not offered participation in a program.[8] See Appendix A.1 for a description of the recruitment and assignment process that led to households being in each program.

Table 1. Summary of Household Programs/Groups

| Programs/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 program/group.

[8]We therefore estimate event-level treatment effects for compliers with each program offer. Though selected, the group of households in each program are representative of the groups that utility companies are interested in until such time that utility-managed electricity consumption is something that customers can be defaulted into. See Section 4.2 for discussion of alternative estimators. See Appendix A.3 for a discussion of the comparability of household-level characteristics across program participants.

Table 1 summarizes our demand response programs. Households in the **Manual** program earned financial rewards for demand reductions during events but did not have any load controller equipment installed by the Utility to manage their consumption via the App. They had to respond to events manually.[9] Households could monitor their real-time electricity usage information in the Utility's App.

The **Tech** program differs from the Manual program in that the Utility installed load controller equipment on one or more of the household's electric hot water heaters, baseboard thermostats, and level 2 EV chargers, to enable remote electricity consumption ("load") reductions. This equipment allows households to see device-specific electricity consumption and turn on and off devices remotely via the App. Critically, while the Tech program is equipped with load control technology to ease their effort in responding to events, they still must take active action to do so via their phone's App. Both the Manual and Tech programs allow us to test the efficacy of a *decentralized* approach to demand response—one without and one with enabling smart technology, respectively.

The **Central** program received the same equipment installed in their homes as the Tech program. However, during an event the default setting for Central program participants was for the Utility to manage their load-controlled devices by reducing electricity consumption. That is, without any active response, the Central program participants would reduce consumption via demand management initiated by the Utility; they needed to actively choose *not* to respond to an event (i.e., opt out of utility management) by pushing a button on their App. The Central program allows us to examine the efficacy of centralized demand management and, as compared to the Tech program, the difference between having to opt-in to an event versus being defaulted in and able to opt-out.[10]

Finally, we have two groups of households that serve as *never-treated* baselines throughout the study. One is a **Control** group that receives no intervention or messaging regarding the experiment. Another is an **Info** group that is identical to the Control group, but these households have access to real-time consumption information

[9]We can observe if a household independently installs its own equipment and links it to the Utility's App. Only three (out of 242) Manual households installed equipment independently. These three all installed smart thermostats that allow the monitoring and remote control of the household's electric baseboard heaters.

[10]Customers in the Central program received advanced notifications of events just as other program participants, and the language of the notifications reminded them that the Utility would manage their devices. See Appendix B.1 for event notifications for each program. They had the option to opt-out of utility management for all devices ahead of an event or for individual devices during an event. These options are described in Appendix B.2.

for their home via the App after they accepted and installed a device provided by the Utility partner. In contrast, the Control group can, if they choose, only see their household-level consumption with a one-day lag. Both of these never-treated groups do not receive peak events or financial incentives. We passively monitor their consumption.

### 2.4 Data Description

For all households in our experiment, we track hourly household-level consumption (in kWh) from October 1, 2020 until June 30, 2023. We also have information on a number of household characteristics, such as household appliances, 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 an approximate geographical location. We are also provided time-stamped information on household interactions with the Utility's App at the daily level.

We complement the detailed household-level data with demographic 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.[11] These data were accessed at Environment and Climate Change Canada.

In the last month of the experiment (mid-June 2023), we conducted an additional survey that contained questions on participants' experience with their respective demand response program. We provide a subset of questions from the survey in Appendix D.1. We use this subset in our analysis in Section 6.

### 2.5 Acceptance

Table 2 summarizes the number of households invited and the acceptance rates for each program offer. Acceptance rates among all programs were high. In particular, the acceptance rate for the Central offer was 42%, and only marginally statistically different than the acceptance rate for the Tech offer (48%).[12] Compared to Tech

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

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

and Central, the Manual program had a statistically significantly higher acceptance rate of 59%, followed by the Info-only group at 68%.[13] Finally, Control had 100% acceptance because their participation was not subject to an offer. The acceptance rates of the Central and Tech programs were lower than the others due in part to the need for load controllers to be successfully installed in households that accepted these offers.[14]

Table 2. Program Acceptance by Program/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 program/group. "Accepted" is the number of households that accepted our offer and made it through equipment installation process (as applicable, by program/group). "Pct. Accepted" displays acceptance rates relative to the number of households invited.

We take the similarity among final acceptance rates between the Central and Tech programs as the first set of evidence that we can compare our estimated treatment effects between these programs. While the Manual program had a higher final acceptance rate, concerns that the Manual program participants systematically differ from those of the other two programs are mitigated based on a comparison of observable characteristics across programs in our final sample, as are concerns about differences in the composition of households in each program. See Appendix A.3 for a detailed discussion.

## 3 Descriptive Results

We begin our analysis with descriptive evidence that participating households reduce their electricity consumption during peak events and show how this response differs across demand response programs.

[13]Like the Manual program, the info group required actively accepting the offer to join the experiment and installing a device (called the "Hub") that facilitates the monitoring and reporting of real-time consumption.

[14]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.

Figure 1 provides average hourly household-level consumption for the Central, Tech, and Manual programs for the entire sample period during non-event (solid lines) and event (dashed lines) days. The shaded regions reflect the relevant event hours for each event type.

Figure 1a demonstrates that the Central program had 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, or restart EV charging).[15] Comparing High Evening to Evening event consumption, we see no discernible difference in response to this higher reward.

Figures 1b and 1c demonstrate that the Tech and Manual programs show negligible changes in consumption patterns during events. This limited observable response for the Tech program arises despite the fact that this program has access to the same equipment as the Central households. However, unlike the Central program where changes are automated, the Tech households must actively engage with the App or device to turn off the same appliances during an event.

Taken together, these descriptive results suggest that the Central program has a considerably larger response to each event type. Further, we see no visual evidence of greater performance for greater financial rewards. Rather, the largest difference appears to be whether the household is in a centralized versus decentralized program. In the sections that follow, we undertake a formal empirical analysis to quantify these effects and control for potentially confounding factors.

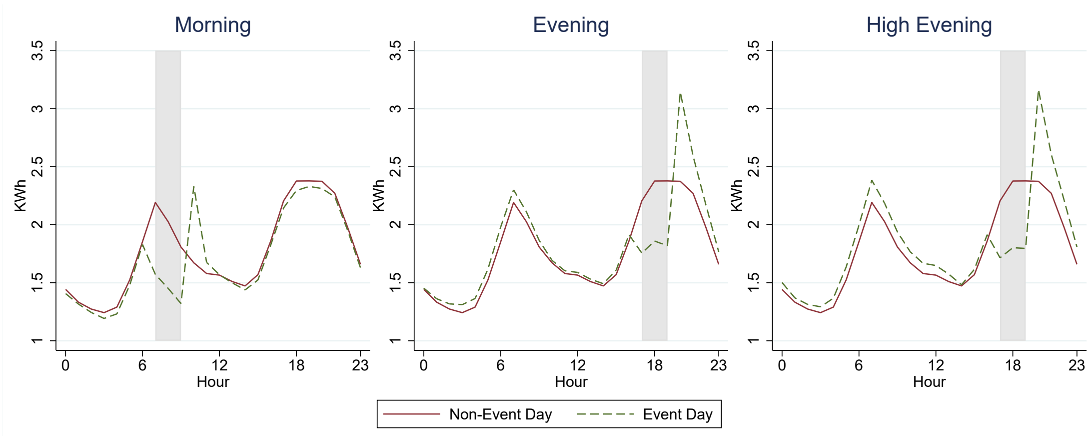
# 4 Empirical Framework

Our estimation strategy relies on the panel nature of our experimental design. The 3-hour treatment events occur randomly within and across all participating households; each household has its own random schedule of events, conditional on them occurring during the set morning (7-10am) or evening (5-8pm) event windows. This, along with the fact that all households in the Central, Tech, and Manual programs were sent events, allow us to estimate the average effect of an event on electricity consumption,

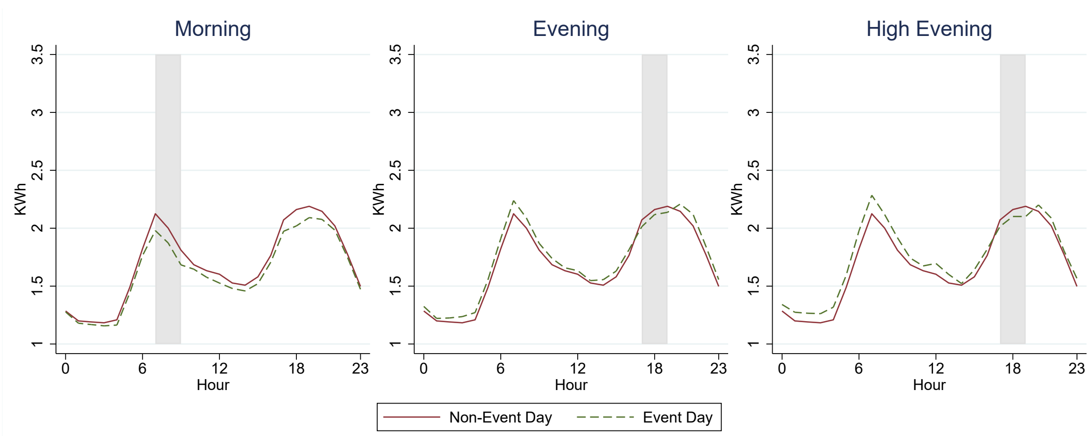
[15] Because we are interested in event-induced demand flexibility among programs, not energy efficiency, we are unconcerned with demand being shifted to another time. However, based on conversations with the Utility, the new problematic peaks in demand created by the observed snap-back could be mitigated by the Utility staggering the beginning and/or end of the load-controlled event across households or only partially adjusting the demand levels on controllable devices. Managing the "snap-back" or "shadow peak" from demand response is an important area for further research.

Figure 1. Average Household Consumption

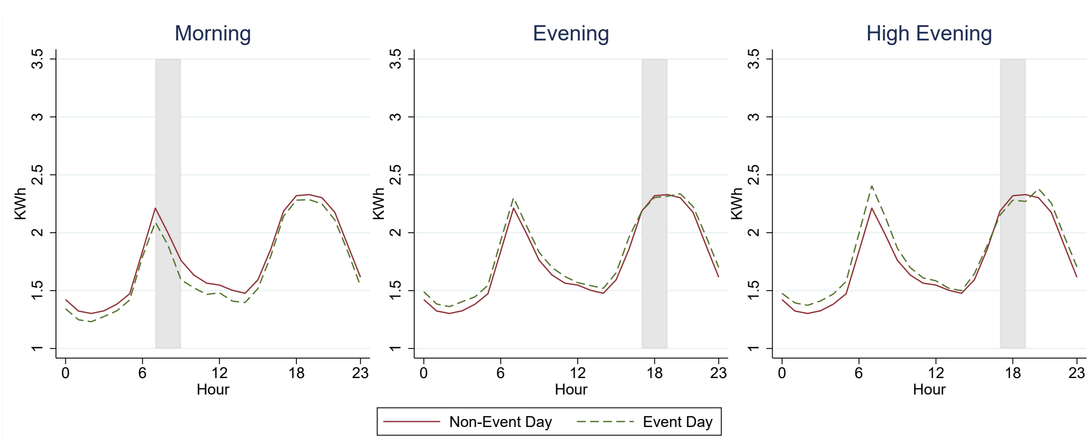
(a) Central Program




(b) Tech Program




(c) Manual Program




Notes: Figure plots mean household consumption for the Central program by hour on weekdays, on event days and non-event days over the period February 1, 2022 - June 30, 2023. Event days are separated by type: Morning, Evening, and High-Evening. The shaded area represents the relevant 3-hour event period.

by program.

### 4.1 Program-Level Regressions

We estimate the average treatment effect of demand response events on electricity consumption by using data from all demand response programs and never-treated households with the following model at the household $i$ and hour $t$ level:

$$ln(c_{it}) \;=\; \sum_{j\in\{C,T,M\}} \beta_j \,\text{Program}_{ji} \cdot E_{it} + \alpha_i + \tau_t + \delta X_{it} + \varepsilon_{it} \qquad (1)$$

in which $ln(c_{it})$ is the log of household electricity consumption ($c_{it}$), $E_{it}$ is the household-specific event indicator that equals one if the household is (randomly) assigned an event in hour $t$, and $\text{Program}_j$ is a categorical variable for which demand response program the household is enrolled in (e.g., Central (C), Tech (T), or Manual (M)). A key advantage of this model is that it allows us to readily test for differences in responsiveness to events across the three demand response programs. We use the log of household electricity consumption on the left-hand side to account for the right-skewed nature of consumption.[16]

We include $\alpha_i$, household fixed effects, which control for time-invariant household characteristics. We also include $\tau_t$, an hour-of-sample fixed effect, which controls for time-varying factors that impact consumption. Household electricity consumption and consumer responses to events may vary with local weather conditions (especially due to thermostat settings). To control for this, 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 these 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. $\varepsilon_{it}$ is the error term. We cluster standard errors at the household level.

We also consider a version of this regression specification where the event indicator variable, $E_{it}$, is adjusted to be a categorical variable for the three potential event-types: Morning, Evening, and High-Evening. This analysis allows us to evaluate if households' responses to events differ by time-of-day and the financial reward for responding, in the case of Evening compared to High Evening events.

For each specification, we report the average marginal effect of an event on house-

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

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

### 4.2 Identification

Our parameters of interest are $\beta_j$ for $j \in \{C, T, M\}$, which measure the change in household-level electricity consumption during peak events for each of the Central, Tech, and Manual demand response programs. To identify each $\beta_j$, the model compares the event-time consumption of households in program $j$ to the non-event time consumption of households in that program, households in the never-treated group, and households in the other demand response programs, conditional on the control variables.

Our empirical framework relies on three identifying assumptions to recover the causal effect of events on household-level consumption. First, events are not correlated with other drivers of household electricity consumption. This is met via our randomization of events. We include weather controls to ensure that our estimated treatment effects can be interpreted as weather-agnostic.

Second, our analysis falls within an emerging literature on experiments called "panel experiments" (Bojinov et al., 2021). Panel experiments involve the treatment of interest (i.e., peak events) that vary in time. Based on this literature, a key identification assumption needed for estimating event-level treatment effects is that our random treatment events do not "carryover" to a persistent change in behavior in similar hours on non-event days. This could occur if, for example, experiencing an event led to a household persistently scheduling consumption reductions during the event hours on all days going forward. Or, conversely, a savvy participant might suspect a financial benefit of "gaming" the baseline by purposely increasing consumption during event hours on non-event days.[17]

We test the validity of the "no carryover" assumption in our context by estimating a DID regression. Separately for each demand response program, we run the following regression that excludes event days in the post-treatment period and includes the

[17] We mitigated this latter effect by providing no information on how the baseline consumption that was used to determine the household's rewards was calculated.

never-treated households:

$$ln(c_{it}) \;=\; \beta\, D_i \cdot \text{EventWindow}_{it} + \alpha_i + \tau_t + \delta X_{it} + \varepsilon_{it} \tag{2}$$

where $\text{EventWindow}_{it}$ equals 1 in the post-treatment period for hours where morning or evening events occur and 0 otherwise. $D_i$ equals 1 if the household is in a demand-response group (i.e., $C$, $T$, or $M$) and 0 otherwise. All other features of the regression analysis are identical to those in Equation (1). This analysis evaluates whether households in each demand response program adjusted their consumption during the event windows on non-event days in the post-treatment period, relative to the never-treated households. If households treated with events did not systematically alter their behavior on non-event days in response to being exposed to events, $\beta$ should be statistically indistinguishable from zero. In addition, we consider a specification that estimates separate effects for the morning and evening event windows on non-event days.

Third, as noted above, for each program, Equation (1) compares event time consumption to non-event time consumption of households in the same demand response program, other demand response programs, and never-treated households. This relies on the assumption that the households in the never-treated and other demand response programs provide a valid counterfactual on non-event days. While observed characteristics are similar across the demand response programs (see Appendix A.3), one may be concerned that the programs are differentially selected, and this impacts our ability to compare event time to non-event time behavior across programs, even after including our various control variables.

We undertake two additional regression specifications that vary the comparison groups used to identify the response to peak events. We adjust the specification in Equation (1) to only consider each demand response program separately with the never-treated households. This specification identifies our parameters of interest by comparing event time behavior to non-event time behavior in the same demand response program and never-treated households. We also consider a specification that only includes data from each demand response program. This specification identifies the response to events by comparing behaviour to non-event time consumption from households within the same demand response program. This is advantageous if there are concerns about differences across demand response programs that preclude the comparison group from providing a valid counterfactual. To the extent that our results are robust to these alternative specifications, this will alleviate concerns that potential differential selection into demand response programs impacts our key conclusions.

Our empirical framework provides estimated average treatment effects by program for households that accepted the demand response program offers. From a policy perspective, our estimates may be the parameters of most interest to utility companies that want to understand what time-specific flexibility will be possible/expected among participants who accept an opt-in demand response program.

Nevertheless, if one is interested in the impact of each *program* on time-specific electricity consumption, one needs to address selection at the program-level. We provide program-specific local average treatment effect (LATE) estimates, leveraging the fact that we also randomized program offers to households and utilizing a "synthetic event" approach for households that did not take up the offer.[18] We also use synthetic events to retrieve another standard experimental parameter that may be of interest: intention-to-treat (ITT) estimates, which are the impact of offering each program on event-level demand reductions. Detailed methods and results are available in Appendix C.

### 4.3 Household-Level Regressions

A unique feature of our setting is our ability to estimate household-level treatment effects. We can leverage the fact that event timing is randomized at the household level, ensuring that events are not correlated with factors that drive consumption decisions. This allows us to do two things: (1) examine heterogeneity in event responsiveness across households and (2) look at factors associated with household-level responsiveness that speak to the role of attention and effort in responding to events or otherwise drive differences in treatment effects across programs.

For each hour $t$ and household $i$ in the demand response programs, we estimate the following model:

$$ln(c_{it}) \;=\; \gamma_i + \beta_i \, E_{it} + T_t + \delta_i X_{it} + \eta_{it} \tag{3}$$

where, analogous to above, $c_{it}$ is household consumption, $E_{it}$ equals 1 when household $i$ has an event and zero otherwise, and $X_t$ includes the same set of temperature controls as the specification in Equation (1). In this specification, $T_t$ is a set of time fixed effects that includes day-of-week, hour-of-day, and year-month to capture time-

[18] More specifically, we created "synthetic event" schedules for households that did not accept our program offer. These schedules were randomized in the same manner as those for households that ultimately participated in our experiment.

varying factors that impact consumption.[19] $\eta_{it}$ is the heteroskedastic-robust error term.

The regression in (3) gives us an estimate, $\widehat{\beta}_i$, for each household in our demand response programs. The identification strategy of this household-level regression compares consumption behavior during event hours to non-event hours within the same household, conditional on time-based fixed effects and weather variables. Analogous to the discussion in Section 4.2, the identification of the treatment effect assumes there is no carry-over effect to consumption during event hours on non-event days.[20] We summarize the distribution of estimated treatment effects for each demand response program using a non-parametric kernel density function.

The ability to estimate household-level treatment effects provides us with the opportunity to understand the potential mechanisms driving our results. In particular, we have data on when a household has interacted with the App on a given day. App interactions are indicative of the attention and effort that households expended to respond to events.[21] We leverage this to estimate separate event treatment effects, by household, for when households do and do not interact with the App.

We run a specification of (3) that interacts $E_{it}$ with an indicator variable App Interact$_{it}$ that equals one if the household has interacted with the Utility's App on the relevant day and zero otherwise. A key benefit of this approach is that it allows us to quantify how a specific household's estimated treatment effect varies by whether or not they interacted with the App on a given day. This helps overcome the sample selection challenge that would arise by running an analogous regression using all households within a given demand response program. With such a regression, it would not be possible to disentangle whether the different treatment effects arise because the household interacted with the App on an event day or whether the households that interact with the App are unique in the way they respond to events.

[19]We cannot include an hour-of-sample fixed effect in the household-level regression because it would absorb the variation we are using for identification in this specification. We include several calendar fixed effects to absorb seasonal, day-of-week, and hour-of-day factors that impact consumption. We estimated our program-level specification detailed in Equation (1) using this set of fixed effects. The results closely reflect the estimates reported below.

[20]See Section 4.2 for details on how we evaluate if the no carry-over assumption is satisfied.

[21]Recall that households can use the App to monitor their hourly household consumption and observe the timing and rewards for upcoming events 21 hours in advance. Households with installed devices can also monitor their device-level consumption in the App. Tech households can adjust their connected devices by pushing a button in the App (e.g., to turn off their use during events). Finally, Central households can adjust their connected devices and opt out of centralized load management during an event.

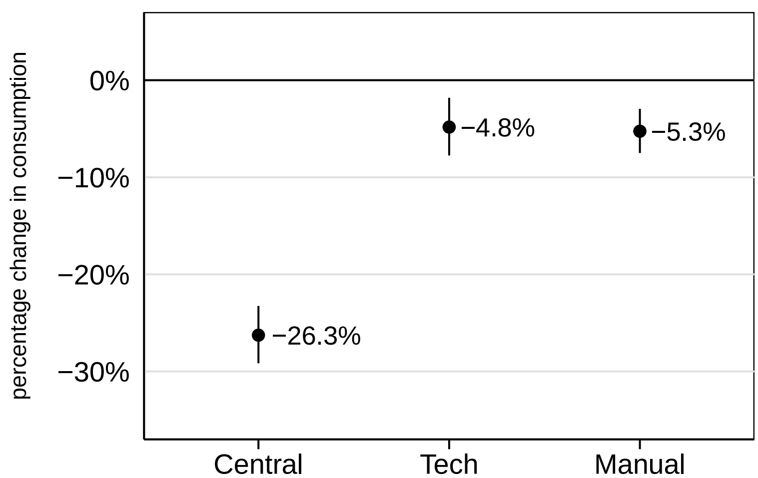
## 5 Empirical Results

This section presents the results of our demand response group-level econometric analyses. In particular, we provide the average treatment effect of events for each program, across all events and then separated by event type.

### 5.1 Program-Level Treatment Effects

Figure 2 provides the estimated average response to events by program as a percentage change in household-level consumption using the specification in equation (1). We observe an average 26% reduction in consumption during events for the Central program. In contrast, the Tech and Manual programs reduced demand by approximately 5% on average during events. Both of these effects are statistically significantly different from zero. Even though the Tech program had the same equipment as the Central program, it demonstrated a significantly lower response to events. Additionally, the average response for the Tech and Manual program are not statistically significantly different from each other.

Figure 2. Average Estimated Treatment Effects of Participants by Program




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

Recall that the Central program has the ability to opt-out of events using the App. We observe an opt-out rate of only 4% at the event-connected device level from the participants in the Central program. When taken together with the results for the Tech and Manual programs, this low opt-out rate suggests that the large reductions for the Central program are primarily attributable to consumers allowing utility management of their devices during events.[22]

These results are consistent with the descriptive data presented in Section 3 that suggests that households in the Tech program were not using the load controller equipment to the same extent as the Central program. They also suggest that smart technology that enables device remote control provides minimal resolution of consumer barriers to responding to short-run electricity prices and peak event notifications, while centrally managed, automated demand response resolves barriers and yields large demand reductions.

Recall from Section 4.2 that we undertake additional analyses to evaluate the validity of our identification strategy. Appendix Table C1 provides the results of our regression analysis when we vary the comparison group during non-event hours to estimate the event treatment effects. These results demonstrate that our estimated treatment effects are highly robust to varying the comparison groups included in the regressions. This suggests that non-event time consumption does not significantly differ between the demand response programs and the never-treated group, and that using the demand response programs' non-event time consumption provides a valid control for event-time consumption to estimate treatment effects. Appendix Table C3 provides the results for our test of the validity of our no carryover assumption. In this analysis, we find no evidence of changes in behaviour during the event window on non-event days for any of our demand response programs.

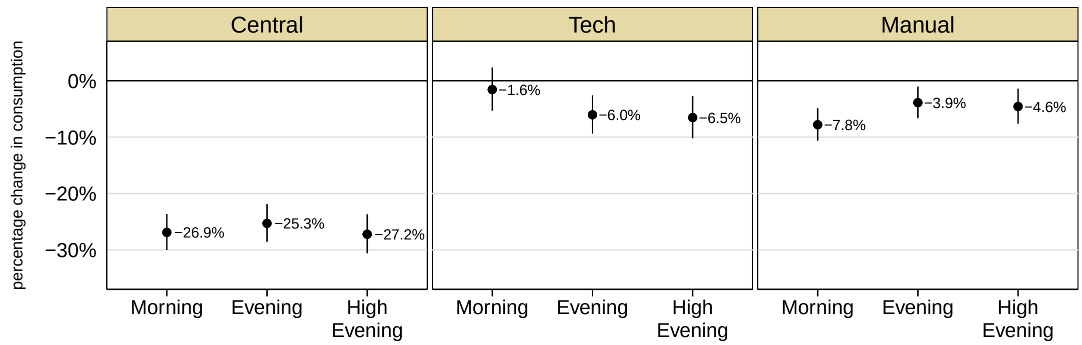
### 5.2 Program-Level Treatment Effects by Event Type

In addition to randomized event timing, we also randomize event types, varying both the time of the event and the financial reward for reductions. This allows us to estimate how consumers respond to different price incentives and event times.

Figure 3 presents the estimated response to events allowing for differential responses by event type. For the Central program, we see a large demand reduction for all event types, with an approximate 27% reduction during morning events, 25% during evening events, and a 27% average reduction during high evening events. This

[22] Notably, when consumers did opt-out of central management, they generally did so for their thermostats: 90% of event-device opt-outs occurred by households adjusting their thermostats during events.

Figure 3. Average Treatment Effect of Participants by Program and Event Type




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

indicates that the Central households allowed central management of demand during both morning and evening times. It also indicates that they were not distinctly more responsive to the greater incentives offered during the High Evening events.

During the Evening and High Evening events, the Tech program reduced its demand by approximately 6%, while the Manual program had a 4% estimated reduction in demand during these event types. These effects are statistically different from zero. The Evening and High Evening Tech and Manual program effects are not significantly different from each other, when compared within each event type.

The Tech program had a response to Morning events that are not statistically different from zero. This differs (statistically significantly) from the Manual program, which had an average estimated reduction of 8% during the Morning events. This is a counter-intuitive result, as the Tech program had all the same information, incentives, and abilities as the Manual program 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.

For all three programs, the change in consumption during High Evening events does not statistically significantly differ from their responses to regular Evening events. This suggests that the increased financial incentives for reduced electricity consumption during these times does not motivate participants to undertake additional effort to make greater reductions in usage. This result suggests the barrier to demand responsiveness may have less to do with the scale of financial rewards and more to do with the hurdle of effort and attention.[23]

Similar to the discussion in the previous section, Appendix Table C2 provides the results of our regression analysis when we vary the comparison group used to estimate event treatment effects. These results continue to demonstrate that our estimated treatment effects are largely robust to varying the comparison groups included in the regressions. The estimated response to morning events for the Tech and Manual programs varies with the relevant comparison group, with a smaller response for the Manual and a larger response for the Tech than our main specification. However, our key conclusions persist. The Central program is considerably more responsive to all event types. Further, no program shows a distinct response to the elevated incentives during the high evening events.

Finally, Appendix Table C4 provides the results for our test of the validity of our no carryover assumption, allowing for differential estimates for the morning and evening event windows. In this analysis, we find no evidence of changes in behaviour during either the morning or evening event windows on non-event days for any of our demand response programs.

## 6 Attention and Effort

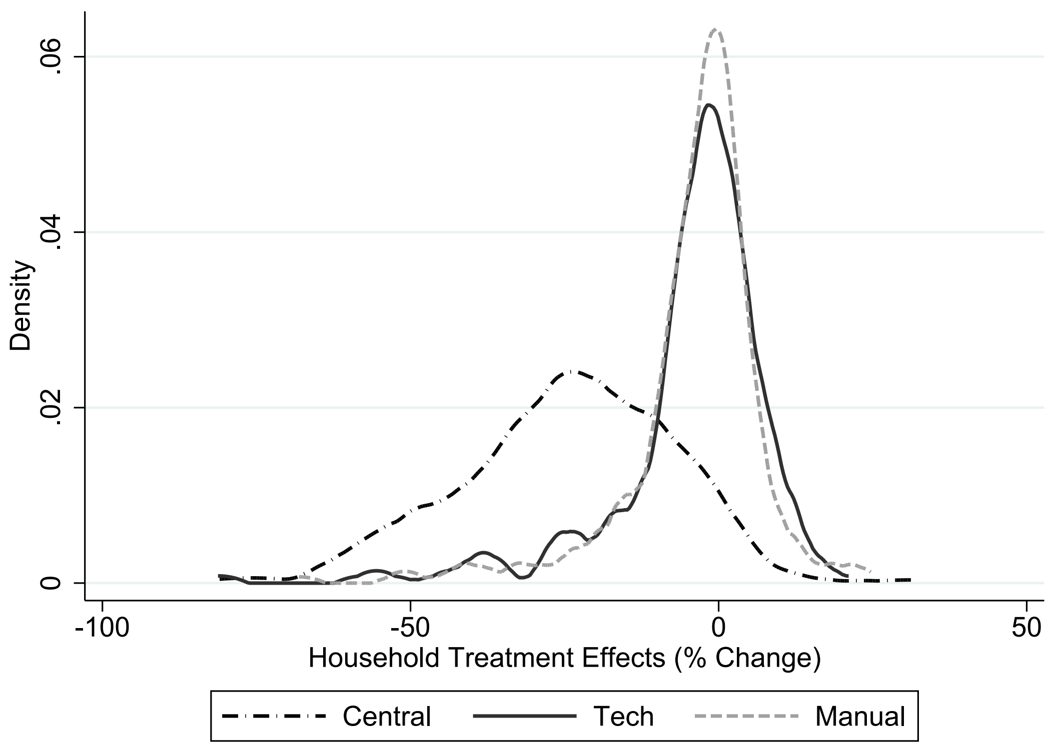
Our results have so far focused on average estimated treatment effects by demand response programs. A natural and important economic question is: What drives these differences across programs? In particular, what behavior underlies the differences between the average treatment effects of event offers applied to the Central vs. the Tech programs? In this section, we estimate household-level treatment effects and investigate the extent to which there is heterogeneity in event responsiveness across households. In particular, we use App interaction data during events to analyze the extent to which attention and effort relate to the household-specific estimated treatment effects.

[23] That said, participants were only eligible to receive $1 more for achieving a 30% reduction in electricity use during a High Evening vs. Evening event. The reward for achieving a 50% reduction was doubled ($6 vs. $3). It is possible that a larger scaling of incentives could induce a greater response. However, given our rewards fall in the range of wholesale price caps observed in practice, it is unlikely that incentives provided in a real-world setting would be considerably larger than the amounts we provided.

### 6.1 Household-Level Treatment Effects

Figure 4 provides the distributions of household-level treatment effects by program. This analysis provides several insights. We observe that the Tech and Manual programs' treatment effects are tightly distributed near zero. On average, households in the Tech and Manual programs reduce their consumption during events by 4.6% and 4%, respectively, which corresponds closely to the estimated program-level treatment effects in the previous section. Only 20% and 18% of the household-level estimated treatment effects are negative and statistically significant at the 5% level for the Tech and Manual programs, respectively.[24]

Figure 4. Household-Level Estimated Treatment Effect Distributions by Program




Notes. The reported results summarize the distribution of estimated household-level treatment effects obtained from estimating specification (3). The results are summarized using a non-parametric kernel density function using the Epanechnikov kernel function. The width of the density window is chosen to minimize the mean integrated squared error in the data.

Figure 4 reveals that, despite the fact that the Tech program has enabling devices,

[24]There is a small subset of households with positive estimated treatment effects. These estimated effects are systematically statistically insignificant, with only 3% being positive and statistically significant.

the *distribution* of their estimated treatment effects closely resembles that of the Manual households. This is striking, as even though the Tech and Manual programs have similar mean responses as seen in Section 5.1, we expected the Tech program would contain a subset of households with higher household-specific average treatment effects given their ability to remotely control their hot water heaters, EV chargers, and/or thermostats during events. However, this is not the case. In both the Tech and Manual programs, we observe a long left tail, suggesting there is a small subset of households with large estimated treatment effects that perform similarly, regardless of whether or not they have installed load controllers. These households are capable of delivering reductions that compare to the typical reductions seen from those in the Central program. In the next section, we leverage detailed App interaction data to provide suggestive evidence of what these higher-performing consumers are doing differently when achieving these larger reductions.

Figure 4 also re-confirms the larger response from households in the Central program. The average household-level treatment effect is a reduction of 24% during events for households in the Central program, closely reflecting the estimated effects in the program-level regressions above. In terms of distribution of household-level effects, Figure 4 shows that a larger number of Central households have considerably large estimated responses to events, with a leftward shifted distribution. 80% of the Central household-level estimated treatment effects are negative and statistically significant.

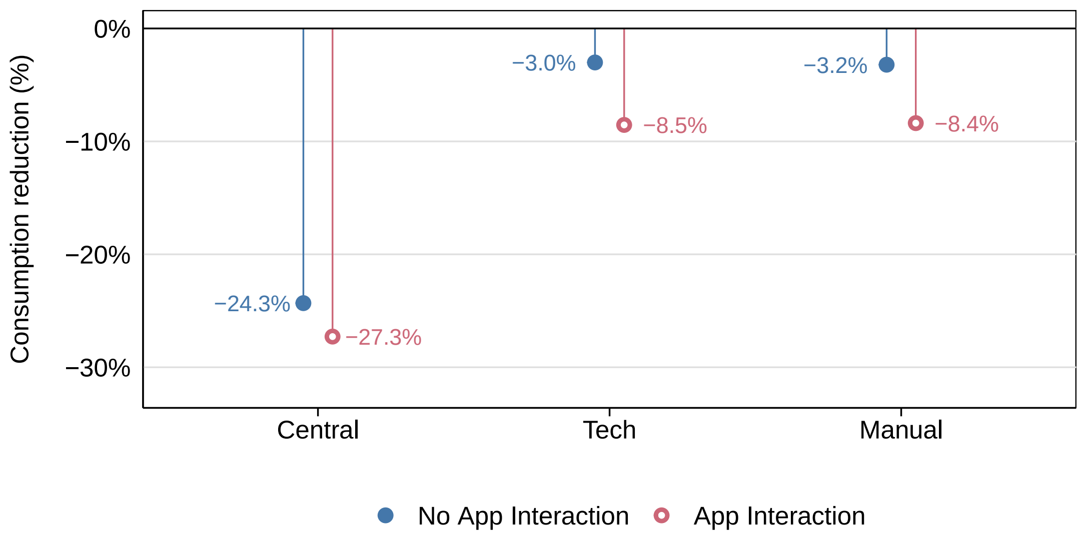
These results suggest, like those above, that the Central program outperforms both the Tech and Manual programs during events by a considerable margin. While there is variation in the estimated treatment effects among the Central households, the demand reductions during events are systematically larger than the other demand response programs.

### 6.2 App Interactions

Despite the lower average response to events for the Tech and Manual households, Figure 4 demonstrates there is a small subset of high performers in the Tech and Manual programs that have large negative estimated treatment effects. In this section, we leverage App interaction data to evaluate if these high-performers differ in their use of the App during events. Interaction with the App could be considered a strong indication that the household is aware of the demand response event and serves as a proxy for attention in this setting.

Recall the App data tell us when users are interacting with the App on a given day, as well as more details on which features of the App (i.e., pages/tabs) they are accessing. For both the Central and Tech households, interacting with the App allows them to control connected devices. For the Central program, the App can be used to opt out of automatic load control before or during events. In all programs, the App allows the household to observe the details of upcoming events 21 hours in advance, detailed information about household consumption in real-time, and performance in previous events.

Figure 5. Household-Level Average Treatment Effects by App Interaction




Notes. The reported results are average household-level treatment effects by whether or not the household interacted with the App on an event day. This represents the specification in equation (3), adjusted to interact the event indicator ($E_{it}$) with an App Indicator$_{it}$ variable that equals 1 when the household interacts with the App on the event day and zero otherwise. All specifications include fixed effects at the year-month, day-of-week, and hourly levels.

Figure 5 reports the average estimated household-level treatment effects, allowing for heterogenous treatment effects by whether or not the household interacted with the App on a given day. For the Central program, the average estimated household-level treatment effect is approximately -24% when the household does not interact with the App, increasing in magnitude to -27% when they do interact with the App during an event day. Whereas, for both the Manual and Tech programs, households only reduce demand on event days by 3% on average when they do not interact with the App, increasing in magnitude to roughly -8.5% when they interact with the App.

Overall, Figure 5 provides two key findings. First, there is a positive relationship between App interactions and demand reductions. Households that were more attentive to the App on event days achieved higher reductions. This differential effect is larger for the Tech and Manual groups, where the change in demand reductions was larger between interacting and not interacting with the App (approx. 8.5% vs 3%). Second, the "no interaction" estimates in the Central program participants and those in the other programs (-24.3% vs roughly -3%) indicates that the Central program received a roughly 21% "headstart" over the other programs, despite perhaps no attention paid to the event. This "headstart" is key to the overall finding of greater demand response by the Central program participants.

Table 3. Average Daily App Interaction Frequency by Program and Performance Quartiles

| Program | Performance Quartile | Household Count | General Interactions | Energy Usage Dial | Devices Tab | Advisor Tab |
|---|---|---|---|---|---|---|
| Central | 1 | 103 (58%) | 0.31 | 0.30 | 0.18 | 0.25 |
| | 2 | 49 (28%) | 0.23 | 0.21 | 0.14 | 0.18 |
| | 3 | 17 (10%) | 0.21 | 0.20 | 0.13 | 0.17 |
| | 4 | 8 (5%) | 0.17 | 0.15 | 0.11 | 0.11 |
| Tech | 1 | 17 (9%) | 0.61 | 0.54 | 0.37 | 0.47 |
| | 2 | 41 (22%) | 0.28 | 0.25 | 0.15 | 0.20 |
| | 3 | 65 (36%) | 0.21 | 0.20 | 0.11 | 0.17 |
| | 4 | 60 (33%) | 0.16 | 0.14 | 0.08 | 0.11 |
| Manual | 1 | 22 (9%) | 0.51 | 0.49 | 0.06 | 0.42 |
| | 2 | 63 (26%) | 0.20 | 0.19 | 0.03 | 0.16 |
| | 3 | 73 (30%) | 0.14 | 0.14 | 0.02 | 0.11 |
| | 4 | 84 (35%) | 0.16 | 0.15 | 0.03 | 0.13 |

Notes. The reported results provide the daily frequency of App Interactions by program and performance quartile. Performance quartiles are determined using the median percentage reduction in demand relative to the household's baseline. Household Count represents the number of households that fall within each performance quartile. The percentages report the percentage of households within a program that falls within each quartile. General Interactions reflect any interactions with the App. Energy Usage Dial displays the energy dial in the App that provides data on real-time usage. The Devices Tab displays a household's connected devices and allows households to adjust the use of the installed devices. The Advisor Tab reports information on upcoming events and historical performance on past events.

As noted above, we also have detailed data on the types of pages/tabs the households interact with in the App. This provides insight into the types of actions households may have taken to respond to events. Table 3 summarizes the average daily App interaction frequency for each demand response program separated by the performance quartile. The performance quartiles are determined by computing the median percentage reduction in the household's consumption relative to its baseline, looking across the three demand response groups and all events over our sample period. This separates households into categories of whether or not they are high or low-performing households during our experiment. We summarize the count of the households and the percentage of households enrolled in each demand response program that fall within a specific performance quartile. We report App interactions by four categories. "General Interactions" indicate whether the household interacted with the App at all on a given day. "Energy Usage Dial" reports whether the household looked at its real-time energy usage. The "Devices Tab" is the interface where households can control their installed devices. Finally, the "Advisor Tab" is the App location that provides information about upcoming events and a household's performance on past events.

Table 3 demonstrates that the vast majority of Central households (86%) fall within the top 50th percentile of performance. In contrast, only 31% and 35% of Tech and Manual households are in the top two quartiles. This is consistent with our results above that the Central program has a significantly greater demand response than the Tech and Manual programs, which show similar results.

Within each program, we see a reduction in the frequency of daily general interactions as we move down the performance quartiles. This suggests that higher-performing households are more attentive and/or undertaking effort to monitor and react to events. In fact, across all columns except for the Devices Tab column for the Central and Manual programs, the frequency of App interactions in the top quartile is significantly greater (at the 5% level) than the frequency in the second-highest quartile.

We see a considerably high frequency of App interactions among the Tech and Manual households that achieve the top quartile of performance, interacting with the App on 61% and 51% of days on average, respectively. In contrast, Central households in the top quartile only interact with the App on 31% of the days, suggesting that achieving this high threshold of performance required less effort and attention. The top quartile of the Central program interacts significantly less with the App than the top quartile of the Tech and Manual programs (at the 5% level). This is true for each App interaction column in the table, except for the Devices Tab.

Looking across the App categories, households interacted with the Energy Dial the most, followed by the Advisor tab. This suggests that when households used the App, they often monitored their real-time consumption and the details of upcoming events and/or their past performance. The top quartile performers in the Tech and Manual group stand out in both these categories, with the frequency of their interactions with both the Energy Dial and Advisor tabs being roughly 2 to 3 times those of the bottom 3 quartiles.

Tech households in the highest-performing quartile interacted with the Devices Tab in the App at a significantly higher frequency than all other households. This suggests that a (small) subset of households in this program were using the installed devices to achieve larger demand reductions. However, looking at the lower performance quartiles in the Tech program, the majority of households interacted with the Devices Tab about as much as those in the Central program.[25]

These results suggest that households in the Tech and Manual programs had to undertake a more active role in managing their consumption to achieve rewards in the top performance quartile. For the Central households, achieving these higher payoffs and demand reductions came with considerably fewer App interactions and, as a result, lower effort and attention utilization.

### 6.3 Opportunity Cost and Time Preferences

As described in Section **??**, we hypothesize that residential electricity consumers' willingness to respond to periodic peak pricing events is a function of whether doing so has a sufficiently high net benefit, given their opportunity cost of time, their preferences over their use of time, and the reward for participation. In this section, we explore whether our main results are consistent with this theory. We utilize data from a survey of participants conducted at the end of the experiment (in June 2023). Specifically, several survey questions were asked to capture respondents' income (a proxy for respondents' opportunity cost of time) and their stated value of participation in events. We use these data to estimate a model to evaluate the relationship between respondents' answers to these questions and their observed responsiveness to events, conditional on important controls such as the demand response program they were enrolled in and their household appliances.

Section 2.4 briefly describes the end of the experiment survey, with more detail provided in Appendix D.1. 71% of participants in the Central, Tech, and Manual groups filled out the survey, though respondents were not required to fill out every

[25] In addition to the installed devices from our experiment, a subset of households had other devices linked to the App. These were primarily smart plugs linked to lighting in the home. This helps explain why we observe interactions with the Devices Tab for the Manual program. Only 3 Manual households had installed and linked devices to control and monitor electric baseboard heaters during our sample.

question to complete the survey.[26]

To obtain information about income, we asked participants "What is your approximate household income?" and allowed them to choose between the following options: (1) Less than \$50k per year, (2) \$50-99k per year, (3) \$100-149k per year, (4) \$150-200k per year, (5) Over \$200k per year, and (6) Don't know/Rather not say.

To assess the extent to which participants felt that participating in events was worth their time, we asked the following ("Worth Time") question: "For the events you noticed, how often was it worth your time to participate by attempting to reduce your electricity consumption?" and gave them the following choices: (1) Never, (2) Sometimes, (3) About half the time, (4) Most of the time, (5) Always and (6) Don't know/Not Applicable. This allows us to assess how much of household, event-level reductions in electricity consumption are correlated with the perceived value that respondents placed on participating in events. Note that this value is impacted not only by participants' opportunity cost of time but also by their preferences; that is, it attempts to extract a net value of participation, as Section **??** suggests should explain the participation/effort put forth to respond to events.

Using data generated from survey responses, we estimate the following equation for each household $i$ in our three demand response groups:

$$Y_i = \beta_0 + \beta_1 I_i + \beta_2 Z_i + \beta_3 G_i + \beta X_i + \epsilon_i \tag{4}$$

in which $Y_i$ is the estimated treatment effect estimated for household $i$ (see Section 6.1), $I_i$ is a household's reported income from the survey, $Z_i$ is the household's response to the Worth Time question described above. $G_i$ is an indicator variable for the program in which the household is enrolled, and $X_i$ is a vector of additional control variables that include: an indicator variable for whether a household has an electric hot water heater, a categorical variable for whether a household has an electric vehicle and their corresponding charger type (No, Level 1, Level 2), a categorical variable for air conditioning in the home (No, Window Unit, Central Air), a categorical variable for electric baseboard heating (No, $1-3$ Units, 4 or more units), and a house/duplex dummy variable (1 for house/duplex, 0 for row home). These variables control for a household's ability to respond to events with electricity reductions, given the devices that they have at home.

The first coefficient of interest, $\beta_1$, indicates the relationship between responsive-

[26] This translates into 75%, 71%, and 69% of Central, Tech, and Manual group, respectively. Respondents were paid \$20 upon completion of the survey.

ness to events and income, conditional on the program in which a participant is enrolled as well as household appliances. The other coefficient of interest is $\beta_2$, which gives the relationship between responsiveness to events and participants' responses to the Worth Time survey question described above, conditional on a household's income. It yields the relationship between participants' perceived net benefit of participating in events and electricity reductions during events, conditional on the opportunity cost of time as measured by their income and the other controls. We report results with heteroskedastic robust standard errors.

In Appendix D.2, we evaluate if there are systematic differences across observable characteristics by survey response. In short, we find evidence to suggest that survey respondents had smaller levels of consumption on average compared to non-respondents. To the extent that consumption is correlated with income, this suggests that non-respondents had potentially higher income and a higher opportunity cost of time. The results in this section are applicable to households that filled out the survey and on average may have a lower opportunity cost of time.

Table 4 shows the results of estimating the Equation (4). For income, the baseline excluded category is income less than \$50,000. We see that, for survey respondents, higher income is associated with larger (positive) treatment effects, or smaller electricity consumption reductions during events. The magnitude of the coefficients are increasing in income. Coefficients for income brackets of \$150,000 per year or more are statistically different than the excluded lowest income bracket (less than \$50,000 per year). This is consistent with smaller household-level treatment effects from participants with a higher opportunity cost of time.

All coefficients on the Worth Time variable categories are negative and statistically different than the excluded "Never" category (with the "Sometimes" coefficient being marginally statistically different). Additionally, the coefficients are monotonically more negative in answers (higher frequencies of whether event participation was worth participants' time). This suggests that participants' perceived net benefits of event participation are correlated with greater electricity consumption reductions during events, conditional on income (opportunity cost of time). More intuitively, participants who reported viewing their response to events as worthwhile more often, perhaps due to their preferences for recurring electricity savings, reduced more electricity during events on average than others. This is consistent with the theory we present in Section **??**. More broadly, this points to an explanation for the difference in event-level treatment effects we see between the Central and Manual groups (Sec-

Table 4. Household Treatment Effect on Worth Time and Income (with controls)

| | Coefficient | Std. Error | P-Value |
|---|---|---|---|
| **Worth Time** | | | |
| Sometimes | -3.49 | 2.13 | 0.10 |
| Half of the Time | -6.30 | 2.39 | 0.01 |
| Most of the Time | -9.03 | 2.35 | 0.00 |
| Always | -15.05 | 2.86 | 0.00 |
| **Income** | | | |
| 50 - 99k | 5.53 | 4.56 | 0.23 |
| 100 - 149k | 7.32 | 4.55 | 0.11 |
| 150 - 200k | 9.89 | 4.55 | 0.03 |
| >200 k | 9.74 | 4.50 | 0.03 |
| **Group Indicators** | | | |
| Central | -17.34 | 1.74 | 0.00 |
| Tech | -1.37 | 1.57 | 0.38 |
| **Controls** | | | |
| **Electric Hot Water Heater** | -6.33 | 1.51 | 0.00 |
| **Electric Vehicle** | | | |
| Yes, Level 1 | -1.07 | 2.48 | 0.67 |
| Yes, Level 2 | -3.65 | 2.12 | 0.09 |
| **Air Conditioning** | | | |
| Yes, Window Unit | 1.69 | 1.74 | 0.33 |
| Yes, Central Air | 0.93 | 1.85 | 0.61 |
| **Baseboard Heating** | | | |
| Yes, 1 - 3 Units | 3.01 | 2.23 | 0.18 |
| Yes, 4 or more | 1.75 | 1.76 | 0.32 |
| **Home/Duplex** | 3.70 | 2.25 | 0.10 |

Notes. The reported results present the estimates for Equation (4).

tion 5.1): Automated responses to events can overcome the multitude of reasons why event participation is not worth consumers' time and effort.

# 7 Conclusion

As the electricity supply includes a growing share of variable renewable sources, the ability to alter electricity demand in time will become more valuable. Moreover, residential consumption from devices such as EVs, which allow consumption to be detached from device usage, provides an opportunity for such flexible demand (Bailey et al., 2023). However, consumer inattentiveness to dynamic electricity prices has long posed a problem for flexible demand to be meaningful.

Suspecting that inattention to dynamic pricing is rational—that consumers' rewards for paying attention to dynamic electricity prices and learning about how to respond to them are not worth the associated costs—we run a novel, large-scale framed field experiment that tests the efficacy of utility-managed ("centralized") electricity demand on consumer responses to dynamic prices. Centralized demand management has the potential to take the burden of actively responding to price signals off of consumers' shoulders while allowing them (as well as other consumers and grid operators, in critical conditions) to reap the rewards of timing consumption with changing electricity system conditions.

We find that customers participating in a centralized demand management program, the "Central" program, reduced consumption by 26% on average during critical "peak events". In contrast, participants in the "Tech" program, who had the same smart technology as those in the Central program to remotely control baseboard thermostats, hot water heaters, and EV chargers, but had to initiate reductions themselves, only reduced consumption by 5% during events. This difference indicates that centralized electricity demand management has large potential to help consumers overcome barriers to respond to electricity prices. We find that the take-up rates between the Central and Tech programs are not appreciably different, suggesting that centralized electricity management is not as unpalatable as one might expect.

Somewhat surprisingly, we find that participants in the Tech program reduce consumption during events no more than those in our "Manual" program, who do not have smart, remote device adjustment capability. This indicates that smart home energy technology was not sufficient on its own to induce demand flexibility; overcoming the key barrier of attention requires switching the default response, as per the Central program. The fact that both the Tech and Manual program participants reduce consumption by on average 5% during events is broadly consistent with the electricity demand response literature that has found similar magnitudes when consumers are faced with critical peak prices (Faruqui and Sergici, 2010; Yan et al., 2018). In summary, financial incentives motivate consumers to take some action, but consumers still face barriers that a centralized demand response program can resolve.

Interestingly, consumers in our context were not motivated to reduce consumption more when rewards were increased during periodic "high peak events". It is possible that consumers would need more than what we offered them to overcome the barriers that the Central program does. However, since our rewards were in line with peak electricity system prices, larger offers would not likely be economically efficient.

Given our experimental design using randomized household event schedules, we are able to estimate household-specific treatment effects to events. The distributions of household treatment effects across programs reveal that households in the Central program have a symmetrically distributed set of treatment effects, with the central mass of effects less than zero. In contrast, the Tech and Manual program participants display a distribution of household-level treatment effects that are centered around zero with a long left tail. This suggests that "high achievers" in these programs drove average treatment effects to events. Additionally, it suggests there is something about the Central program that facilitates the average household to respond to events by reducing consumption.

To understand the mechanisms behind our results, we use suggestive evidence from data on participant interaction with the experiment electricity management phone App. Across all programs, App interaction is correlated with larger household-level treatment effects. Average household-level treatment effects when households do not interact with the App are about 3% for the Tech and Manual programs and 24% for the Central program. When households *do* interact with their App, these numbers increase to about 8.5% for the Tech and Manual programs, and 27% for the Central program. This highlights the Central program's "headstart", whereby its participants achieve consumption reduction even in the absence of App interaction. We find that "high achievers" in the Tech and Manual programs, who drive the average consumption reductions during events for these programs, interact with their App on 61% and 51% of days on average during the experiment, whereas the high achievers in the Central program interacted with the App on average significantly less (31% of days on average). This suggests that high achievers in the Tech and Manual program devoted a lot of attention to their electricity consumption and effort in reducing it during events. Taken together, this evidence points to attention and effort (in the form of app interaction) being an important component of responsiveness to events, and that the Central program relieved participants of needing to devote such attention and effort to electricity management to make large consumption reductions during events.

Given our results, we surmise that programs and policies that relieve consumers of cognitive, time, and other burdens that contribute to rational inattention will have large potential to lead to welfare improvements. In the case of residential electricity, we see centralized demand as one such program, having the potential to both save money for consumers and facilitate flexible demand to meet emerging needs of electricity markets.

## References


**Bailey, Megan R, David P Brown, Blake C Shaffer, and Frank A Wolak**, "Show Me the Money! Incentives and Nudges to Shift Electric Vehicle Charge Timing," Technical Report, National Bureau of Economic Research 2023.

**Becker, Gary S**, "A Theory of the Allocation of Time," *The Economic Journal*, 1965, *75* (299), 493–517.

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

**Bojinov, Iavor, Ashesh Rambachan, and Neil Shephard**, "Panel Experiments and Dynamic Causal Effects: A Finite Population Perspective," *Quantitative Economics*, 2021, *12* (4), 1171–1196.

**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.

**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.

**Schneider, Ian and Cass R Sunstein**, "Behavioral Considerations for Effective Time-Varying Electricity Prices," *Behavioural Public Policy*, 2017, *1* (2), 219–251.

**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 Empirical Framework Material

## A.1 Recruitment and Assignment

The study sample was drawn from the population of residential customers in the Utility's service territory in and near a large metropolitan city in Canada. We employed a two-step recruitment strategy. In Phase 1, starting in August 2021, the Utility invited households to join an App operated by a third-party company in partnership with the Utility. The App provides households with household-level hourly consumption posted at a one-day lag. 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.

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. Recall, the set of controllable electric devices includes 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 programs.

In Phase 2 of recruitment, we randomized the eligible households into our treatment programs and never-treated groups.[27] Starting in October 2021, we sent program- and 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 program or 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 were also randomly offered a small sign-on incentive of the amounts $10 or $20, or no incentive. All households faced a yes/no decision regarding accepting our program- or group-specific offer. The never-treated Control group that received no equipment, price incentives, or real-time usage information (recall Table 1) received no further communication beyond joining the App in the first phase of recruitment.

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

This two-phase recruitment process occurred over the months of August 2021 - February 2022. The second phase of recruitment occurred in five waves starting in October 2021. As additional households joined the App, we collected the survey responses, identified eligible households, randomized households into programs and 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.

Finally, during the invitations to join each program or group, we randomized the upfront incentive. While we find a higher rate of initial acceptance with higher upfront incentive payments, the differences are small and not significantly different.[28]

[27] Specifically, we used a randomization procedure designed to balance important observable characteristics over programs and 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 the 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 program and group assignments so households within a cluster were balanced across programs and groups.

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

## A.2 Comparison of Household Characteristics Upon Randomization

We evaluate if there are differences in pre-treatment characteristics across our various programs and groups to assess the quality of our randomization. Table A1 provides summary statistics by program or group for a number of variables, including those used in the clustering procedure during randomization (recall the discussion in Footnote 27). The sample presented in this Table represents all 1,661 households invited to participate in the experiment. For all variables, we report the p-values from a one-way ANOVA test to evaluate if there are statistical differences in means across the programs/groups.[29]

Table A1 shows that we do not find statistically significant differences in key characteristics pre-treatment across our programs and groups. These results indicate that our randomization approach effectively achieved balance on observables pre-treatment. In addition, Table A1 demonstrates that the majority of households in our sample have electric hot water heating and use baseboard heating as the primary heat source. In contrast, 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 in summer. This demonstrates the potential for larger opportunities for load shifting during these months.

## A.3 Comparison of Household Characteristics After Acceptance

We compare the pre-treatment means in observable characteristics by program and group, including only the households that accepted our invitation to join each program. Large differences in observable characteristics would raise questions about the comparability of our estimated treatment effects from the main specifications.

Table A2 shows observables across groups for the final set of households in each program/group. We observe limited differences in these characteristics across programs/groups. The exceptions are that we find a statistically significant difference 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 program than

[29] The seasonal cumulative consumption and load factor data only contain households with a full year's worth of historical consumption. We computed analogous statistics for the entire sample of households 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 programs/groups using this data.

Table A1. Comparison of Means by Programs/Group - Initial Randomization

| | Central | Tech | Manual | Info | Control | ANOVA (p-value) |
|---|---|---|---|---|---|---|
| Cumul. kWh | | | | | | |
| Winter | 5,279 | 5,268 | 5,442 | 4,859 | 5,265 | 0.27 |
| | (2,694) | (3,032) | (3,076) | (2,748) | (2,950) | |
| Spring | 3,760 | 3,773 | 3,818 | 3,503 | 3,712 | 0.48 |
| | (1,924) | (2,112) | (1,911) | (2,116) | (1,974) | |
| Summer | 2,845 | 2,836 | 2,708 | 2,614 | 2,729 | 0.54 |
| | (1,742) | (1,872) | (1,539) | (1,861) | (1,710) | |
| Fall | 3,633 | 3,670 | 3,700 | 3,458 | 3,623 | 0.66 |
| | (1,663) | (1,945) | (1,974) | (1,796) | (1,860) | |
| Load Factor | | | | | | |
| Winter | 24.66 | 24.98 | 25.41 | 24.73 | 24.67 | 0.81 |
| | (8.20) | (8.15) | (8.80) | (8.29) | (8.63) | |
| Spring | 19.52 | 20.12 | 20.01 | 19.28 | 19.91 | 0.65 |
| | (7.25) | (6.97) | (6.70) | (7.73) | (7.41) | |
| Summer | 16.82 | 16.55 | 16.73 | 16.12 | 16.32 | 0.82 |
| | (7.89) | (6.30) | (5.93) | (8.11) | (8.29) | |
| Fall | 18.56 | 18.90 | 19.34 | 18.42 | 19.06 | 0.42 |
| | (5.89) | (6.23) | (6.00) | (6.48) | (6.50) | |
| Electric Vehicle | 0.27 | 0.27 | 0.27 | 0.33 | 0.27 | 0.41 |
| | (0.44) | (0.45) | (0.45) | (0.47) | (0.45) | |
| Baseboard Heating | 0.61 | 0.64 | 0.61 | 0.63 | 0.63 | 0.95 |
| | (0.49) | (0.48) | (0.49) | (0.48) | (0.48) | |
| Air Conditioning | 0.52 | 0.51 | 0.50 | 0.51 | 0.54 | 0.95 |
| | (0.50) | (0.50) | (0.50) | (0.50) | (0.50) | |
| Electric Hot Water | 0.70 | 0.66 | 0.70 | 0.66 | 0.72 | 0.38 |
| | (0.46) | (0.47) | (0.46) | (0.47) | (0.45) | |
| House/Duplex | 0.77 | 0.76 | 0.81 | 0.78 | 0.84 | 0.17 |
| | (0.42) | (0.43) | (0.39) | (0.42) | (0.37) | |
| Median Income | 86,376 | 88,291 | 85,931 | 87,470 | 85,948 | 0.48 |
| | (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 programs/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. 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. Median Income reports the median household-level income of the Census Dissemination Area where the household is located. ANOVA reports the p-value from one-way ANOVA tests for differences in means across programs/groups. Statistical significance: $^{*}\,p < 0.10$, $^{**}\,p < 0.05$, and $^{***}\,p < 0.01$.

in other programs/groups, in particular. We also observe a difference across programs/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 due to the initial randomization largely remains in the final sample.

Table A2. Comparison of Means by Program/Group - Final Accepted Households

| | Central | Tech | Manual | Info | Control | ANOVA (p-value) |
|---|---|---|---|---|---|---|
| Cumul. kWh | | | | | | |
| Winter | 5,507 | 5,302 | 5,422 | 5,037 | 5,265 | 0.71 |
| | (2,706) | (2,737) | (3,240) | (2,768) | (2,950) | |
| Spring | 3,900 | 3,739 | 3,797 | 3,642 | 3,712 | 0.85 |
| | (1,934) | (1,791) | (1,939) | (2,159) | (1,974) | |
| Summer | 2,851 | 2,672 | 2,766 | 2,702 | 2,729 | 0.93 |
| | (1,869) | (1,759) | (1,547) | (1,849) | (1,710) | |
| Fall | 3,754 | 3,550 | 3,677 | 3,547 | 3,623 | 0.86 |
| | (1,733) | (1,659) | (1,992) | (1,788) | (1,860) | |
| Load Factor | | | | | | |
| Winter | 24.62 | 25.56 | 24.93 | 24.93 | 24.67 | 0.90 |
| | (8.68) | (8.21) | (9.04) | (7.76) | (8.63) | |
| Spring | 19.33 | 20.48 | 19.80 | 19.59 | 19.91 | 0.72 |
| | (7.43) | (6.30) | (6.45) | (7.05) | (7.41) | |
| Summer | 16.33 | 16.87 | 16.95 | 16.61 | 16.32 | 0.91 |
| | (8.54) | (6.02) | (5.95) | (7.80) | (8.29) | |
| Fall | 18.17 | 18.97 | 19.11 | 18.53 | 19.06 | 0.65 |
| | (6.27) | (5.78) | (6.29) | (6.11) | (6.50) | |
| Electric Vehicle | 0.25 | 0.21 | 0.30 | 0.34 | 0.27 | 0.07* |
| | (0.43) | (0.41) | (0.46) | (0.47) | (0.45) | |
| Baseboard Heating | 0.68 | 0.70 | 0.60 | 0.59 | 0.63 | 0.12 |
| | (0.47) | (0.46) | (0.49) | (0.49) | (0.48) | |
| Air Conditioning | 0.46 | 0.46 | 0.51 | 0.51 | 0.54 | 0.41 |
| | (0.50) | (0.50) | (0.50) | (0.50) | (0.50) | |
| Electric Hot Water | 0.75 | 0.74 | 0.68 | 0.65 | 0.72 | 0.16 |
| | (0.43) | (0.44) | (0.47) | (0.48) | (0.45) | |
| House/Duplex | 0.82 | 0.77 | 0.89 | 0.84 | 0.84 | 0.02** |
| | (0.39) | (0.42) | (0.32) | (0.37) | (0.37) | |
| Median Income | 84,978 | 88,274 | 86,718 | 89,504 | 85,948 | 0.23 |
| | (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 programs/groups for households that were in our final programs/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. Median Income reports the median household-level income of the Census Dissemination Area where the household is located. ANOVA reports the p-value from one-way ANOVA tests for differences in means across programs/groups. Statistical Significance $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

# B Treatment Details

## B.1 Program-Specific Event Notifications

Each treatment program experienced event notifications tailored to their treatment. Each program received a notification 21 and 2 hours before an event. All participants were shown a short notification according to their device and in-app notification settings. If participants touched and pressed the notification, they were shown the long notification specific to their program, featured below, with event incentive details.

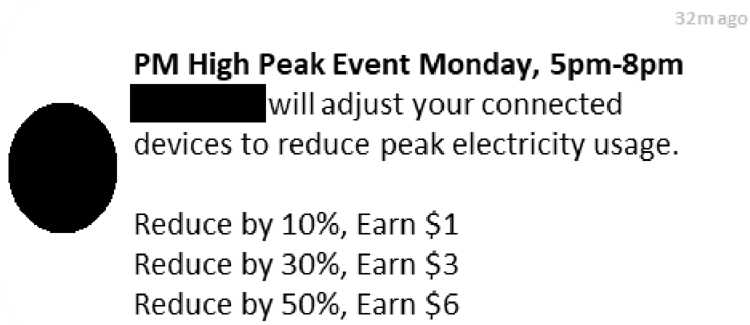



Figure B1. Long Notification for Central program

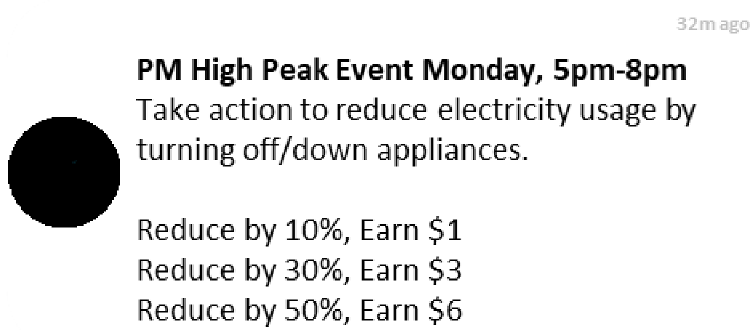



Figure B2. Long Notification for Tech program

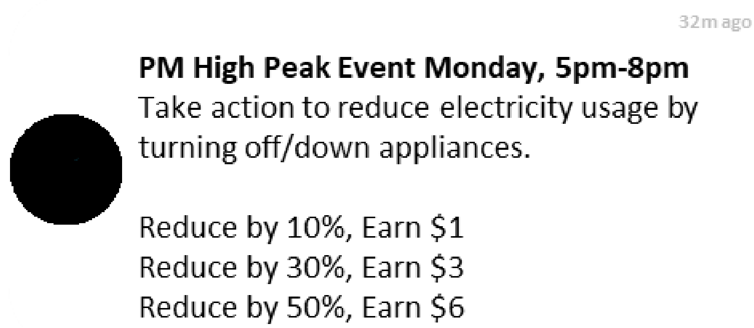



Figure B3. Long Notification for Manual program

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

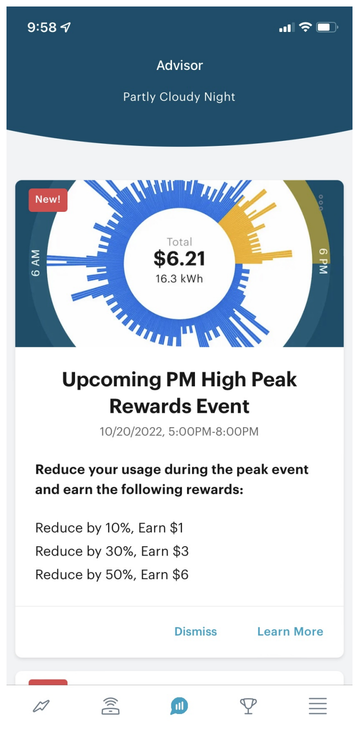



Figure B4. Event info in App

## B.2 Treatment Program-Specific App Functionality

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

### B.2.1 Central Program

The Central program 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 program 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 (Figure B5). (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.

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, via the icons above the text "You are opted in", "Event", and "Complete" (Figure B6).

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

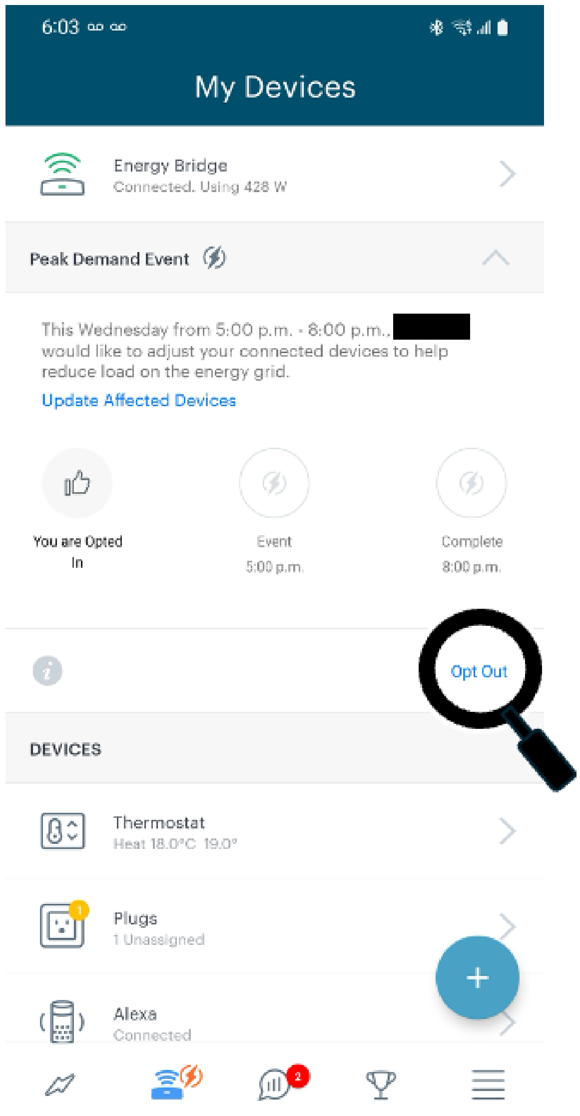



Figure B5. Central Program Opt-Out Functionality

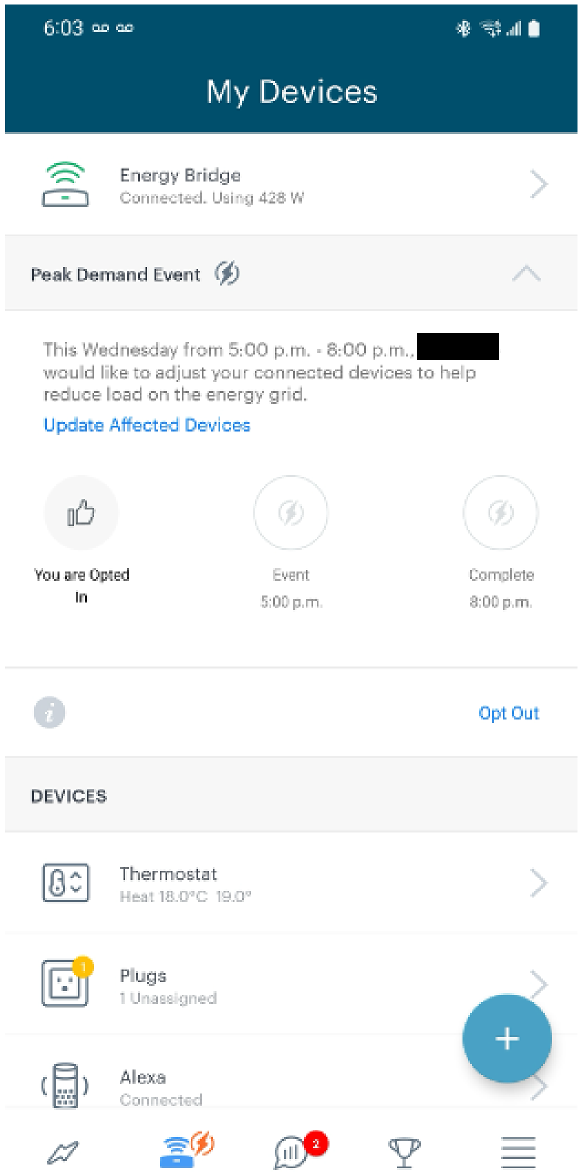



Figure B6. Central Program Event Experience

### B.2.2 Tech Program

The Tech program 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 program can remotely control any device that has an installed load controller through the App. 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 B7 below for the instructions sent to participants that explain these actions.) Tech program 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.)

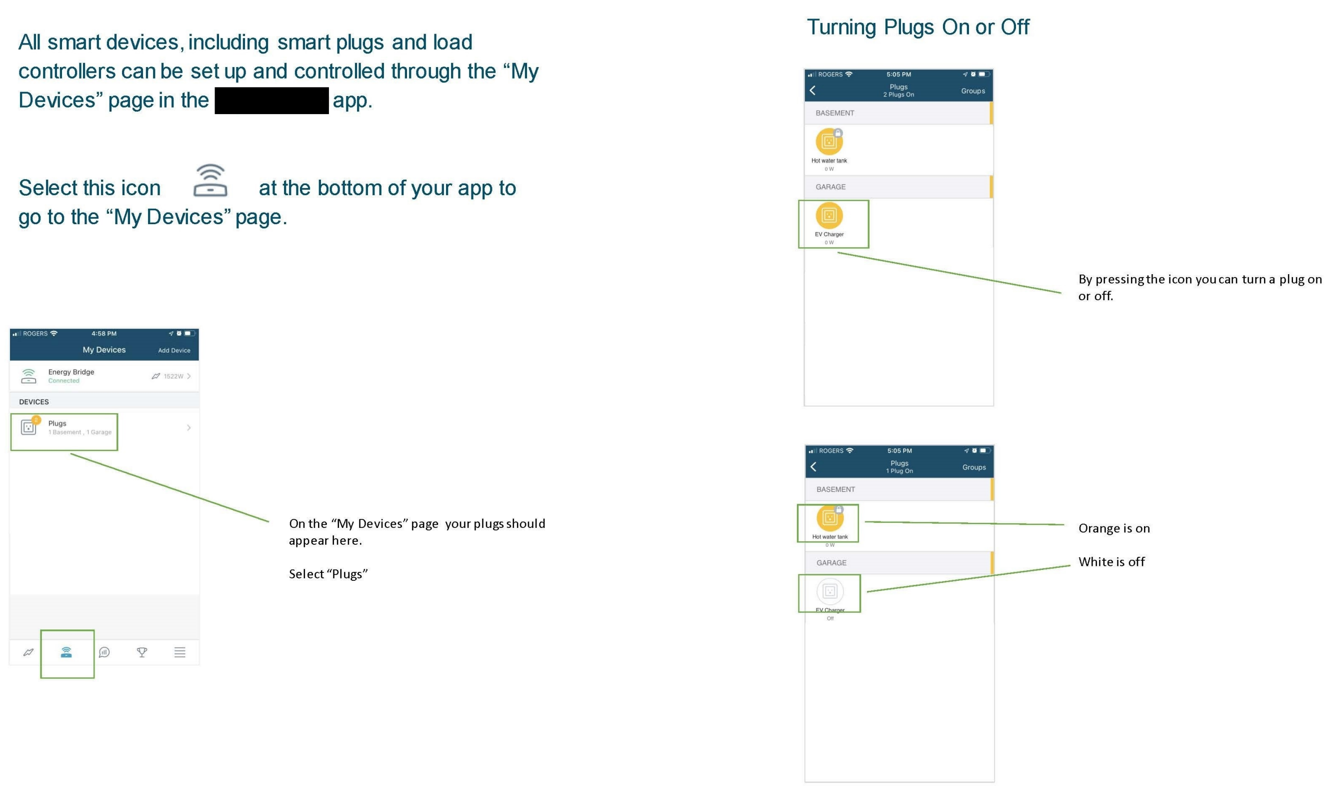



Figure B7. Controller Guide for Tech Program

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

### B.2.3 Manual Program

The Manual program 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 program 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 program participants install their own smart home devices, they may be able to link them to the smart electricity consumption technology ecosystem used in this experiment. If so, they may have the capabilities of the Tech program to observe the real-time consumption of those devices/devices individually and adjust them remotely through the App. (Only three households in the Manual program installed their own smart thermostats over our sample period.)

### B.2.4 Central, Tech, and Manual Programs

After each event, all three of the Central, Tech, and Manual programs 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 B8.

From this rewards screen, participants can select “Event History” and see their recent history of event rewards, as shown in Figure B9.[30]

[30] Figure B9 was created for illustrative purposes using a series of simulated events. As a result, the event times differ from the event times considered in our study (i.e., 7:00 AM - 10:00 AM and 5:00 PM - 8:00 PM).

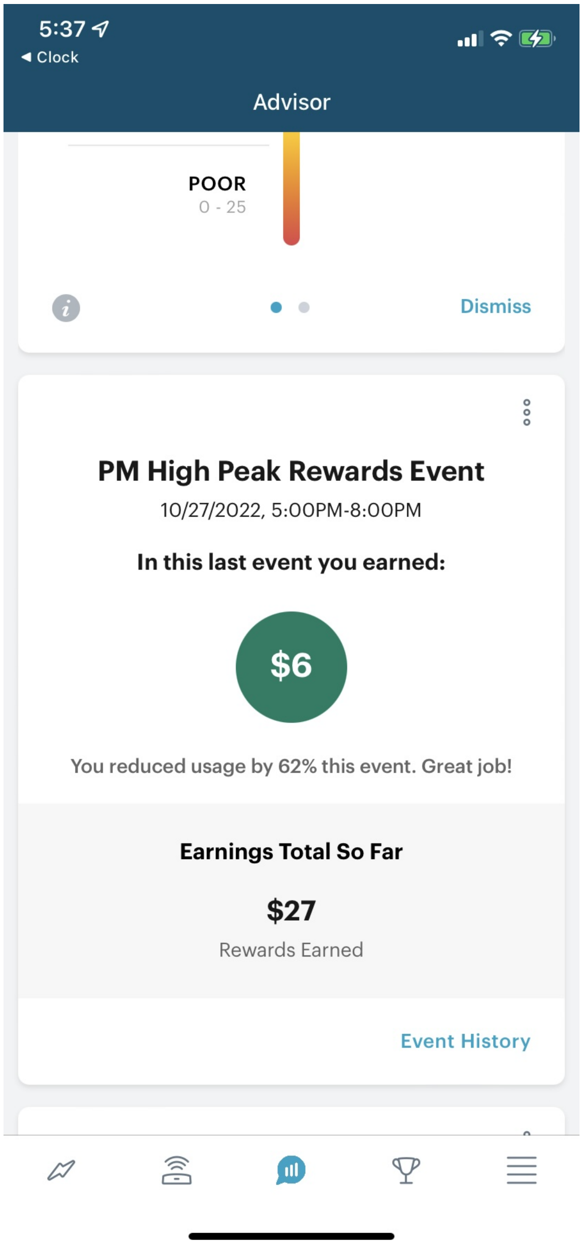



Figure B8. Rewards Screen

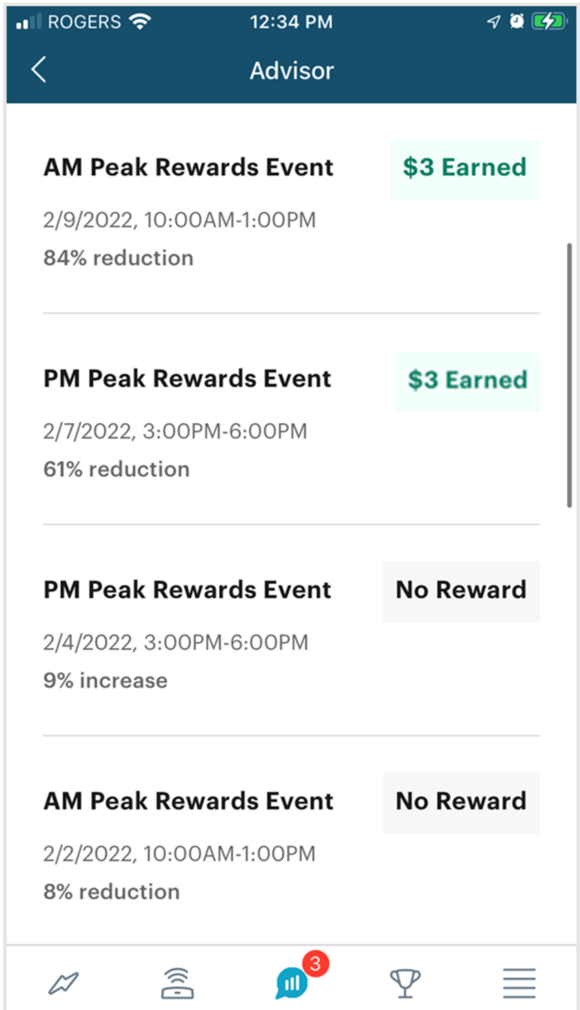



Figure B9. Event History

# C Extensions and Robustness

## C.1 Comparison Group in Program-Level Regression

In our main specification in Equation (1), our analysis includes all three demand response groups and never-treated households. As discussed in Section 4.2, this approach compares event time consumption to non-event time consumption of households in the same demand response program, other demand response programs, and never-treated households. Tables C1 and C2 present our estimated treatment effects of participants by program to events and separated by event type, respectively, allowing for regressions only including households in the same demand response group (Column (1)), same demand response group and the never-treated (Column (2)), and the results from our main specification in Column (3) for comparison purposes. Our results are consistent across all three specifications.

Table C1. Treatment Effects of Participants by Program

| Program | (1) | (2) | (3) |
|---|---|---|---|
| Central | -0.3151*** | -0.3007*** | -0.3047*** |
| | (0.0206) | (0.0204) | (0.0204) |
| Tech | -0.0661*** | -0.0475*** | -0.0495*** |
| | (0.0132) | (0.0155) | (0.0159) |
| Manual | -0.0507*** | -0.0459*** | -0.0540*** |
| | (0.0092) | (0.0117) | (0.0122) |
| **Comparisons** | | | |
| Own Program | Y | Y | Y |
| Other Treated | | | Y |
| Never Treated | | Y | Y |

Notes. The reported results are program-specific treatment effect coefficients. Standard errors are reported in the parentheses and clustered at the household level. Column (1) reports the regression results including only within demand response program comparisons, column (2) includes both within demand response program and the never-treated (Info and Control) groups, and column (3) reports results when all programs/groups are included. All specifications include fixed effects at the household and hour-of-sample levels. Statistical Significance $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table C2. Treatment Effects of Participants by Program and Event-Type

| Program | (1) | (2) | (3) |
|---|---|---|---|
| **Central** | | | |
| Morning | -0.3198*** | -0.3160*** | -0.3133*** |
| | (0.0221) | (0.0224) | (0.0223) |
| Evening | -0.3040*** | -0.2850*** | -0.2916*** |
| | (0.0218) | (0.0226) | (0.0227) |
| High Evening | -0.3291*** | -0.3105*** | -0.3177*** |
| | (0.0237) | (0.0241) | (0.0243) |
| **Tech** | | | |
| Morning | -0.0495*** | -0.0273 | -0.0158 |
| | (0.0128) | (0.0188) | (0.0199) |
| Evening | -0.0713*** | -0.0545*** | -0.0623*** |
| | (0.0155) | (0.0181) | (0.0185) |
| High Evening | -0.0786*** | -0.0604*** | -0.0675*** |
| | (0.0166) | (0.0199) | (0.0205) |
| **Manual** | | | |
| Morning | -0.0488*** | -0.0689*** | -0.0812*** |
| | (0.0097) | (0.0150) | (0.0159) |
| Evening | -0.0488*** | -0.0332** | -0.0396*** |
| | (0.0108) | (0.0143) | (0.0149) |
| High Evening | -0.0567*** | -0.0403** | -0.0467*** |
| | (0.0126) | (0.0159) | (0.0166) |
| **Comparisons** | | | |
| Own Program | Y | Y | Y |
| Other Treated | | | Y |
| Never Treated | | Y | Y |

Notes. The reported results are program-specific treatment effect coefficients by event type. Standard errors are reported in the parentheses and clustered at the household level. Column (1) reports the regression results including only within demand response program comparisons, column (2) includes both within demand response program and the never-treated (Info and Control) groups, and column (3) reports results when all programs/groups are included. Each regression is adjusted to include an event-type-specific categorical variable. All specifications include fixed effects at the household and hour-of-sample levels. Statistical Significance $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

### C.2 No Carryover Assumption

As discussed in Section 4.2, an assumption in our identification strategy is that our randomized events do not impact (or "carryover" to have a treatment effect on) persistent changes in behavior in the event hours on non-event days. Table C3 presents the results of our no carryover assumption DID test, described in Section 4.2. Table C4 presents the results when we allow for differential effects across the morning and evening event windows on non-event days. In both specifications, we find no evidence of changes to non-event day consumption during the event windows.

Table C3. Carry Over DID Estimates by Program

| | Central | Tech | Manual |
|---|---|---|---|
| Event Window | 0.0241 | 0.0302 | -0.0014 |
| | (0.0169) | (0.0181) | (0.0155) |

Notes. The reported results are the program-specific Event Window coefficients from equation (2). For each demand response program, the sample includes households from their own treatment program and the never-treated groups. Standard errors are reported in the parentheses and clustered at the household level. All specifications include fixed effects at the household and hour-of-sample levels. Statistical Significance $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table C4. Carry Over DID Estimates by Program and Event Type

| | Central | Tech | Manual |
|---|---|---|---|
| Morning Event Window | 0.0188 | 0.0353 | -0.0226 |
| | (0.0206) | (0.0235) | (0.0201) |
| Evening Event Window | 0.0295 | 0.0251 | 0.0196 |
| | (0.0212) | (0.0210) | (0.0185) |

Notes. The reported results are the program-specific Event Window coefficients from equation (2), allowing for differential effects during the morning and evening event windows. For each demand response program, the sample includes households from their own treatment program and the never-treated groups. Standard errors are reported in the parentheses and clustered at the household level. All specifications include fixed effects at the household and hour-of-sample levels. Statistical Significance $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

### C.3 Intention-to-Treat and Local Average Treatment Effect Methodology

Because we randomized program offers to households as described in Appendix A.1, we can estimate program-level treatment effects. We estimate the effect of each program offer on event-level consumption (an intention-to-treat (ITT) estimate) as well as the effect of each program on event-level consumption (a local average treatment effect (LATE) estimate). Because our events where randomized at the household-level, we modify the standard approaches to estimating these by using "synthetic events" for households that were offered but not enrolled in a program.

#### C.3.1 Intention-to-Treat

We estimate a regression specification that allows us to estimate an effect similar to an intention-to-treat (ITT) estimate. Unlike many experimental settings, our treatment (events) are randomized within demand response programs for households that participated in our experiment. Therefore, unlike standard ITT specifications, we cannot have a binary treatment indicator that turns on for all households assigned to a specific experimental program, regardless of whether they accept that program. Because we have random, periodic events that are randomly assigned to only those households that chose to participate, we create an analogous environment in our setting for all households invited to each program. To do this, we assign households that were randomized to receive the Central, Tech, and Manual program offers but did not accept the offer a distribution of randomized "synthetic" 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 synthetic events for households that were invited to the Central, Tech, or Manual programs but did not accept our offer.

We estimate the following equation, separately for each demand response program and including the never-treated households that were not subject to events:[31]

$$ln(c_{it}) \;=\; \gamma + \beta_{ITT}\, \text{Assigned}_i \cdot \widehat{E}_{it} + \alpha_i + \tau_t + \delta X_t + \varepsilon_{it} \tag{5}$$

where $\text{Assigned}_i$ is an indicator variable that equals one if the household was assigned to receive a program offer (i.e., the Central, Tech, or Manual offer) and zero otherwise.

[31] We run two additional specifications of this analysis where we only include households within each demand response program, and all demand response programs and the never-treated simultaneously as in equation (1). Our ITT results are robust to these alternative specifications.

$\widehat{E}_{it}$ is the household-specific event indicator that equals one if household $i$ experiences an event in hour $t$ (or is assigned a synthetic event, for households that did not accept the program offer) and zero otherwise. For each program, 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 fixed effects at the household and hour-of-sample levels; $X_t$ the same vector of hourly weather controls. We cluster standard errors at the household level.

We also consider a modified version of equation (5) that permits event type-specific treatment effects. More specifically, $\widehat{E}_{it}$ is now a household-specific set of indicator variables that denote whether the household is experiencing a morning, evening, or high evening event. For households that did not enter into our final demand response programs, the (synthetic) allocation of the three event types is randomized with the same frequency distribution as for households that did accept the program offers and receive the event treatments. All other details of the regression specification remain the same.

#### C.3.2 Local Average Treatment Effect

In addition to estimating the ITT, we also estimate a Local Average Treatment Effect (LATE) for each of our demand response programs. The same empirical challenges discussed in the ITT estimation apply here. That is, households that did not choose to participate in the experiment did not receive the (randomized) events. These households are included in this regression, and we use the same synthetic events approach detailed above.

Since our program assignment was randomized, we could estimate the following equation to identify a program-level estimate of the effect of each demand response program on a event-time consumption if we had 100% compliance with program assignment. However, since we had less than a 100% take-up rate, the following equation serves as our second-stage in an instrumental variables (IV) framework in which we instrument treatment assignment for take-up. We estimate the following second-stage equation, separately for each demand response program and including the never-treated households that were not subject to events:

$$ln(c_{it}) \;=\; \gamma + \beta_{LATE}\,\text{Treat}_i \cdot \widehat{E}_{it} + \alpha_i + \tau_t + \delta X_t + \varepsilon_{it} \qquad (6)$$

where $\text{Treat}_i$ is an indicator variable that equals 1 if household $i$ was assigned to and accepted a demand response program invitation and zero otherwise. $\widehat{E}_{it}$ is our synthetic event indicator variable that equals 1 if the individual was subject to a treatment event and zero otherwise. All other aspects of the regression specification are analogous to that specified in equation (5) above.

Because program offer acceptance is endogenous with event responsiveness, we must use an IV approach to establish a causal estimate with $\beta_{LATE}$. We use the randomized, initial invitation to join the demand response program as our IV. More specifically, we rely on the exclusion restriction that the initial (randomized) assignment and invitation to join a specific demand response program affects consumption only via its impact on participation in treatment events.

More formally, for each demand response program, we estimate the following first-stage equation:

$$\text{Treat}_i \cdot \widehat{E}_{it} \;=\; \mu + \theta\,\text{Assigned}_i \cdot \widehat{E}_{it} + \pi_i + \kappa_t + \omega X_t + \eta_{it} \tag{7}$$

where $\text{Assigned}_i$ equals 1 for households assigned to the relevant demand response program and zero otherwise. $\pi_i$ reflects our household fixed effects, $\kappa_t$ is our hour-of-sample fixed effects, and $X_t$ are our weather covariates. $\widehat{E}_{it}$ is our synthetic event variable. The main objective of the first-stage equation is to establish an IV for the variable $\text{Treat}_i$ that interacts with the exogenous (randomized) variable $\widehat{E}_{it}$ in our LATE specification. Consequently, the first-stage equation estimates this endogenous interaction using the instrument $\text{Assigned}_i$.

We also consider a modified version of equations (6) and (7) that permits event type-specific treatment effects. More specifically, in both equations, $\widehat{E}_{it}$ is a household-specific set of indicator variables that denote whether the household is experiencing a morning, evening, or high evening event. This implies that we have three endogenous regressors and instruments for each event type. Analogous to the ITT analysis above, for households that did not enter into our final demand response programs, the (synthetic) allocation of the three event types is randomized in the same manner as was done for those households that did receive the event treatments. All other details of the regression specification remain the same.

## C.4 Intention-to-Treat and Local Average Treatment Effect Results

### C.4.1 Intention-to-Treat

Table C5 provides the ITT estimates. The coefficients imply that offering households the opportunity to join the Central, Tech, and Manual demand response programs results in an average reduction in consumption of -13%, -3%, and -3% during events, respectively, from each program of invited households.[32] These percent changes are what an electric utility company can expect "at the end of the day" in terms of event-time consumption reductions after both making offers such as these and conducting event-time alerts. That is, these estimates include both the extensive margin of program offer acceptance and intensive margin of household electricity consumption reductions from those who accepted the offer.

These results, like our main event-level treatment effect results above, indicate centralized electricity consumption reductions result in an expected average reduction in demand during events that is several orders of magnitude larger than non-centralized programs, without or without the same technology. The ITT estimates here show uniquely that the relative advantage of centralized demand is still realized, even when one considers the differential take-up rates across programs. In addition, the Tech and Manual programs have ITT estimates that are similar in magnitude, consistent with the findings in our main treatment effect estimates.

Table C5. ITT Estimates by Program

| | Central | Tech | Manual |
|---|---|---|---|
| Assigned × Event | -0.1435*** | -0.0342*** | -0.0322*** |
| | (0.0141) | (0.0102) | (0.0088) |

Notes. The reported results are program-specific ITT coefficients from equation (5). For each demand response program, the sample includes households from their own treatment program and the never-treated groups. Standard errors are reported in the parentheses and clustered at the household level. All specifications include fixed effects at the household and hour-of-sample levels. Statistical Significance $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

The estimates for the ITT are considerably lower than the average treatment effect of participants in our main analysis. The events are randomized, and the households that did not accept the offers to enroll in the demand response programs

[32] These calculations reflect the relevant coefficient estimate $\widehat{\beta}$ reported in Table C5, transformed to percentage change in consumption as follows: $100 * (\, exp(\widehat{\beta}) - 1\,)$.

had no knowledge of the events. These households are expected to have (noisy) null responses during events. The ITT estimates can be approximately recovered by multiplying the average treatment effects on participating households by the probability that a household accepted our offer to join a demand response program. Using the acceptance rates in Table 2 and the estimated average treatment effects described in Section **??**, this calculation yields approximate ITT estimates, in percentage terms, of 11%, 2%, and 3% for the Central, Tech, and Manual programs, respectively.[33]

Finally, Table C6 reports the ITT estimates by event type. The findings parallel the qualitative conclusions from our main average treatment effects on the households that participated in our experiment, with smaller estimated effects due to the ITT including the extensive acceptance margin, as described above. The estimated ITT coefficients represent an average reductions in consumption during events ranging from 12% to 15% for the Central program, 3% to 4% for the Tech program, and 2% to 5% for the Manual program, respectively. We continue to find modest differences across event types for each demand response program. Further, for each program, we do not observe any evidence of a statistically significant higher response to increased financial incentives during High Evening events.

Table C6. ITT Estimates by Program and Event Type

| | Central | Tech | Manual |
|---|---|---|---|
| Assigned × Morning Event | -0.1659*** | -0.0276** | -0.0481*** |
| | (0.0156) | (0.0131) | (0.0114) |
| Assigned × Evening Event | -0.1279*** | -0.0366*** | -0.0237** |
| | (0.0155) | (0.0122) | (0.0109) |
| Assigned × High Evening Event | -0.1437*** | -0.0382*** | -0.0275** |
| | (0.0168) | (0.0137) | (0.0122) |

Notes. The reported results are program-specific ITT coefficients from equation (5), allowing for event type-specific estimates. For each demand response program, the sample includes households from their own treatment program and the never-treated. Standard errors are reported in the parentheses and clustered at the household level. All specifications include fixed effects at the household and hour-of-sample levels. Statistical Significance $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

[33]For the Central, Tech, and Manual programs, these values reflect $26\% * 0.42 = 11\%$, $5\% * 0.48 = 2.4\%$, and $5\% * 0.59 = 3\%$, respectively.

### C.4.2 Local Average Treatment Effect

Table C.4.2 presents results from estimating Equation 6. The large F-statistics indicate that our first stage (Equation (7)) is strong. Coefficients for each program can be interpreted as the average impact of each demand response program on event-time consumption. Based on these coefficients, the Central, Tech, a Manual programs induce their participants to reduce consumption by 25.5%, 5.7%, and 4.8%, respectively. The striking similarity between these results and our main results (Table C1), which can be interpreted as the average causal effect of event notifications on event-time consumption by selected program, suggests that selection into our demand response programs does not play a major role in the differences of the event-level treatment effects across programs.

These results continue to demonstrate that automated options (such as the Central program) can help consumers respond much more deeply to electricity prices than standard programs (like the Manual program). Smart technology that enables customers to view device-specific consumption and control devices remotely, like load controllers that were given to the Tech program, will not necessarily resolve barriers to price response. From the perspective of a utility company, these LATE results are useful in understanding the event-time consumption reductions that they can expect from similar programs when selection may not be involved in customer participation, such as when customers are defaulted into such programs.[34]

Table C8 shows that the impact of each program on event-time consumption does not vary by the timing of the event or the amount of incentives offered. The relative advantage of the Central program over the other two remains across event times. That story does not change when one focuses on the events with larger incentives, the High Evening events.

[34]Note that in our setting, however, defaulting customers into the Central and Tech programs would not have been possible, given the installation required for the load controllers. As appliance and automation technology evolves, it is quite possible that defaulting customers into such programs may be possible in the future. Given that our setting required this installation, we very cautiously interpret these results as indicative of causal program effects in a setting that would not require customer opt-in.

Table C7. LATE Estimates by Program

| | Central | Tech | Manual |
|---|---|---|---|
| Treated × Event | -0.2938*** | -0.0583*** | -0.0489*** |
| | (0.0233) | (0.0174) | (0.0133) |
| F-Stat | 333.6153 | 441.0571 | 698.1311 |

Notes. The reported results are program-specific LATE coefficients from equation (6). For each demand response program, the sample includes households from their own treatment program and the never-treated groups. Standard errors are reported in the parentheses and clustered at the household level. All specifications include fixed effects at the household and hour-of-sample levels. F-Stat represents the F-statistic from the first-stage IV regression. Statistical Significance $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table C8. LATE Estimates by Program and Event Type

| | Central | Tech | Manual |
|---|---|---|---|
| Treated × Morning Event | -0.3394*** | -0.0468** | -0.0732*** |
| | (0.0276) | (0.0223) | (0.0173) |
| Treated × Evening Event | -0.2624*** | -0.0625*** | -0.0361** |
| | (0.0269) | (0.0207) | (0.0165) |
| Treated × High Evening Event | -0.2935*** | -0.0652*** | -0.0418** |
| | (0.0291) | (0.0232) | (0.0184) |
| F-Stat 1 | 118.6108 | 151.1503 | 232.0957 |
| F-Stat 2 | 113.2662 | 145.8802 | 234.5316 |
| F-Stat 3 | 115.0813 | 149.1091 | 233.7128 |

Notes. The reported results are program-specific LATE coefficients from equation (6), allowing for event type-specific estimates. For each demand response program, the sample includes households from their own treatment program and the never-treated groups. Standard errors are reported in the parentheses and clustered at the household level. All specifications include fixed effects at the household and hour-of-sample levels. F-Stat represents the F-statistic from the first-stage IV regressions. Statistical Significance $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

# D End of Experiment Survey

## D.1 Survey Details

The following is text from a voluntary, paid survey sent out to participants in the Central, Tech, and Manual groups via email, in mid-June, 2023.

*Survey instructions*:
"This short survey is designed to hear about your experience in the Peak Rewards Trial through [APP NAME]. All homes had a different experience, and we want to hear about yours.

We appreciate the time and thought you put into this survey.

Properly completed surveys will be rewarded with $20 on bill credit as a token of our gratitude."

*Survey questions used an analysis above:*

"What is your approximate household income?

- Less than $50k per year
- $50-99k per year
- $100-149k per year
- $150-200k per year
- Over $200k per year
- Don't know/Rather not say"

"For the events you noticed, how often was it worth your time to participate by attempting to reduce your electricity consumption?"

- Never
- Sometimes
- About half the time
- Most of the time
- Always
- Don't know/Not Applicable"

### D.2 Survey Response

In this section, we compare the observable characteristics of participants who filled out the end-of-experiment survey to those who did not. The Table below recreates Table A2 from our main analysis (using pre-treatment data), but separates the results by whether or not the household responded to the exit survey. The p-value corresponds to a difference in means test.

Table C9 demonstrates that the non-respondents had larger cumulative consumption during the pre-treatment period. Non-respondents also were more likely to have an electric vehicle. All other characteristics are similar across the two groups.

Table C9. Balance by Exit Survey Response (Pre-Treatment Data)

| | Yes | No | p-value |
|---|---|---|---|
| Cumul. kWh | | | |
| Winter | 5,229 | 5,892 | 0.04 |
| | (2,810) | (3,199) | |
| Spring | 3,675 | 4,167 | 0.02 |
| | (1,788) | (2,104) | |
| Summer | 2,614 | 3,155 | 0.01 |
| | (1,492) | (2,142) | |
| Fall | 3,528 | 4,012 | 0.02 |
| | (1,721) | (2,016) | |
| Load Factor | | | |
| Winter | 24.96 | 25.21 | 0.78 |
| | (8.61) | (8.89) | |
| Spring | 19.82 | 19.98 | 0.83 |
| | (6.64) | (6.93) | |
| Summer | 16.56 | 17.21 | 0.37 |
| | (6.87) | (6.76) | |
| Fall | 18.66 | 19.12 | 0.51 |
| | (5.91) | (6.72) | |
| Electric Vehicle | 0.24 | 0.32 | 0.05 |
| | (0.42) | (0.47) | |
| BaseBoard Heating | 0.66 | 0.64 | 0.71 |
| | (0.47) | (0.48) | |
| Air Conditioning | 0.47 | 0.51 | 0.32 |
| | (0.50) | (0.50) | |
| Electric Hot Water | 0.72 | 0.70 | 0.56 |
| | (0.45) | (0.46) | |
| House Duplex | 0.83 | 0.82 | 0.71 |
| | (0.37) | (0.38) | |
| Median Income | 86,377 | 87,434 | 0.55 |
| | (19,853) | (19,836) | |
| Observations | 429 | 174 | |

Notes. This table compares pre-treatment average values by whether or not the household participated in the exit survey. 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. Median Income reports the median household-level income of the Census Dissemination Area where the household is located. ANOVA reports the p-value from one-way ANOVA tests for differences in means across programs/groups. Statistical Significance $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.