NBER WORKING PAPER SERIES

THE HUMAN PERILS OF SCALING SMART TECHNOLOGIES: EVIDENCE FROM FIELD EXPERIMENTS

Alec Brandon
Christopher M. Clapp
John A. List
Robert D. Metcalfe
Michael Price

Working Paper 30482
http://www.nber.org/papers/w30482

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2022, Revised November 2024

We thank Dylan Brewer, Fiona Burlig, Ken Gillingham, Michael Greenstone, Koichiro Ito, David Jimenez- Gomez, Juanna Joensen, David Novgorodsky, Mar Reguant, Florian Rundhammer, Casey Wichman, and seminar participants at the Advances with Field Experiments Conference, Heartland Environmental and Resource Economics Workshop, Federal Communications Commission, EPIC Faculty Workshop, NBER Environmental and Energy Economics Summer Institute, and Harvard Seminar in Environmental Economics and Policy for helpful comments and suggestions. Lillian Bartholomew, Ariel Listo, David Liu, Uditi Karna, and Jack Ogle provided excellent research assistance. We would like to thank Richard Caperton, Augustin Fonts, Ilan Frank, Kevin Hamilton, Nicholas Payton, Jim Kapsis, and many others at Opower for sharing data and offering insights. Opower provided the data analyzed in this paper to the authors under a nondisclosure agreement. The authors and Opower structured the agreement in a way that maintains the authors' independence. In particular, the agreement stipulates that Opower has the right to review the publication prior to public release solely for factual accuracy. None of the authors were paid by Opower for this research. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Human Perils of Scaling Smart Technologies: Evidence from Field Experiments
Alec Brandon, Christopher M. Clapp, John A. List, Robert D. Metcalfe, and Michael Price
NBER Working Paper No. 30482
September 2022, Revised November 2024
JEL No. D01, O10, Q4

ABSTRACT

Smart-home technologies have been heralded as an important way to increase energy conservation. While in vitro engineering estimates support optimism, little has been done to explore whether such estimates scale beyond the lab. We estimate the causal impact of smart thermostats on energy use via two novel framed field experiments in which a random subset of treated households have a smart thermostat installed in their home. Examining 18 months of associated high-frequency data on household energy consumption, yielding more than 16 million hourly electricity and daily natural gas observations, we find little evidence that smart thermostats have a statistically or economically significant effect on energy use. Using almost four million observations of system events including human interactions with their smart thermostat, we find that user behavior dampens energy savings and explains the discrepancy between estimates from engineering models, which assume a perfectly compliant subject, and actual households, who are occupied by users acting in accord with behavioral economists' conjectures. In this manner, our data document a keen threat to the scalability of new user-based technologies.

Alec Brandon
Carey School of Business
Johns Hopkins University
100 International Drive
Baltimore, MD 21202
alec.brandon@gmail.com

Christopher M. Clapp
Harris School of Public Policy
University of Chicago
1307 E. 60th St.
Chicago, IL 60637
cclapp@uchicago.edu

John A. List
Department of Economics
University of Chicago
1126 East 59th
Chicago, IL 60637
and Australian National University
and also NBER
jlist@uchicago.edu

Robert D. Metcalfe
Columbia University
and NBER
rdm2176@columbia.edu

Michael Price
Department of Economics, Finance,
and Legal Studies
The University of Alabama
250 Alston Hall
Box 870224
Tuscaloosa, AL 35487
and Australian National University
and also NBER
mkprice2@ua.edu

Innovation is the market introduction of a technical or organisational novelty, not just its invention - Joseph Schumpeter

Nearly every problem has been solved by someone, somewhere. The frustration is that we can't seem to replicate [those solutions] anywhere else. – President Bill Clinton

1 Introduction

Economists have long argued that an essential driver of economic growth is innovation (see, e.g. Romer, 1990; Aghion and Howitt, 1992). Indeed, in his seminal work, Young (1995) argues that differences in production technologies represent an important source of disparities in patterns of long-run economic growth across countries. For instance, some estimates suggest that roughly 50 percent of U.S. annual GDP growth can be attributed to innovation (U.S. Chamber of Commerce Foundation, 2012).[1] Not surprisingly, policymakers have thus focused a great deal of attention on policies designed to stimulate innovation and the supply of new technologies. Yet, as Schumpeter's quote in the epigraph suggests, innovation includes not only creation, but also the diffusion of new technologies and products in the marketplace. Schumpeter's insight has motivated various disciplines to explore the diffusion process (Skinner and Staiger, 2007). For their part, economists have explored both pecuniary and non-pecuniary aspects of technology adoption (see, e.g., the excellent survey by Hall, 2005).

Given that an important policy input is to measure the total impact of new technologies, Schumpeter's insight is unduly narrow. Beyond diffusion, there are many technologies whose impacts depend upon appropriate *use* upon adoption. In this manner, while "innovation is the market introduction of a technical or organisational novelty, not just its invention," an effective innovation should be measured by its returns at scale. A nascent literature has begun to recognize that scale underlies all social and technological progress, since deeply impactful innovations are those that reach the largest number of people and remain effective at scale (List, 2022).

As President Clinton noted in the epigraph, solutions in one setting are often frustrated when transferred to another. We denote this frustration as part of the scale-up problem (Al-Ubaydli, List and Suskind, 2017; Al-Ubaydli et al., 2017; Muralidharan and Niehaus, 2017), which revolves around several important questions such as do research findings persist in larger markets and broader settings? When we scale the intervention to these populations, should we expect the same level of efficacy that we observed in the small-scale setting? If not, then what are the important threats to scalability? Without a proper understanding of these, and related questions, the scale-up problem can lead to a vast waste of resources, a missed opportunity to improve people's lives, and a diminution in the public's trust in the scientific

[1] Technology and technological progress are also central to climate policy and the formation of international environmental agreements (Barrett, 2006; Hoel and De Zeeuw, 2010; Acemoglu et al., 2012; Harstad, 2012; Acemoglu et al., 2016; Battaglini and Harstad, 2016; Goeschl and Perino, 2017; Harstad, 2020).

method's ability to contribute to policymaking (Al-Ubaydli, List and Suskind, 2020).

In this study, we explore the scale-up problem for an important class of new technologies in the energy space that leverage "smart" functionalities. Partnering with Opower and Honeywell in conjunction with Pacific Gas and Electric (PG&E) – the second largest residential energy provider in the United States – our goal is to explore the effect that smart thermostats have on home energy usage. To do so, we examine data from two framed field experiments, wherein the 1,385 households that volunteered to participate in the study were randomized into either a treatment group that received free installation of a Honeywell two-way programmable smart thermostat or a control group that did not receive such a smart device and kept their existing thermostat.[2] We evaluate the effect of the smart thermostat on subsequent energy consumption using high-frequency data over an 18-month period that includes more than 16 million hourly electricity use records and almost 700 thousand daily observations of natural gas consumption.

Non-experimental methods predict substantial energy savings from the adoption of smart thermostats. For instance, the ecobee (2019) website touts savings of "up to 23%" on heating and cooling costs. The Nest (2019) website advertises a 10 to 12% savings on heating and a 15% savings on cooling costs. These claims inflate savings by using heating- and cooling-specific energy use as the denominator and are agnostic to the local climate. However, even more pertinent engineering estimates from the California Technical Forum also predict that smart thermostats will produce substantial reductions in energy consumption. The most relevant estimates to our experimental sample come from Department of Energy (DOE) Technical Reference Manuals (TRM), which are annual reports produced by energy providers and regulators (DOE, 2017).[3] These reports primarily rely on engineering simulations and survey data to predict the effects of energy efficiency programs at scale. These predictions are then used by energy providers to justify expenditures on energy efficiency programs. Mapping these predictions for Californians, which vary by climate zone and the size of a home, to our experimental samples we find that savings of 1.3% and 4.0% are respectively predicted for overall electricity and natural gas consumption (California Municipal Utilities Association, 2017).

Our experimental estimates provide several insights into whether the *petri dish* estimates of engineers hold when technology is scaled beyond the lab. First, we find that smart ther-

[2]In addition to the ability to schedule permanent temperature setpoints and interact with the thermostat remotely, the smart thermostat given to households in our experiment provided households with a social norm framing of their setpoint choices. Framing of setpoints is an increasingly common feature of more modern smart thermostats, and there is an extensive literature documenting the responsiveness of household energy consumption to social norm framing (e.g., Allcott, 2011; Ferraro and Price, 2013; Ayres, Raseman and Shih, 2012; Costa and Kahn, 2013; Allcott and Rogers, 2014; Dolan and Metcalfe, 2015). Given this finding and the Peffer et al. (2013) result that most individuals do not use the programmable features of their thermostats as intended, this feature should provide the best chance for the smart thermostats used in our experiment to cause a reduction in energy consumption.

[3]A related set of econometric estimates can be found in white papers produced by utility-commissioned consultants. Both these and the engineering approaches have known issues. See Allcott and Greenstone (2012) for a general discussion, and Section 2 for a review of the smart thermostat-specific literature.

mostats fail to deliver the expected energy savings; our results show that such technologies have neither a statistically nor economically significant effect on energy use. For example, using a specification that includes household and time effects, as well as controls for weather, our point estimates suggest that smart thermostats only decrease electricity consumption by 0.09% and *increase* gas consumption by 1.70%. The failure of engineering estimates to accurately predict measured responses is broadly consistent with a growing body of research that documents real-world effects of energy efficient technology that pale in comparison to the effects predicted by engineers (Davis, Fuchs and Gertler, 2014; Levinson, 2016; Zivin and Novan, 2016; Houde and Aldy, 2017; Fowlie, Greenstone and Wolfram, 2018; Alpízar, Bernedo and Ferraro, 2019; Davis, Martinez and Taboada, 2020; Christensen et al., 2021).

Second, to investigate whether this aggregate result masks significant, but offsetting, heterogeneous effects that may have implications for how the intervention scales to different settings, we estimate the model across different subsamples such as day of the week, hour of the day, by ambient temperature/humidity quintiles, and when there is a peak-load alert. We find almost no evidence of heterogeneous treatment effects.[4] The overall pattern across all our results consistently indicates that smart thermostats under-deliver on the savings promised by engineers.

Third, we explore mechanisms that may explain the drop in smart thermostat effectiveness when moving from engineering studies to our field experiments. Using almost four million observations of treatment group heating, ventilation, and air conditioning (HVAC) system activity, three insights emerge. First, while the average scheduled setpoints are consistent with DOE energy efficiency recommendations, there is substantial heterogeneity. Second, households frequently override their setpoints, with an average of nearly 1.7 overrides per day. Third, overrides are typically less energy efficient than the previously scheduled setpoint. Finally, combining the setpoint and override data with energy consumption data, we find that our experimental data can reproduce engineering predictions when the control group is compared to a treated group comprised of households that choose efficient setpoints. Specifically, we define the efficiency type of a household based on its relative position in the distributions of permanent setpoints and temporary overrides (e.g., high types are those above the median number of programmed setpoints and low types are those below).[5] Collectively, these results suggest that engineering predictions fail because they assume unrealistic levels of compliance with the intended use of smart technologies. That is, the households who adopt the smart technology use its features in ways that undo the purported benefits, suggesting that human behavior is a peril to scaling such technologies.

We view our results as speaking to several literatures and stakeholders. For example, for policymakers, empiricists, and theorists interested in scaling insights from the small to the large,

[4]Similar to other studies in the weather-energy use literature (Deschênes and Greenstone, 2011; Auffhammer and Mansur, 2014), we find that temperature has a U-shaped impact on energy consumption, but that smart thermostats do not attenuate this relationship. As climate change will result in more extreme temperature days, smart thermostats are not a panacea for climate change mitigation.

[5]As these are post-treatment measures of adherence to treatment, we caution against interpreting this analysis as causal. Rather, it is designed to replicate the assumptions implicit in engineering estimates.

we present a novel case study that holds import for the recent evidence-based policy movement. Over the past several decades, empirical methods have evolved to be a key contributor to the scientific knowledge base from which policymakers draw insights. Indeed, in most governmental circles, evidence-based programs were once an aspirational goal and now they are the expectation (Abraham et al., 2017). Yet, whether, and to what extent, insights from any research study scale to the level of the broader public is, in many situations, based on blind faith. A recent literature has emerged in economics that explores the economic underpinnings of the scale-up problem (see, e.g., Al-Ubaydli, List and Suskind, 2020). Our results pinpoint a key feature for user-based technologies; humans may not use the technology as envisioned and assumed by engineers.[6]

In this sense, our framed field experiments and analysis of the underlying mechanisms that drive our results provide fresh insights into a key issue that medical practitioners have grappled with for centuries: patient non-adherence to prescribed medications. This problem has spawned a large literature in the medical sciences regarding the best practices for improving medication adherence, and the results are directly relevant to economists seeking to tackle the key component of the scaling problem that arises in the efficacy of new technologies. While distinct from the features that implementation scientists tend to focus on – how lack of fidelity causes treatment effect sizes observed in research studies to diminish substantially when the program is rolled out at larger scale (see, e.g., Kilbourne et al., 2007; Weiss, Bloom and Brock, 2014; Supplee and Meyer, 2015; Supplee and Metz, 2015; Gottfredson et al., 2015; Cheng et al., 2017; and in economics, Al-Ubaydli et al., 2017; Brandon, 2019) – we view our work as complementing this literature in that it highlights the multi-dimensional nature of the scale-up problem.

The non-adherence we highlight as the key hindrance to scaling engineering estimates occurs because human behavior is not appropriately accounted for in those models. In their most naive form, engineering studies compare the energy use of an HVAC system simulated under two different scenarios: a smart thermostat optimally programmed for energy savings and a traditional thermostat set to maintain a fixed temperature (Urban, Elliott and Sachs, 2012; Urban and Gomez, 2013; Daken, Meier and Frazee, 2016). Both the experimental and baseline scenarios are unrealistic as they treat users as automatons and thus ignore how people actually use their thermostats.[7] As such, these studies estimate the upper bound on true energy savings. Thus, it is not surprising that device producers often justify their energy-saving claims based on the results of engineering studies. In contrast, our study is based on a field experiment that captures how individuals actually use both smart and traditional thermostats and allows us to estimate real-world savings as opposed to a hypothetical upper-bound.

[6] We are more concerned with "vertical" scaling (e.g., moving from the lab to the field) as opposed to "horizontal" scaling (e.g., moving from one place or sample or another) (List, 2022). Allcott (2015) is an example of the latter, but to the best of our knowledge, we are the first to examine the former in the smart technology space.

[7] For instance, the ecobee (2019) website makes the aforementioned claim of 23% in HVAC energy savings from its smart thermostat relative to a constant temperature setting. This methodology is akin to implicitly assuming ideal energy-conservation behavior in the treatment group and no optimizing behavior in the control group.

Moreover, issues with non-adherence are likely to be compounded at larger scales than our framed field experiments. Even though our sample is externally valid with respect to base energy use, it is comprised of those who expressed interest in a smart thermostat.[8] If selection into our experiments is driven by anticipated energy savings (i.e., gains) then scaling adoption of the smart thermostat would be less likely to yield the savings predicted by engineers and policymakers. This, of course, is subject to future research because it assumes that such households override setpoints more than households that are interested in the technology.

More generally, while null results like ours have posed a challenge for researchers in terms of their informativeness (Abadie, 2020), we show that people's interaction with the smart technology is the reason why we observe the null effect and the resulting departure from engineering model predictions. While our findings are "statistical" nulls, they are not "policy" nulls because they are counter to widely-held prior beliefs based on engineering estimates, and these estimates are driving policymakers and energy producers alike to subsidize smart technologies based on misleading information. For example, according to the Environmental Protection Agency (EPA), 170 energy providers subsidize the purchase of smart thermostats (EPA, 2019). In 20 states, over half of all households are eligible for a smart thermostat rebate (Bloomberg Finance L.P., 2019).[9] These subsidies are justified both by the aforementioned TRMs and the joint EPA and DOE ENERGY STAR program. This program grants certain types of energy efficient technologies with a ENERGY STAR certification (Houde and Aldy, 2017). Energy providers then subsidize the purchase of certified products with funds that would be better spent on more effective interventions.

Regarding energy efficiency policies more broadly, our results also speak to the literature on the potential benefits of smart grid investments (Joskow, 2012). Between 2009 and 2014, the DOE invested $7.9 billion in smart technologies under the Smart Grid Investment Grant (SGIG) program by providing matching grants to competitively chosen projects (DOE, 2016). Much like the smart thermostats we study, many of the grant funds were allocated to projects that were justified on the basis of engineering estimates and targeted smart technologies that rely on households conforming to the behavioral expectations of engineers.[10] Our analysis of the mechanisms underlying the effects of smart thermostats highlights one reason why investments in these related smart grid technologies may fail to scale if they are applied more broadly: because engineering models do not properly account for how individuals actually use them.

The remainder of this study is organized as follows. In Section 2 we describe the details of the

[8]Based on data from the Energy Information Administration's (EIA) Residential Energy Consumption Survey (RECS) for the year 2009, Californians eligible for our thermostats experiment used 1.2 kWh per hour. This number is extremely similar to our sample, which ranges between 1.0 and 1.3. kWh per hour.

[9]In the most generous case, all of the residents in Nevada are eligible to receive a smart thermostat for free.

[10]A non-trivial fraction of the projects targeted the development and dissemination of technologies such as smart thermostats that allow individuals to remotely communicate with their appliances. At the same time, more than two-thirds of the grants went towards other projects such as outfitting households with complimentary technologies that include smart meters and systems that allow utilities to better monitor and communicate grid conditions to customers with the goal of influencing their consumption decisions (e.g., via demand response messaging).

field experiment, the sample of households in the study, and our data. The following section formalizes our empirical specification. Section 4 presents our model estimates, and Section 5 explores the mechanisms that drive our findings. The final section concludes.

2 Field Experimental Design

Before discussing our experimental design, it is worthwhile to briefly summarize the current literature pertaining to energy savings estimates. The existing econometric literature primarily consists of white papers that thermostat producers use to claim energy savings of 10% or more based on a combination of observational and experimental data (Apex Analytics, LLC, 2014, 2016; Aarish et al., 2015; Ho, 2014; Kelsven, Weber and Urbatsch, 2016; Nest Labs, 2014, 2015; Schellenberg, Lemarchand and Wein, 2017; Stewart and Jackson, 2015; Robinson et al., 2016; Ward, Stewart and Jackson, 2014). Importantly, to our best knowledge, few, if any, of these have been subject to peer review. And, to varying degrees, all are unclear about salient features of the study, have methodological flaws (primarily related to selection), and/or draw incorrect conclusions from their estimates. These issues are likely to lead to upwardly biased estimates of savings, thus it is not surprising that device manufacturers are eager to advertise the results of these studies.[11] Such studies have likely influenced energy producer and federal policies on smart thermostats.

Exceptions in terms of both clarity and quality are the white papers due to Broaddus, Ryan and Marrin (2016, 2018) and Park et al. (2017).[12] However, in both cases, the observed outcome is based on aggregate energy consumption data: Broaddus, Ryan and Marrin (2016, 2018) observe monthly energy billing data and Park et al. (2017) observe weekly smart meter data. Following Agnew and Goldberg (2013), both studies include coarse measures to control for ambient weather conditions: counts of heating and/or cooling degree days. In contrast to the existing econometric literature, we use high-frequency energy use data to estimate a difference-in-differences instrumental variables (DDIV) model. Ghanem and Smith (2021) formalize the benefits of using high-frequency hourly data over a more aggregate analog. They show that fixed effects estimators based on aggregate data are inconsistent when there is high-frequency temporal heterogeneity in the effects. While their focus is on high-frequency

[11] The aforementioned Nest (2019) website claims of 10 to 15% in savings is based on an internal study. Nest Labs (2015) reports estimates from a difference-in-differences (DD) regression model that compares the monthly energy use of a self-selected group of households that were early adopters of the Nest smart thermostat and enrolled in an energy-monitoring program to those who only enrolled in the monitoring program. The study's authors acknowledge potential sources of bias in their estimates, but fail to provide evidence that the change in the energy use of their comparison group is a reasonable counterfactual for that of those who decide to install a Nest.

[12] These studies acknowledge self-selection in the treatment group and estimate ITT models on all those encouraged to install a smart thermostat in their experiment. The latter uses four different methodologies to estimate the effect of a smart thermostat on energy use, including a small-scale field experiment that uses a matched-pair randomization design to address selection after randomization. While significant, we note that estimated savings effects in these studies are generally smaller than in the previously cited studies and in-line with the predictions from the TRMs, on the order of 1% to 6%.

heterogeneity in treatment effects, the same concern extends to potential confounders. Accounting for this variation is particularly important in our context because smart thermostats are designed to allow individuals to vary energy use in response to within-day changes in temperatures. Thus, in addition to our model specification addressing the selection issues that bias much of the existing literature towards findings of significant savings, our high-frequency data allows us to better control for differences in ambient weather conditions and more accurately estimate our coefficient of interest than existing studies.[13] With these advantages in mind, we turn to our data generation procedure.

2.1 Smart Thermostat

The intervention in our framed field experiments occurs when a given household's existing thermostat is replaced by a smart device.[14] Smart thermostats are designed to increase consumer utility by improving the efficiency of the home's HVAC system and reducing adjustment costs. To these ends, the device in our experiment has two primary features common to most smart thermostats. First, the thermostat allows the user to program an extensive schedule of permanent temperature setpoints for each day of the week. Second, the user can either interact with the device directly or remotely via a web portal or smartphone app. Both lower the cost of adjusting temperature settings.[15]

While the effect of these features on energy usage is theoretically ambiguous depending the schedule the user sets and how she interacts with the device, there are several additional features of the thermostat used in our experiment that are designed to reduce energy consumption compared to a traditional thermostat. First, our smart thermostat is able to learn about how HVAC system operations affect indoor temperatures, then optimize the transition between programmed temperature setpoints. Second, when choosing setpoints, users receive messages that compare their settings to those of similar households. Analogous to the social comparison module studied in Allcott (2011), the thermostat interface presents: (i) descriptive norms with information on peer setpoint choices and (ii) injunctive norms with efficiency ratings of setpoints. Third, the thermostat app interface is designed to facilitate toggling to a less energy intensive setting when the user leaves home and toggling it back to the previous setting when the user returns. Finally, when a user overrides a permanent setpoint to make a temporary change that is more energy efficient than the scheduled one, she is prompted by a

[13] Novan, Smith and Zhou (2022) use similar high-frequency smart meter data to reexamine the effect of building codes on energy use. In contrast to the existing literature (Levinson, 2016; Kotchen, 2017), which analyzes lower-frequency data, the authors find that residential energy efficiency standards reduce electricity consumption.

[14] Specifically, surrogates of Opower/Honeywell installed a Honeywell Z-Wave Touchscreen Thermostat that communicates with a website portal and smartphone app designed and hosted by Opower. We do not observe anything about the pre-existing thermostat.

[15] Appendix Section A provides a more detailed description of the device. Panel (a) of Appendix Figure 12 displays the thermostat and associated applications. Panel (b) shows a screen-shot of scheduling using the smartphone app.

query asking if she wants to make this more energy efficient setting permanent.[16]

Some newer smart thermostats have additional energy saving features.[17] While we cannot say whether our experiment tests the efficacy of all smart thermostats or the combination of features in the experimental thermostat, we note that the thermostat in our experiment has all the core features of current smart thermostat models. Additionally, the analysis in Section 5 indicates that individuals make use of these features and do so largely as intended. Taken together, this suggests that our results are unlikely to be specific to the particular device installed as part of the experiment.

2.2 The Framed Field Experiments

Subjects were recruited in public places (e.g., malls, markets, and festivals) in two waves (or experiments) following the spirit of a framed field experiment (Harrison and List, 2004).[18] Recruitment for the first field experiment took place across four counties in Northern California from July through October of 2012. Subjects in the second experiment were recruited from December of 2012 to February of 2013 in three Central California counties.[19] Appendix Figure 15 depicts the locations of homes in the experiments and provides visual evidence that treatment and control groups are spatially balanced across locations.[20]

Figure 1 illustrates the execution of the field experiments. It describes the assignment of households to treatment and control groups, as well as the subsequent installation decisions of treatment households. A total of 1,379 eligible households agreed to participate in our experiments: 815 as part of the Northern California experiment and 564 in the Central California experiment.[21] They were randomized into either a treatment or control group. After group assignment, the experimenter had no further contact with the total of 690 control households across both experiments. The 689 total households assigned to the treatment

[16]Appendix Figure 13 highlights features of the smart thermostat. Panel (a) illustrates the social norm framing displayed when households choose setpoints. Panel (b) shows how households can remotely toggle the thermostat in response to leaving and returning home via a smartphone or personal computer.

[17]For instance, Daken, Meier and Frazee (2016) explain that some smart thermostats use the location of user cell phones to automatically adjust settings when users are away from home and/or optimize HVAC system settings in response to local weather conditions.

[18]To be eligible, an individual had to own her residence and have central air conditioning, a smart phone, and high-speed Internet. See Appendix Section I.1 for a summary of the eligibility requirements. For more information on canvassing, see Appendix Section I.2 for the original recruitment and enrollment guide.

[19]Subjects for the experiment in Northern California were recruited from the greater San Francisco/Sacramento area (Contra Costa, San Joaquin, Solano, and Yolo counties). Households in the Central California experiment are located in and around Fresno and Bakersfield (Fresno, Kern, and Madera counties).

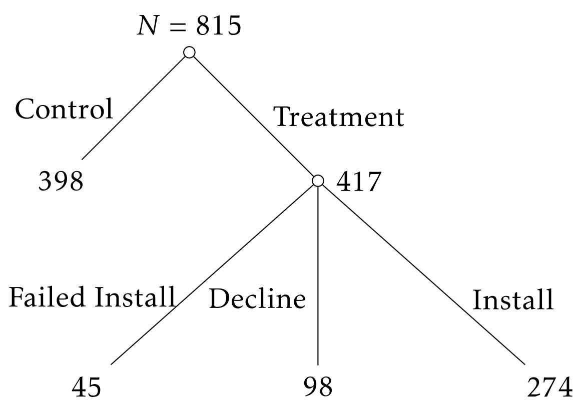
[20]We formally test balance in Section 2.6 and fail to reject the null of spatial balance in the counties where households are located.

[21]All household counts in this section are based on the households for which we observe electricity consumption. Aggregating across both experiments, there are a total of 1,379 unique households in the electricity sample, a total of 1,369 unique households in the natural gas samples, and a total of 1,385 unique households across both energy-type samples. Stated another way, we observe 16 households with electricity consumption data, but not natural gas information and another six households that consume natural gas, but for which we have no electricity consumption information.

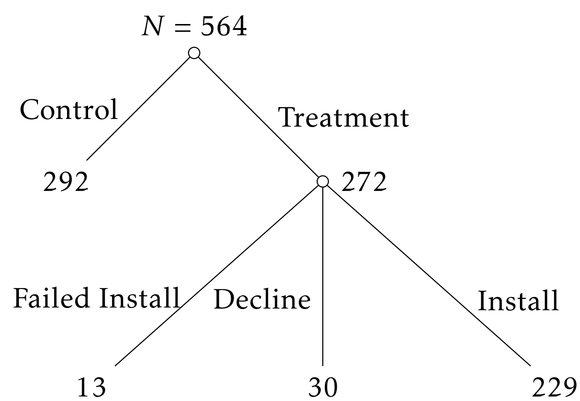
groups were offered the smart thermostat described in the previous section and installation at no cost.

Professional installation of the smart technology is an important feature of our experiments over the encouragement or self-installation designs common to other experiments. Peffer et al. (2011, 2013) provide evidence that programmable thermostats are often installed incorrectly and list flawed installation as a reason they are not more effective. Additionally, Apex Analytics, LLC (2016) find that although cheaper, their self-installation design "led to substantial attrition among interested and qualified customers." In contrast to the 35% take-up rate in their experiment, on average across our experiments, the smart thermostat was successfully installed in 73% of treatment group homes. Of the remaining treatment homes, 19% percent declined, and 8% had complications that prevented installation (e.g., compatibility issues).[22]

Figure 1: Randomization of Sample



(a) Northern CA Experiment



(b) Central CA Experiment

Note: This figure presents the number of households in treatment and control for each experiment. Treatment group installations of the smart thermostat, failed installations, and declined installations are also reported. Panel (a) reports the sample sizes for the Northern California experiment and panel (b) reports the sample sizes for the Central California experiment.

2.3 Energy Data

All households in the study were equipped with smart meters that enabled PG&E to record household-level data on hourly electricity use and daily natural gas consumption. The quantity of electricity consumed is measured in kilowatt hours (kWh), and the unit of measure-

[22] Appendix Figure 14 plots the cumulative density function (CDF) of the difference in time between assignment and installation dates that illustrates how long it takes households in the treated groups to install the smart thermostat (conditional on eventual installation of the smart thermostat). Most households had the smart thermostat installed shortly after being assigned to the treatment group: 50% of households had their thermostat installed within 5 days, and 95% had it installed within 30 days.

ment for natural gas is a therm (thm).[23] As we cannot observe temperature setpoints directly for control households with a traditional thermostat, and energy is the policy-relevant good, these measures are the main outcome variables in our analyses. In total, we observe an average of 11,908 hourly electricity use decisions for the 1,379 households in electricity sample and 495 natural gas use decisions for the 1,369 households in the natural gas sample over an 18 month period from July 2012 through December 2013.

2.4 Timing

Figure 2 presents two visual depictions of important timing issues associated with the experiments and data. Panel (a) of Figure 2 plots the flow of households into treatment and control groups over time. The horizontal axis spans the period of time over which we observe energy data. The grey shaded areas illustrate the periods of subject recruitment in each of the two experiments. The subfigure shows that treatment and control households are temporally balanced, as they were assigned at similar rates over time, and that there is very little attrition over the year and a half study period.[24]

Unfortunately, we only observe energy readings starting on the first day of recruitment in Northern California experiment. Panel (b) illustrates the effect of this issue by plotting the number of electricity readings per day for each experiment relative to event time (where assignment to the treatment or control group occurs at time zero).[25] The figure shows that we do not observe a substantial pre-period for all households in the Northern California experiment, but we do for the Central California experiment. We report estimates in Section 4 both separately by experiment and based on a sample that combines data from both experiments to account for this issue. Estimates are not qualitatively different across specifications.

2.5 Additional Data

2.5.1 External Data

We supplement the main experimental dataset with information from several external sources and additional data collected as part of the experiment. First, we compile hourly temperature, humidity, and heat index readings for each county in the study from the National Oceanic and Atmospheric Administration (NOAA).[26] Appendix Table 3 summarizes the weather data.

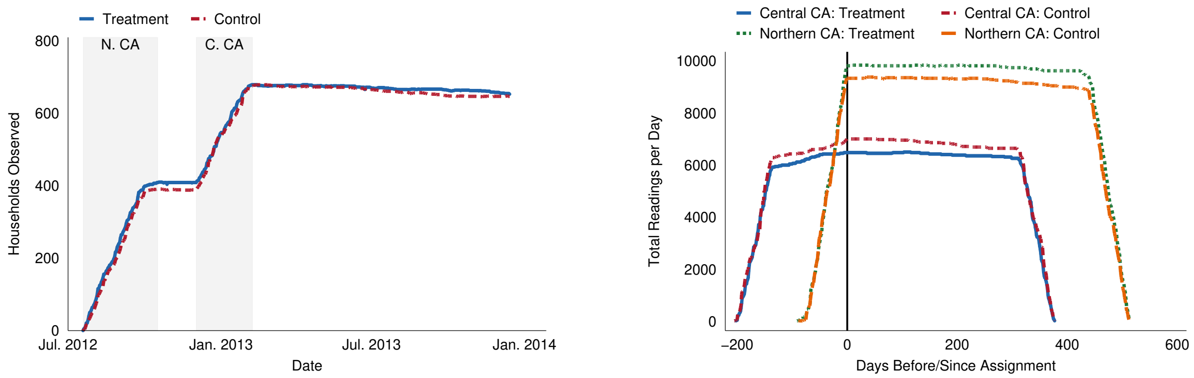
[23] A therm is a unit of heat energy equivalent to 100,000 BTUs.

[24] We formally test balance in Section 2.6 and fail to reject the null of temporal balance in the month of assignment to experimental group.

[25] Plotting an analogous graph for natural gas readings changes the scale of the vertical axis but produces the same overall pattern.

[26] We are missing values for 0.09% of the temperature and 0.5% of the humidity observations in the sample. We interpolate these missing values using the predicted values from separate regressions of the given weather variable on location, day, and hour fixed effects. We calculate the heat index from the temperature and humidity readings (see: https://www.wpc.ncep.noaa.gov/html/heatindex_equation.shtml for the formula).

Figure 2: Timing of Recruitment and Observation



(a) Number of Households Observed by Experimental Status and Date

(b) Total Electricity Readings by Experiment, Experimental Status, and Event Time

Note: This figure reports the timing of recruitment and observation in the two experiments. Panel (a) reports the flow of treatment and control households in and out of the experiments. Panel (b) reports the number of electricity readings per day relative to a household's assignment to treatment or control in the experiments.

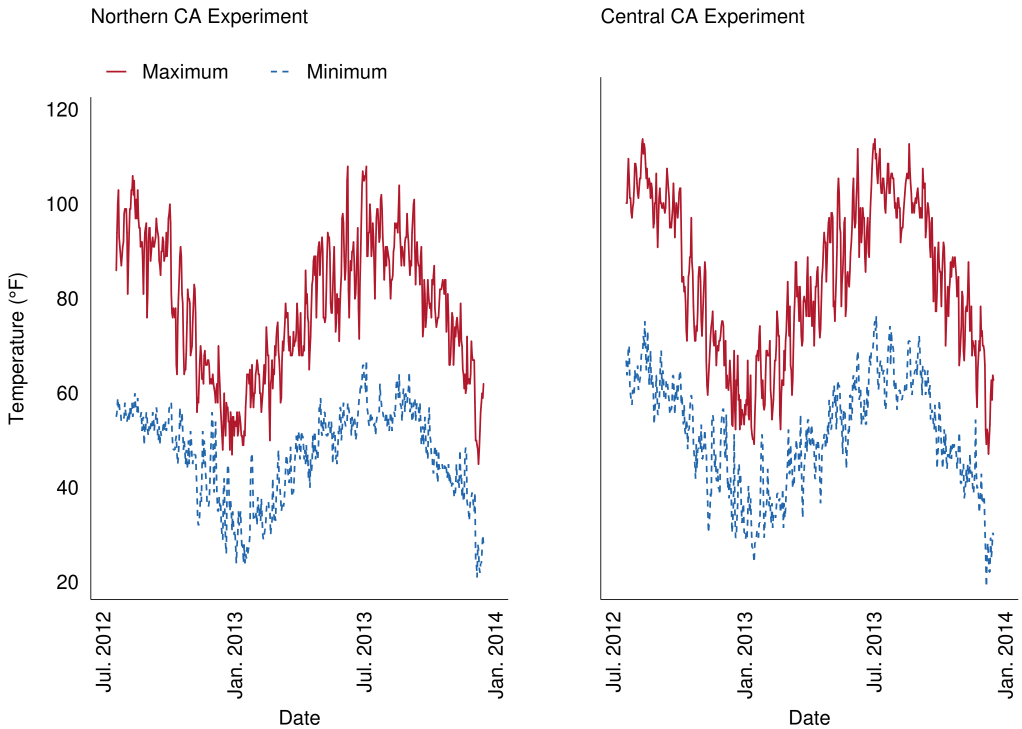
Temperatures in the combined sample (Panel C) average 63.7 degrees Fahrenheit (F), but range from below freezing to well over 100 degrees F. The various *Daily Measure* statistics for each of the three weather measures indicate that there is both spatial (between-county) and seasonal (within-county) variation in the data. The *Minimum* and *Maximum* statistics indicate that there is also daily variation in all three weather variables. Figure 3 visualizes this variation by plotting experiment-specific time series of the minimum and maximum daily outdoor temperatures over the sample period. The table and figures indicate that despite our sample being drawn from a temperate part of the country, there is substantial variation in the weather data. Summers are hot, humid and likely to require the use of air conditioning to ensure comfortable indoor temperatures. While the rest of the year is more moderate, there are many days cold enough to necessitate home heating.

To confirm that this is the case and that we are able to identify the effects of HVAC system use in our smart meter data, Figure 4 plots the relationship between mean daily energy consumption and mean daily temperature for homes in the control group.[27] The blue markers represent electricity use (the energy source used for cooling; denoted on the left-hand vertical axis), and the red markers represent natural gas consumption (the predominate energy source for heating; denoted on the right vertical axis).[28] As one would expect, electricity use increases, and natural gas use decreases, with the temperature. Both relationships are non-linear, and the fitted-value lines indicate that quadratic models predict the data well.

[27] Analogous scatter plots based on treated households produce the same patterns. Additionally, Ge and Ho (2019) analyze high frequency, smart thermostat event log data (similar to the data we analyze in Section 5) and find that the home heating and cooling decisions of smart thermostat users are affected by weather conditions.

[28] The area of both markers are weighted by the number of observations in the given cell.

Figure 3: Minimum and Maximum Daily Outdoor Temperatures (°F) by Date



Note: These figures display average daily minimum and maximum temperature in Fahrenheit with temperature readings matched to the samples in each experiment.

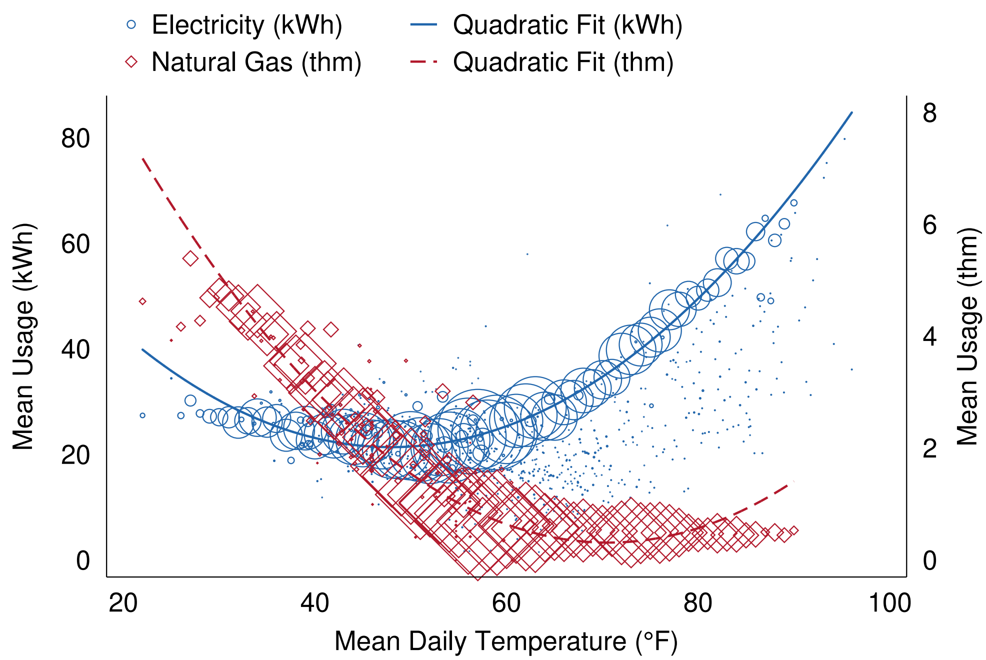
These descriptive analyses indicate that there is sufficient variation in weather conditions in our sample and energy use responds to that variation, so our experimental setting meets the necessary conditions for assessing the efficacy of smart thermostats. They also inform our model specification. We estimate separate models of the effects of smart thermostats on electricity and natural gas use. For robustness, we include outdoor temperature and humidity measures, as well as location and time effects, as controls to mitigate the effects of residual variation on our estimates.

Second, we supplement the household electricity use measure with data from two sources that allow us to test whether smart thermostats have a differential effect on usage when there is critical demand load. To do so, we collect data on the average hourly real-time price of electricity from the California Independent System Operator (CAISO).[29] Electricity is produced from many sources with different production and external costs.[30] Appendix Figure 16 is a box and whisker plot of hourly spot prices by quintile that illustrates the variation in these costs in our data. Spot prices are relatively consistent over the first four quintiles, but

[29]The real-time market for electricity in California clears every five minutes. We use this data to calculate the average spot price each hour.

[30]California instituted a cap-and-trade carbon emissions program in 2012 (Shobe, Holt and Huetteman, 2014), so the price of electricity on the state's wholesale market reflects both the marginal cost of production and the prevailing market price for emissions as reflected in the price of carbon permits.

Figure 4: Average Daily Energy Use by Outdoor Temperature (°F)



Note: This figure plots mean usage of electricity and natural gas as a function of mean daily temperature. A quadratic fit is plotted for each outcome and larger points indicate more observations for that mean daily temperature.

increase substantially from the fourth to the fifth quintile. This is consistent with what we would expect during peak-load times, but the long whisker in the fifth quintile suggests that peak-demand times may comprise only a small fraction of the observations in our dataset.

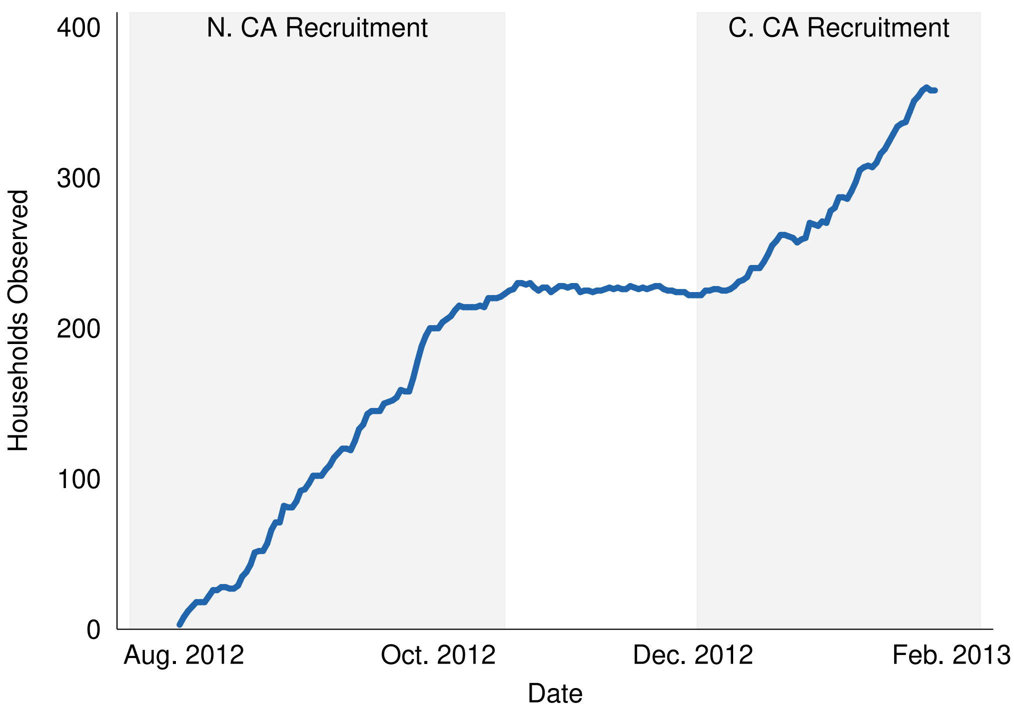
To further identify times when the system is most strained, we also collect data on system-wide peak-alert messages from CAISO and utility-wide alerts from PG&E. The latter alerts (referred to as "SmartDays" by the utility) are issued at a finer spatial scale, but a more granular temporal level (daily) than the former. In contrast, the CAISO alerts are issued hourly, but apply to a broader area. Since system-wide alerts may occur on days when it is less obvious that there is a need to reduce demand to avoid brownouts based on local conditions, we also identify CAISO alerts that were broadcast by local media outlets (e.g., the *Fresno Bee* or the *Bakersfield Californian*) to ensure that they reach a reasonable level of publicity to be salient to the households in our experiments. Conditioning on the additional information from these sources allows us to test whether smart thermostats reduce demand when the cost of electricity production to society is the greatest.

2.5.2 Internal Data

In addition to the external data we collect, we also observe a high-frequency, exact-time log of 3,967,558 HVAC system events, including user interactions with their smart thermostat, from

372 households. The unbalanced panel dataset spans from July 2012 to January 2013, and Figure 5 illustrates the number of households observed by calendar date. Recruitment and installation of smart thermostats first began in Northern California in July of 2012, whereas those in Central California began in December of 2012. Since this dataset is truncated in January of 2013, the majority of the observations in this dataset are generated by homes from Northern California, while only about 5% of the observations are from Central California homes.

Figure 5: Number of Households Observed in Events Data by Date



Note: This figure plots the number of households with observed HVAC system events by date. We observe HVAC system events over a shorter time-period than our experimental data: from July 2012 to January 2013.

The system events and user interactions we observe include ambient temperature, HVAC state, and heating/cooling setpoints (which we classify into permanent setpoints and temporary overrides).[31] Permanent setpoints are thermostat temperature settings previously scheduled to occur automatically at specific times on a periodic basis. Temporary overrides are changes to the current setpoint which result from a concurrent interaction with the thermostat.[32] We aggregate these measures to hour-level observations. Appendix Table 4 summarizes the data. The table shows that while there are more observations from the Northern

[31]Unfortunately, we do not observe who or how many people in the household have access to the app and/or interact with the thermostat.

[32]We do not directly observe whether system temperature changes are due to permanent setpoints or temporary overrides, but we are able to infer event types based on the precise timing of when the changes occur. See Appendix Section D for details.

California experiment, settings in the two locations are remarkably similar.[33]

Finally, Opower and Honeywell conducted an online survey to collect baseline information on both treatment and control households in the experiments. We do not use these time-invariant household characteristics in our main analysis because they are redundant to household fixed effects, but we use them to test the validity of Opower and Honeywell's randomization process.

2.6 Balance

To test for balance, we estimate a linear probability model with an indicator for assignment to treatment as the dependent variable. Appendix Table 5 reports estimates from that model that summarize the results of our balance tests. Column (1) reports estimates based on a sample comprised of households from both experiments, and the estimates in Columns (2) and (3) are from models estimated on subsamples by experiment. The significance of each coefficient estimate represents the results of a single hypothesis test against a null of balance, and the reported F-statistics test the null hypothesis that all parameters in the given model are jointly equal to zero. We fail to reject the null for all single and multiple hypothesis tests across all three models. This indicates that control and treatment households are statistically balanced across observable, pre-experiment measures and is consistent with an appropriate randomization process.

We note that that households in the treatment group in the Northern California experiment used 5.5% less electricity per hour in the pre-period on average than those in the control. Accounting for means that are based on less than two weeks of data in Appendix Table 5 indicates that this difference is driven by the subset of households for which we observe only a limited number of pre-period electricity observations (see Section 2.4). Regardless, out of an abundance of caution, we estimate double-difference models to account for any potential pre-period imbalance.

2.7 Time-Trend and Event-Study Analyses

To illustrate basic temporal patterns in the data and the effect of experimental assignment on energy use Appendix Figure 20 plots the mean of residual energy consumption against event time (days before/after assignment to the treatment or control group) for each of the two experiments.[34] Panel (a) displays electricity use, and Panel (b) illustrates the patterns in natural gas consumption. The figure shows that being assigned to receive free installation of

[33] Average ambient temperatures are higher in Northern than Central California because of seasonal variation. The Northern California panel spans July through January, whereas the Central California panel runs from December through January.

[34] We project out household fixed effects prior to taking daily averages to adjust for pre-period differences in electricity use in a subset of Northern California homes for which we observe only a limited number of pre-period energy values.

a smart thermostat has no discernible impact on subsequent patterns of use. However, the raw data is too noisy to be visually conclusive.[35]

To provide further evidence of the validity of the experimental randomization and additional evidence of parallel pre-trends, Figure 6 plots the coefficient estimates and 95% confidence intervals from event studies of the effect of assignment to treatment.[36] Panels (a) and (b) plot electricity and natural gas estimates, respectively, based on data from the Northern California experiment. Panels (c) and (d) plot the Central California experiment analogs. Consistent with Appendix Figure 20, the event study plots show evidence of parallel pre-trends, but do not indicate large, persistent effects of being assigned to treatment on energy use.[37]

We note that these figures do not account for incomplete take-up of the treatment, and they are based on temporally aggregated, day-level data. For these reasons, in the next section, we outline empirical models that allow us to instrument for smart thermostat installation and take advantage of the high-frequency nature of the electricity consumption data.

3 Empirical Model

Our field experiment randomizes receipt of a smart thermostat among eligible applicants. We observe a long time series of household-level energy use for treatment and control groups before and after experimental assignment. Both motivate our empirical strategy. Given the potential pre-period imbalance in electricity use discussed in Section 2.6, we estimate difference-in-differences (DD) models. To address noncompliance with experimental randomization, we augment our DD model with instrumental variables (IV) modeling techniques. We begin by formalizing our model specification, then discuss identification issues.

3.1 Model Specification

We model the effect of a smart thermostat on household i's consumption of energy type $j \in$\{kWh, thm\} (electricity, natural gas) in time period t (e^j_{it}) using a DD model:

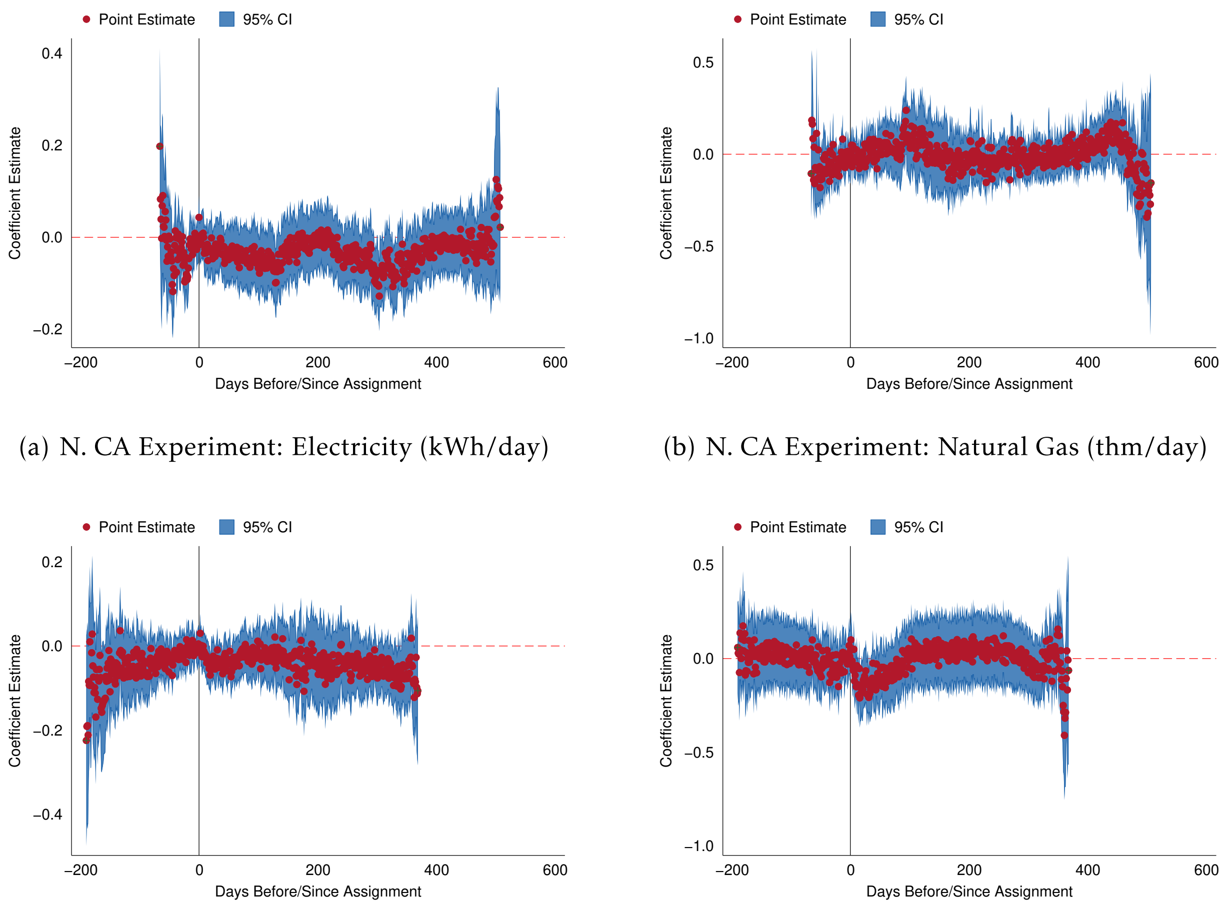
$$e^j_{it} = \alpha^j_i + \beta^j_t + \gamma^j S_i P_{it} + X_{it}\delta^j + u^j_{it}, \tag{1}$$

[35]For instance, the seasonal effects of summer for electricity use and winter for natural gas can be seen in the patterns in the data.

[36]The models are estimated on daily energy use data and include both household fixed effects and daily time effects. Confidence intervals are based on standard errors clustered by household. To mitigate the visual effects of noisy coefficient estimates resulting from unbalanced lags and leads at the endpoints of the time window, we bin all lags and leads that are based on fewer than 30 observations (Schmidheiny and Siegloch, 2019; Clarke and Schythe, 2020).

[37]We note that the parallel trends assumption required for identification in the empirical model we define in Section 3 is satisfied by the experimental randomization of our smart thermostat treatment. We use a DDIV specification because the high-frequency nature of our data affords controlling for pre-trends. Doing so reduces residual variation, improves precision, and helps address concerns about a noisily estimated null.

Figure 6: Event Study Estimates of Energy Use by Experiment



(a) N. CA Experiment: Electricity (kWh/day)

(b) N. CA Experiment: Natural Gas (thm/day)

(c) C. CA Experiment: Electricity (kWh/day)

(d) C. CA Experiment: Natural Gas (thm/day)

Note: These figures plot the daily effect of assignment to the treatment group on electricity (kWh/day) and natural gas (thm/day) consumption. The omitted baseline period is one week before a household was randomly assigned to the treatment or control group. Red circles indicate the point estimate and blue shading indicates the 95 percent confidence interval constructed with standard errors clustered by household that are robust to heteroskedasticity.

where S_i is an indicator equal to one if household i installs a smart thermostat, P_{it} is an indicator for household i's post-assignment status in time period t, X_{it} is a vector of controls, α_i^j is a household fixed effect, β_t^j is a vector of time effects, and u_{it}^j is a household/time varying unobservable.[38] We cluster standard errors at the household level to account for serial correlation (Bertrand, Duflo and Mullainathan, 2004) and estimate the model separately for each energy type. When j denotes electricity, energy is measured in kWh and the time period is an hour. If j denotes natural gas, the energy unit is a therm and observations are recorded daily.

Our parameter of interest is γ^j, which measures the differential change in energy use across pre- and post-intervention periods for smart relative to traditional thermostat households. This specification implicitly assumes that smart thermostats have a constant effect for all households. Given that individuals in our treatment sample are each optimizing over their household's expected energy savings and installation costs when deciding whether or not to follow through with installation of the smart thermostat, our treatment is likely to result in heterogeneous effects and Roy (1951) selection on gains. Consistent with this underlying model of behavior, there is incomplete installation compliance among the treated households in our experiment (see Figure 1). To address concerns of bias from noncompliance, we estimate a DDIV model that uses the experimental randomization as an instrument for the installation of a smart thermostat. Formally, we estimate γ^j using two-stage least squares (2SLS) methods with $E\left[Z_{it}^j u_{it}^j\right] = 0$, where $Z_{it}^j = \left(\alpha_i^j, \beta_t^j, T_i P_{it}, X_{it}\right)'$, and T_i is an indicator for household i's treatment status in our experiment.[39]

3.2 Identification

If the assumption of parallel trends holds in our DD setting, our instrument is relevant and valid, monotonicity holds, and there is one-sided noncompliance in our experiment, our DDIV coefficient of interest, γ^j, identifies the ATT of a smart thermostat (Cornelissen et al., 2016). This is the average impact of a smart thermostat on the energy use of households that

[38] We obtain similar results when estimating the model on the natural log of energy consumption ($\ln(e_{it}^j)$). If the randomization in our experiment is valid, our coefficient of interest is identified regardless of whether or not we include household fixed effects (α_i^j), time effects (β_t^j), or additional controls (X_{it}). Thus, we begin by estimating a basic specification of the model without any additional covariates that replaces α_i^j with $\alpha^j S_i$ and β_t^j with $\beta^j P_{it}$. Subsequent specifications add controls for the weather (which cannot be randomized a priori), household fixed effects, and various time effects to demonstrate robustness and improve the statistical precision of our estimates. Since post-assignment status, P_{it}, varies across households over the relatively narrow recruitment periods in our experiments (see Figure 2), we retain the indicator as a control in X_{it} in specifications with time effects. Results are qualitatively similar across all specifications.

[39] Equation 1 is the second-stage equation, and the first stage is modeled as

$$S_i P_{it} = \theta_i^j + \kappa_t^j + \lambda^j T_i P_{it} + X_{it}\pi^j + w_{it}^j. \tag{2}$$

install one. We discuss our identifying assumptions in more detail in Appendix Section F.[40]

4 Results

We begin by reporting estimates of the parameters in Equation 1 for electricity and natural gas in the next section. We then re-estimate the model on restricted subsamples of the data to investigate whether our main results mask significant, but offsetting, heterogeneous treatment effects. In the subsequent section, we estimate the model separately by quintile of ambient weather conditions, day of the week, hour of the day, hour of the day by weekday/weekend, quintile of the price of electricity, and during peak-use alerts.

4.1 Main Estimates

Table 1 summarizes multiple estimates of the effect of a smart thermostat on energy use based on each of the two experiments and the combined sample of all households recruited during both experiments. Panel A reports estimated effects on hourly electricity usage, and Panel B reports analogous estimates based on daily consumption of natural gas. Each ATT estimate reported in Columns (1) through (6) is based on a separate DDIV regression corresponding to the experimental sample indicated in the given row and the controls indicated at the bottom of the table.[41] Column (1) reports estimates of a basic version of the DDIV model without any fixed effects, time effects or other additional controls.[42] Column (2) reports estimates from a similar model that adds linear and quadratic county temperature and humidity readings to control for ambient weather conditions, as well as indicator for the Northern California experiment in the combined sample to control for potential differences across the two experiments. In Column (3), we add household fixed effects to the previous model that control for all of the time-invariant, unobserved characteristics of the home and household (e.g., age and square footage of the home, number of family members).[43] Column (4) reports estimates

[40]First-stage results in Appendix Section G provide strong support for instrument relevance. Appendix Table 5 and Figure 6 provide evidence in favor of instrument validity and parallel trends. Monotonicity is a standard assumption in IV settings that rules out irrational behavior. Finally, our experimental environment suggests that one-sided non-compliance is a reasonable assumption. In our context, this means that while some households randomized into treatment do not install a smart thermostat, no households in the control group install one. At the time of our experiment, smart thermostats were a nascent technology. According to data from the EIA RECS, two to three years after our experiment, only 4.09% of all households in the survey and 10.58% of observationally similar households owned a smart thermostat. Regardless, note that if we relax the one-sided noncompliance assumption to one of just monotonicity, our DDIV specification instead recovers the Local Average Treatment Effect (LATE) estimate of γ^j (Imbens and Angrist, 1994).

[41]See Appendix Tables 6, 7, and 8 for identical estimates with full regression diagnostics. The rk *LM* and Wald *F* statistics reported in those tables are first-stage diagnostic tests of under and weak identification, respectively, in models with non-i.i.d. errors. In all specifications, we reject the nulls of an under or weakly identified model. See Kleibergen and Paap (2006) for details.

[42]Relative to Equation 1, the model in Column (1) replaces α_i^j with $\alpha^j S_i$, β_t^j with $\beta^j P_{it}$, and restricts $\delta^j = 0$.

[43]Since the experiment indicator is perfectly collinear with recruitment wave, we drop the indicator from this and subsequent specifications.

from a model that adds month-of-year (MOY) effects to the previous specification in order to control for aggregate, time-varying effects such as seasonal variation in weather patterns. In Column (5), we add day-of-week effects to the previous model in order to control for variation in daily usage patterns due to occupant work and schooling schedules. Finally, in Column (6), we replace the time effects in the previous model with day-by-hour or day effects depending on which type of energy use is being modeled.[44]

The interpretation of these coefficients is straightforward. For example, the coefficient estimate of −0.001 reported in Column (6) of the third row in Panel A indicates that, across the two experiments, a smart thermostat causes a 0.001 kWh decrease in electricity usage per hour. The cluster-robust estimate of the standard error of 0.022 reported in parentheses indicates that this estimate is statistically insignificant.[45] To put the magnitudes of these effects in context, Column (7) reports mean energy use in the control group in the corresponding sample. The estimated effect is equivalent to 0.09% of the control group energy use of 1.140 kWhs per hour. The corresponding natural gas coefficient estimate in Panel B is equivalent to an increase of 1.70% of control energy use. Across all specifications in both panels, the lack of economic or statistical significance indicates that smart thermostats do not reduce energy usage.[46]

The estimated coefficients in Table 1 also cast doubt on the validity of the savings predicted by engineers. For example, they fall well short of the 10 to 23 percent savings predicted by engineers for smart thermostat manufacturers. Furthermore, these coefficients fall short of the savings predicted by engineers for policymakers in California in the TRM reports. While only the estimates in Column (6) reject these predictions for electricity at traditional levels of statistical significance, every estimate rejects the predicted savings for natural gas.

[44]The effects noted in Column (6) are day-by-hour effects in the model of *hourly* electricity meter readings (Panel A) and are day effects in the models of *daily* natural gas usage (Panel B). Estimates based on models that instead include week-of-year (WOY), month-by-year, week-by-year, and both day and hour-of-day effects result in qualitatively similar results. Furthermore, estimates based on models that include weather controls, day-of-week effects, and household-by-MOY (or household-by-WOY) effects do not affect our findings. The specification identifies off of hourly (electricity) or daily (gas) variation in usage within a household at a particular time of year. Intuitively, identification comes from the change in consumption in a given month of a the year for a treated home before and after treatment, relative to that same change for a control home. We also estimate models that include ZIP Code-by-MOY and ZIP Code-by-WOY effects that similarly identify off of variation within a neighborhood at a particular time of year. Again, results are qualitatively similar.

[45]Standard errors are clustered at the household level.

[46]We note that in some specifications, the "Both Experiments" estimate is not bounded by the individual experiment estimates (e.g., Column (1) of Panel B). This occurs because we model the Northern and Central California samples as two waves of the same experiment. In practice, this means we include minimal (if any) controls for differences between the two waves when estimating the combined sample models. Since this "single-experiment specification" of the "Both Experiments" models does not guarantee that the combined result is a convex combination of the individual Northern and Central California estimates, we also estimate a "multiple-experiments specification" that includes an indicator for the Northern California experiment and interacts it with the relevant DDIV control variables in each column. We report the results of this specification in Appendix Table 9. Results are qualitatively similar.

Table 1: ATT Estimates of the Effect of a Smart Thermostat on Energy Use

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Power Use (kWh/hour or thm/day)						Mean
Panel A: Electricity (kWh/hour)							
N. CA Experiment: ATT ($\hat{\gamma}^{kWh}$)	-0.055	-0.061	-0.016	-0.016	-0.016	-0.014	1.103
	(0.058)	(0.058)	(0.046)	(0.046)	(0.046)	(0.046)	(1.196)
C. CA Experiment: ATT ($\hat{\gamma}^{kWh}$)	0.009	0.005	0.002	0.002	0.002	0.003	1.191
	(0.029)	(0.028)	(0.025)	(0.025)	(0.025)	(0.025)	(1.273)
Both Experiments: ATT ($\hat{\gamma}^{kWh}$)	-0.031	-0.031	-0.003	-0.001	-0.001	-0.001	1.140
	(0.036)	(0.035)	(0.022)	(0.022)	(0.022)	(0.022)	(1.230)
Panel B: Natural Gas (thm/day)							
N. CA Experiment: ATT ($\hat{\gamma}^{thm}$)	-0.009	0.009	0.085	0.075	0.075	0.067	1.422
	(0.061)	(0.063)	(0.068)	(0.066)	(0.066)	(0.065)	(1.761)
C. CA Experiment: ATT ($\hat{\gamma}^{thm}$)	-0.003	0.004	-0.002	-0.001	-0.001	0.002	1.129
	(0.044)	(0.030)	(0.026)	(0.026)	(0.026)	(0.025)	(1.332)
Both Experiments: ATT ($\hat{\gamma}^{thm}$)	0.062	0.061	0.024	0.021	0.021	0.022	1.298
	(0.060)	(0.049)	(0.027)	(0.026)	(0.026)	(0.026)	(1.599)
Weather Controls		x	x	x	x	x	
HH Fixed Effects			x	x	x	x	
Month-of-Year Effects				x	x		
Day-of-Week Effects					x		
Day-by-Hour or Day Effects						x	

Note: Standard errors in parentheses are clustered at the household level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. Columns (1) through (6) report ATT estimates of the effect of a smart thermostat on energy use ($\hat{\gamma}^{j}$) based on separate DDIV regressions corresponding to the experimental sample indicated in the given row and the controls indicated at the bottom of the table. To put the magnitudes of these effects in context, Column (7) reports mean energy use in the control group in the corresponding sample. The samples used to produce the estimates in Panel A are based on *hourly* electricity meter readings in kWh, while the samples underlying the estimates in Panels B are based on *daily* natural gas meter readings in thm. Thus, the day-by-hour effects noted in Column (6) are included in the electricity model (Panel A) only and are day effects in the natural gas model (Panel B). Note that the estimates reported in Column (2) for samples that combine data from both experiments include an indicator equal to one for observations in the Northern California experiment. This indicator is perfectly co-linear with household fixed effects, so it is dropped from subsequent models. See Appendix Tables 6, 7, and 8 for full regression diagnostics. Based on the values of the rk *LM* and Wald *F* statistics reported in those tables, we reject the nulls of an under or weakly identified model across all specifications.

4.2 Heterogeneity in Treatment Effects

In order to investigate the possibility of significant, heterogeneous effects that are not apparent in the aggregate, we estimate the model conditional on various sub-sample selection criteria. Given that the results in Table 1 do not indicate any substantial differences between experiments and given that all subsequent specifications include household fixed effects, all results presented in this section start from a sample that pools the observations from both experiments. This should also give our models the best chance of recovering a significant heterogeneous treatment effect. We treat the model specification that includes household, month-of-year, and day-of-week effects (reported in Column (5) of Table 1) as our preferred specification because it is applicable to both samples with hourly- and daily-level variation. We then use this model as the basis for our subsequent analyses, but note that we deviate from this specification when the context warrants.

First, since smart thermostats will only have an effect on energy usage when there is a need for the HVAC system to heat or cool the house, moderate ambient temperature observations may attenuate a significant effect. To address this concern, Appendix Table 10 reports estimates by ambient temperature quintile. If the effect of a smart thermostat is only apparent when the HVAC system is in use, we would expect to find significant effects in the upper quintiles of temperature for electricity use and in the lower quintiles for natural gas. This is not the case. Only one of the 10 estimates is statistically significant, and the significant effect occurs in the second quintile of temperature for electricity consumption. Given the overall pattern of results, this finding is likely spurious.

Similarly, Appendix Table 11 reports estimates by ambient humidity quintile. In contrast to the results by temperature quintile, the estimates in Columns (4) and (5) of Panel A indicate that smart thermostats have a significant, negative effect on electricity use when the humidity is high (but not necessarily the temperature). We would expect to find this pattern of results if smart thermostats are successful at reducing the level of humidity in treated homes without deviating from a pre-programmed schedule, but individuals in the control group are prone to over-adjusting their traditional thermostats to less energy-efficient setpoints in order to mitigate the discomfort caused by high humidity. Consistent with this explanation, we do not find similar, significant effects on the consumption of natural gas (in Panel B). Alternatively, as it takes more energy to cool humid air than dry air, the pattern temperature and humidity results is consistent with smart thermostats providing small energy-efficiency gains that are only evident when the HVAC system has to work hardest.[47]

Next, since smart thermostats may only have an effect on energy use during the weekdays when individuals have more predictable schedules, Appendix Table 13 reports estimates by day of the week and by weekday/weekend. Across all days of the week and when we aggregate to the weekday/weekend level, we find no evidence that smart thermostats reduce energy

[47] Appendix Table 12 reports estimates from analogous models that condition on quintiles of the heat index (the perceived temperature) to rule out effects by the combined effects of temperature and humidity on comfort. We do not find significant results.

consumption.

Similarly, smart thermostats may only have an effect during the times of day that individuals typically schedule permanent temperature changes (e.g., before leaving for work/school or after returning home). Appendix Table 14 reports estimates by hour of the day. We are only able to calculate estimates conditional on the hour of the day for the effects of a smart thermostat on electricity usage, as we observe natural gas use at the daily level. Again, there is scant evidence that smart thermostats have a significant effect on energy use. To further test whether smart thermostats have effects only during certain hours of the day on certain days of the week (e.g., weekdays), Appendix Table 15 reports estimates by hour of the day separately for weekdays (Panel A) and weekends (Panel B). Consistent with our previous findings, we do not find evidence of significant effects on electricity use by hour of the day and day of the week.

Finally, since a potential benefit of smart technologies is that they enable consumers to better respond to spikes in demand and network congestion that lead to increased wholesale prices and brownouts (Joskow, 2012), we estimate two sets of models that condition on times when the social benefits of reduced electricity consumption are the greatest. Appendix Table 16 reports results by quintiles of electricity spot prices. If smart thermostats save energy at the most beneficial times, we would expect to see a statistically significant, negative effect in the fifth quintile of prices when production and external costs are the greatest, but none of the reported effects are statistically significant. To further isolate periods of high demand, Appendix Table 17 conditions on times when the system operator or utility issued a peak-usage alert. Column (1) reports estimates based on a sample of hours when CAISO issued hourly alerts. Holladay, Price and Wanamaker (2015) find that media coverage impacts consumer responses to such utility issued conservation appeals, so the estimates in Column (2) further condition on a sample of hours when there was both a CAISO alert and local media coverage of that alert. Neither of the resulting coefficient estimates is statistically significant.

In contrast, we find significant, negative effects when PG&E issues a daily, utility-wide alert.[48] Column (3) is based on a sample of all hours on alert days. PG&E advises their consumers to conserve electricity between 2:00pm and 7:00pm on these days, so the results in Columns (4) and (5) disaggregate this effect into off-peak and peak hours, respectively. The coefficient estimate in Column (3) (Column (4), Column (5)) indicates that smart thermostats reduce electricity consumption by 0.071 kWh (0.038 kWh, 0.122 kWh) relative to mean electricity use of 1.775 kWh (1.402 kWh, 2.893 kWh) in the control group. These estimates imply that smart thermostats result in electricity conservation of 4.0 percent (2.7 percent, 4.2 percent), although the off-peak estimate is not statistically significant.

[48]Why smart thermostats have significant effects in response to some types of alerts, but not others, remains an open question. One possibility is due to the different time periods covered by the alerts (hourly versus daily). The frequency of the alerts may also play a role: there are 173 CAISO hourly alerts, 121 of which received local media coverage, and 18 PG&E daily alerts observed over our sample period. Alternatively, the PG&E alert distribution system may be more robust or their messaging may be more salient. Brewer and Crozier (2022) find that the amplification of a similar peak-alert request by a state Governor increased conservation effects among smart-thermostat users.

While these estimates of savings are of the same order of magnitude as the overall savings predicted by engineers, we note that we find mixed evidence of these alert-period effects only during a narrow time horizon over which our PG&E estimates are obtained.[49] There are only 18 PG&E peak-alert days observed in our dataset. Additionally, the impact of smart thermostats on peak energy use is not uniform over the course of the alert period. Appendix Figure 21 further disaggregates these effects by hour of the day. While the estimates are rarely statistically different from zero, the hourly pattern suggest that households with a smart thermostat reduce electricity use before the peak alert period begins, but that they undo some of this benefit by ramping up their energy use before the peak period ends.

We are encouraged that there is some evidence that smart thermostats have their biggest impact on energy use when conservation is most beneficial, but our results suggest that these peak-period effects are nuanced. Both a better understanding of the differential responses of smart thermostat users to distinct types of alerts and a greater understanding of why smart thermostat households undo conservation benefits accrued over the course of the day are warranted. These could help manufacturers and policymakers design energy efficiency technology and programs that replicate the limited success of smart thermostats over broader time horizons. Overall, despite these encouraging findings, the complete set of results leaves us unable to reject the conclusion that smart thermostats under-deliver on their promised energy efficiency claims.

5 Why Did the Effects of Smart Thermostats Fail to Scale?

In this section, we supplement our experimental analysis by analyzing user interactions with the smart thermostat to explore why the benefits predicted by engineering studies did not manifest at scale. As aforementioned, there is a new line of research exploring why new technologies may not deliver their purported benefits (see, e.g., List, 2022). Examples of this phenomenon are many fold. For instance, Levitt (2008) shows that scaling-up car seat technologies did not produce predicted benefits because people did not fasten them in correctly. A more recent and timely example comes from the use of facemasks to prevent the spread of COVID-19 (Abaluck et al., 2022). This phenomenon is also found in the context of energy. Fowlie, Greenstone and Wolfram (2018) show that a broad set of technologies offered by the Weatherization Assistance Program do not deliver their purported energy savings. Yet, little is known about why the benefits of an important class of new technologies in the energy space that leverage "smart" functionalities may fail to scale.

To inform this issue, we turn to data on how users interacted with their smart thermostat (described in Section 2.5.2). Our analysis of smart thermostat interactions focuses on the Northern California experiment. As shown in Figure 5 and Appendix Table 4, our interactions data is heavily drawn from that experiment, as opposed to the Central California experiment,

[49] This favorable comparison does indicate that our experiments are sufficiently powered to detect the savings predicted by engineers when these savings actually occur in the data.

so we focus our analysis on those households.

Our analysis considers five questions. First, do users program their smart thermostat? Second, are programmed setpoints energy-efficient setpoints? Third, do users deviate from their programmed schedules? Fourth, do users deviate from their programmed schedule towards energy-efficient setpoints? Fifth, do users consume less energy than control households when they use the smart thermostats scheduled setpoints and overrides as intended by engineers?

By answering these questions we aim to sharpen our understanding of the failure of the smart thermostat to deliver energy savings. While engineering models assume households will utilize the functionality of a smart thermostat and do so to conserve energy, economic models are ultimately agnostic and emphasize the potential for preferences to interfere with the response desired by an engineer. We hypothesize that user optimization over more than just energy conservation explains why the engineering results do not scale. The first four questions consider whether households comply with the intended use of smart thermostat functionality. We ask these questions to inform the plausibility of different explanations for our experimental null results. The fifth question considers whether the subsamples of households that interact with their smart thermostat as an engineer may presume obtain the predicted energy savings. In contrast to the previous questions, the analysis that speaks to this question is designed to show that our proposed mechanism produces similar results in the field as in the engineering lab if we make assumptions akin to those engineering models implicitly make.

We find that households schedule setpoints and that these setpoints are broadly in line with energy-efficient suggestions. However, the setpoint overrides made easy by the smart thermostat are common and these overrides are biased towards more energy use: warmer setpoints in the winter and cooler setpoints in the summer. Finally, we find that households using the setpoint functionality as an engineering model assumes save as much as 8 to 19 percent on their consumption of natural gas. However, these savings are not found for electricity consumption, nor for households using the override functionality to obtain more energy-efficient setpoints.

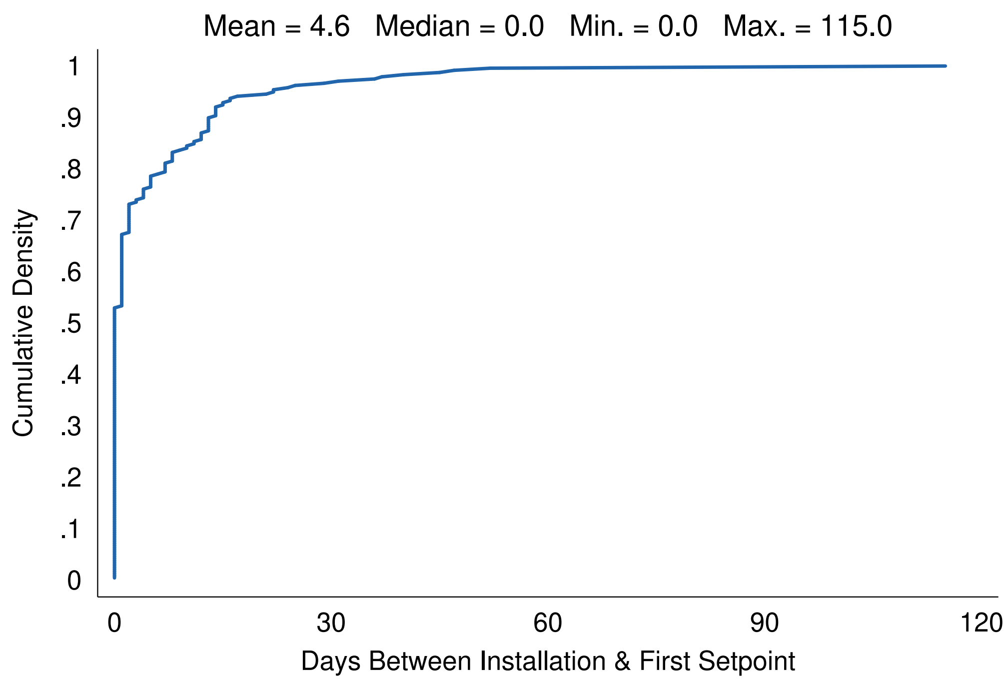
5.1 Do Users Program Their Smart Thermostats?

Peffer et al. (2013) find that programmable thermostats fail to achieve their advertised savings due, in part, to poor usability.[50] If users do not program schedules for their smart thermostats to follow because the interfaces are too complicated or they do not understand how thermostats and/or their HVAC systems work, we would not expect the installation of a smart thermostat to affect energy consumption.

[50] Programmable thermostats are a precursor technology to smart thermostats. The two types of thermostats share the ability to schedule permanent temperature setpoints in advance, but users cannot interact with programmable thermostats remotely, nor do they offer built-in setpoint framing. Peffer et al. (2013) report that they were so difficult to program that most users disabled their defining feature, and the ENERGY STAR program stopped certifying them in December 2009.

To determine what fraction of households who install the smart thermostat use the programmable features of the device and how long it takes them to begin doing so, Figure 7 plots the CDF of the time between the installation date and the first scheduled setpoint. The figure shows that almost all users who install a smart thermostat program at least one permanent setpoint, and most households do so almost immediately. The median time from installation to the first permanent setpoint is zero days.

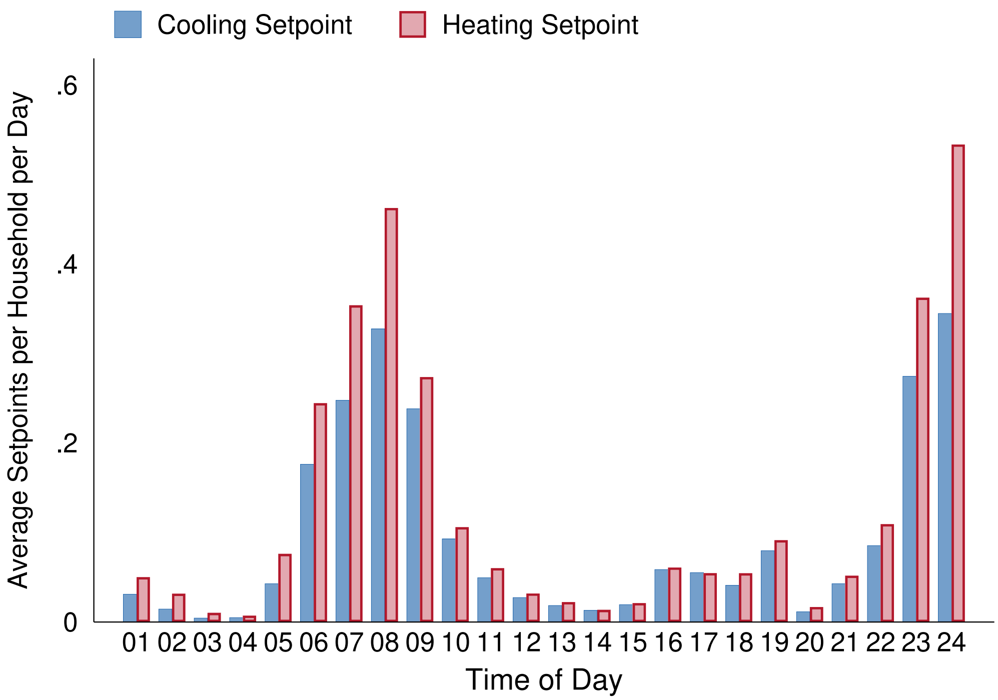
Figure 7: Distribution of Time from Installation to First Scheduled Setpoint



Note: This figure plots the cumulative density of time from the installation of a smart thermostat to the first scheduled setpoint, conditional on observing a household in the HVAC event data.

Additionally, users do not just quickly schedule a permanent setpoint, then fail to continue to use the smart features of the device. Individuals who have a smart thermostat installed as part of our experiment set an average of 3.749 (heating and cooling) setpoints per day. Figure 8 plots a measure of the frequency of permanent setpoints by hour of the day (denoted in military time) for both heating (red bars) and cooling (blue bars) setpoints. The figure provides visual evidence that setpoints occur frequently and when we would expect them: in the morning from about 5:00 AM until 10:00 AM when most users wake and leave for work and/or school. Similarly, there is a small increase in frequency of setpoints during the afternoon from 4:00 PM until 7:00 PM when users return home at the end of their days. Consistent with scheduling setpoints when most users go to sleep, we also observe frequent setpoints in the evening from about 10:00 PM until 12:00 AM. Thus, our analysis suggests that users do program their smart thermostats both quickly and frequently, consistent with engineering model assumptions.

Figure 8: Average Permanent Setpoints per Household per Day by Time of Day



Note: This figure plots the average number of daily setpoints for each hour of the day. Blue bars denote cooling setpoints, and red bars indicate heating setpoints.

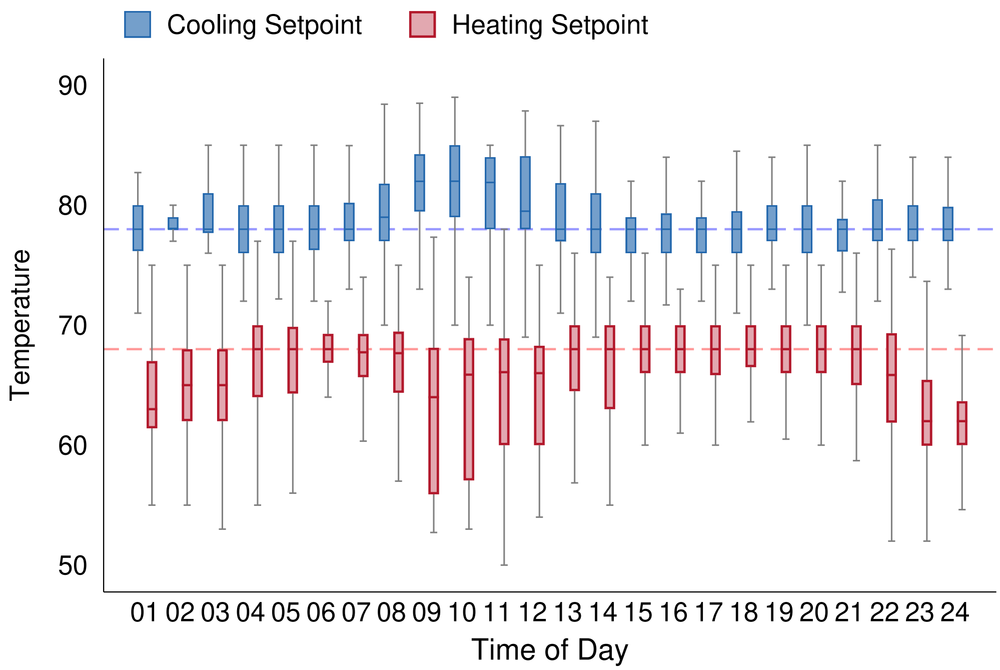
5.2 Are Programmed Setpoints Energy-Efficient Setpoints?

The previous analysis is consistent with users taking advantage of their device's scheduling feature, but is inconclusive as to whether or not they are programming setpoints to achieve energy savings. To inform the latter, Figure 9 is a box and whisker plot of heating and cooling setpoints by hour of the day. The dashed lines represent the cooling and heating temperature settings the DOE recommends for energy savings of 78 degrees F for cooling and 68 degrees F for heating (DOE, 2020). The figure illustrates that median (as well as the 25th and 75th percentiles of) temperatures are in line with the DOE's recommendations.[51] According to Appendix Table 4, cooling setpoints average 78.80 degrees F and are higher than heating setpoints, which average 63.95 degrees F. Additionally, the figure illustrates that there is temporal variation in setpoints over the course of the day consistent with individuals adjusting settings when they leave the house: cooling setpoints increase slightly starting at around 9:00 AM and drop back to baseline around 3:00 PM. Heating setpoints follow a similar, but opposite pattern with a more pronounced discrepancy between evening and daytime temperature setpoints. Overall, while the figure illustrates variation in setpoints across households, our

[51] The horizontal lines in the shaded boxes represent the median temperature setting, the ends of the boxes indicate the first and third quartiles, and the ends of the whiskers denote the upper/lower adjacent values.

analysis suggests that users program their smart thermostats to save energy.[52]

Figure 9: Box and Whisker Plots of Permanent Setpoints by Time of Day



Note: This figure presents a box and whisker plot of permanent setpoints by hour of day and whether the setpoint is for cooling or heating. The horizontal lines in the shaded boxes represent the median temperature setting, the ends of the boxes indicate the first and third quartiles, and the ends of the whiskers denote the upper and lower adjacent values.

5.3 Do Users Deviate from Their Programmed Schedules?

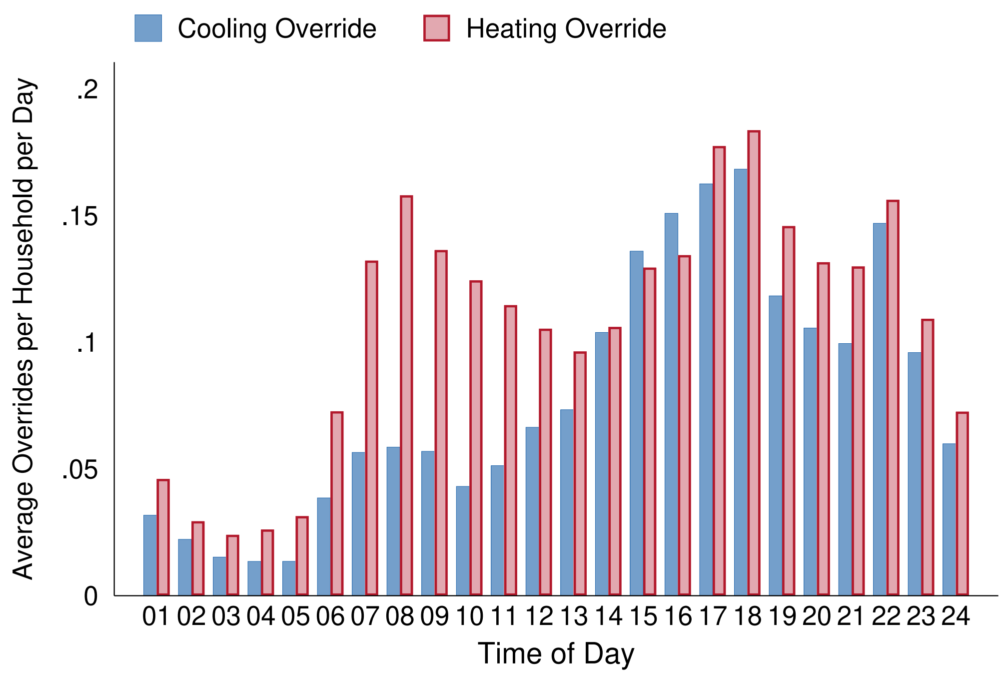
Given the evidence that indicates users program their smart thermostats and do so with energy savings in mind, we turn to an alternative explanation for our null findings. The remote features of the thermostat reduce the costs associated with both permanent and temporary setpoint changes. If users program their thermostats to reduce energy usage, but the ability to more easily adjust temperature settings via a computer or smart phone makes individuals more likely to deviate from their schedules, individuals may undo the benefits of their smart thermostat. If so, the effects of the scheduling and override features of smart thermostats have opposing effects on energy use and could result in a net null effect.

To explore this possibility, Figure 10 plots a measure of the frequency of setpoint overrides by

[52]Regarding the variation in setpoints, Appendix Table 4 reports standard deviations of 4.12 degrees for cooling and 5.58 degrees for heating setpoints.

time of the day.[53] As we would expect, overrides are more frequent when most individuals are likely to be awake, from about 6:00 AM to 11:00 PM. Heating overrides peak in the morning and early evening, while cooling overrides rise throughout the day until about 6:00 PM. More importantly given our focus, the figure illustrates that users often override their permanent schedule both when heating and cooling their homes. Compared to the previously noted 3.749 setpoints per day, users in our data temporarily override their permanent setpoints an average of 1.699 times per day. The hourly measures are also substantial relative to the number of permanent setpoints reported in Figure 8.

Figure 10: Average Temporary Overrides per Household per Day by Time of Day



Note: This figure presents the average number of daily overrides for heating and cooling in each hour of the day. Averages for heating are calculated over days in which the HVAC system heated the home and for cooling over days in which the HVAC system cooled the home.

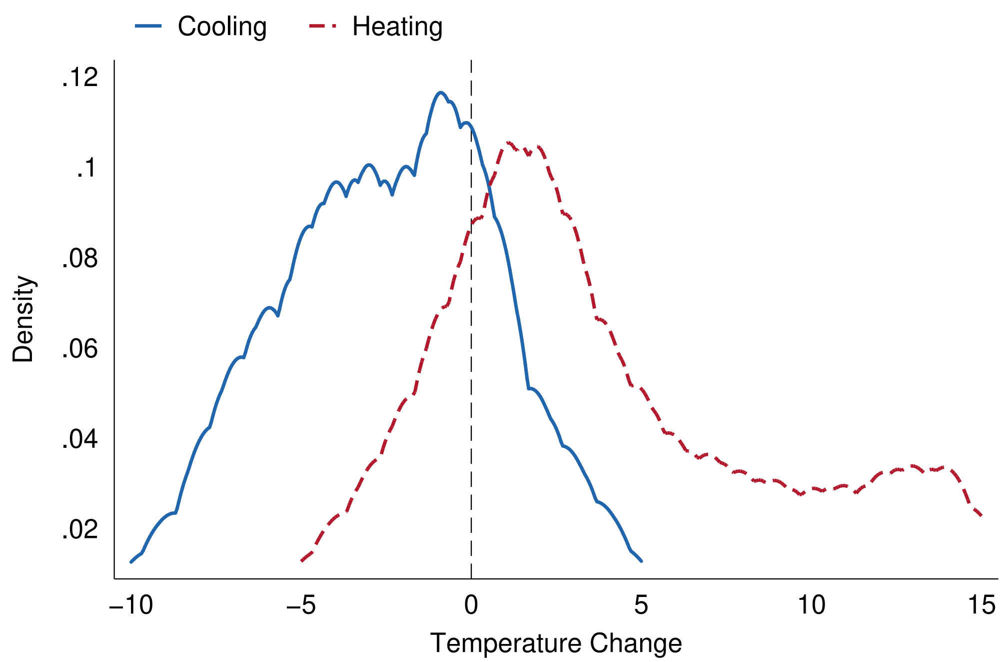
5.4 Do Users Deviate From Their Programmed Schedule Towards Energy-Efficient Setpoints?

Evidence that smart thermostat users frequently override their setpoints offers a potential explanation for our null findings. The features of the smart thermostat that lower adjustment

[53]The figure is the analog to Figure 8 for temporary overrides, save for our definition of "per day." While users program both heating and cooling setpoints every day, we typically only observe heating (cooling) overrides on heating (cooling) degree days. Given that we predominantly observe the HVAC system events data during the fall and winter, failure to address this issue results in heating and cooling override measures that are of different magnitudes. To account for this artifact in the data, we adjust the numerator of our measure to days on which the HVAC system heated or cooled the home to standardize the scales of the heating and cooling override measures.

costs both make it easier to override in ways that increase energy use (e.g., users no longer have to get off the couch or out of bed and walk to the thermostat when they are uncomfortable) and to override to decrease energy use (e.g., by toggling the HVAC system off when leaving home). To determine which effect dominates, Figure 11 plots kernel densities of the difference between the override temperature a user sets and the permanent setpoint, conditional on a temperature override, by temperature setting (cooling or heating). The figure illustrates that when users override their permanently scheduled setpoints, they generally do so in ways that use more energy: when cooling, they set temperatures colder and when heating, they set it warmer.[54] Taken together with the previous figure, our analysis suggests that individuals undo the benefits of their preset smart thermostat schedule when they are uncomfortable in the moment. This suggests a potential explanation for our null experimental findings, and is consistent with existing observational studies (Sachs et al., 2012; Peffer et al., 2013; Pritoni et al., 2015; Huchuk, O'brien and Sanner, 2020).

Figure 11: Density of Difference between Temporary Override and Permanent Setpoint Temperatures by Heating/Cooling



Note: This figure presents the density of the difference between a temporary override and a permanent setpoint for heating and cooling. Densities are truncated at the 5th and 95th percentiles.

[54]There is a non-trivial mass at large override-setpoint temperature differences (e.g., greater than 10 degrees F). This is primarily driven by a small number of households that program setpoints (~55 degrees F) that essentially turn off the HVAC system in the morning and override those setpoints at varying times in the afternoon/evening every day. This is consistent with using the programmable features of the smart thermostat based on a consistent daily departure time and a variable return time. Additionally, we note that the figure plots override-setpoint temperature differences, not override-ambient temperature differences. The ambient temperature may not actually be as low as the setpoint, so the actual temperature change caused by the override may not be so extreme.

5.5 Why Do Engineering Estimates Overstate the Energy Savings of Smart Thermostats?

There is an extensive literature on energy efficient technology that shows engineering estimates overstate the benefits of adoption (Davis, Fuchs and Gertler, 2014; Levinson, 2016; Zivin and Novan, 2016; Houde and Aldy, 2017; Fowlie, Greenstone and Wolfram, 2018; Alpízar, Bernedo and Ferraro, 2019; Davis, Martinez and Taboada, 2020; Christensen et al., 2021). In this subsection, we consider whether engineering estimates are overstated for smart thermostats, in part, because they assume an unrealistic level of compliance among adopters. To evaluate this explanation, we compare the energy consumption of the control group to different subsamples of households in the treatment group. These subsamples are selected to split households who are compliant (i.e., those who use setpoint scheduling and overrides in energy-efficient ways) from those who are not. We then relate the differences obtained with these compliant subsamples to the differences predicted by engineers, with similar magnitudes suggesting that compliance assumptions are a driver of benefits being overstated by engineers.

We classify the compliance of treatment group households by how diligently they use their device to save energy. We do so by defining three energy-efficiency types: high (H), low (L), and unknown types (?). Appendix Figure 18 illustrates how this classification builds on our existing experimental design. The unknown type is necessary both because of experimental non-compliers and because we do not observe all households who install a smart thermostat in the HVAC events data.[55] The high and low types are based on the distributions of two measures of energy-efficiency: the average number of permanent setpoints and temporary overrides observed per hour. For both metrics, we specify models based on various cutpoints between high and low types. Appendix Figure 19 plots the CDFs of both measures of behavior based on all households for which we observe interactions data. As an example, we define high-type households based on the permanent setpoint measure as those above the median and low types as those below the median. In contrast, for the other metric, we define high types as those below the median number of average overrides per hour and low types as those above the median.

Given these classifications, we interact indicators for type with treatment and estimate the following model,

$$e_{it}^{j} = \alpha_i^{j} + \beta_t^{j} + \sum_k \gamma_k^{j} R_i^{k} T_i P_{it} + X_{it}\beta_X^{j} + u_{it}^{j}, \tag{3}$$

where $k \in$\{H, L, ?\} index the three types, R_i^k is an indicator for household i being of type k and all other indexes, variables, and parameters are defined as in Equations 1 and 2. The parameters of interest in this model are γ_k^j which measures the average difference of energy type k between treatment group households of type k and the control group. Of course, this parameter *should not* be interpreted as causal because we are conditioning on post-treatment

[55] Of course, this set of types relies on post-treatment outcomes, which has important implications for the interpretation of any analysis that relies on them. We discuss this in more detail in the following paragraph.

outcomes. We condition on these post-treatment outcomes to explore whether this is the type of error that causes engineering predictions to go astray. If conditioning on post-treatment outcomes like compliance with the treatment reproduces the estimates predicted by engineers then it suggests this error in causal reasoning can explain why engineering estimates so often overstate energy savings.

Table 2 reports estimates of the γ_k^{thm} parameters based on these subsamples. Panel A reports estimated differences from a model based on the permanent setpoint type classification, and Panel B reports analogous estimates based on the temporary override type definition. To provide context for the other estimates, Column (1) reports estimates from a baseline DDITT model that does not differentiate by type. Consistent with our DDIV model estimates, the effects are not statistically significant. Columns (2) through (6) report estimates based on varying definitions of the high- vs. low-type percentile cutpoint.[56] Intuitively, these estimates are comparisons of trends in type-specific treatment group subsamples to the control group. The estimates in Panel A in these columns indicate that households with a smart thermostat that set above the 50th percentile of average permanent setpoints per hour are associated with statistically significantly less natural gas consumption compared to the control group. Additionally, smart-thermostat households with above the 90th percentile of setpoints are associated with the greatest difference in their natural gas use and the control group's. In contrast, low-type smart-thermostat users who program relatively few setpoints consumed a statistically indistinguishable quantity of natural gas relative to the control group.

Interestingly, estimates for high-type households with a smart thermostat who are above the median of permanent setpoints are broadly in line with engineering predictions. For example, Column (4) of Panel A shows the high types use 0.112 fewer thm per day than the control group, which consumes an average of 1.422 thm per day (see Table 1). This implies an 8 percent difference that, coupled with the subsequent estimates in Columns (5) and (6), line up well with the engineering estimates of 10-23% we discussed in Section 1. Moving to Panel B, we see a broadly similar relationship between the energy use of compliant subsamples of the treatment group and the control group's energy use. However, likely because temporary overrides alter energy use over shorter windows of time than permanent setpoints, the difference in energy use is never statistically significant for high types.

These results suggest that engineers overstated the benefits of smart thermostats, in part, because they assumed adopters would comply with the intended use of the devices. Instead, a significant share of the treatment group used their setpoint schedules and overrides in ways more easily explained by a desire for a more comfortable home than reductions in energy use. Engineering estimates effectively assume these types of adopters away, as evidenced by our ability to reproduce the benefits purported by engineers with compliant subsamples of the treatment group.

[56]For instance, the estimates reported in Column (4) of Panel A define high-types as those with more than the median number of setpoints per hour and low-types as those below the median.

Table 2: Estimates of the Association between a Smart Thermostat and Natural Gas Use by Setpoint and Override Type

	(1)	(2)	(3)	(4)	(5)	(6)
		High/Low-Type Percentile Cutpoint				
	Baseline	10	25	50	75	90
		Power Use (thm/day)				
Panel A: Permanent Setpoint Type Classification						
ITT ($\hat{\gamma}^{thm}$)	0.046					
	(0.044)					
High-Type Coef. ($\hat{\gamma}_H^{thm}$)		-0.044	-0.078	-0.112**	-0.147**	-0.266***
		(0.048)	(0.049)	(0.051)	(0.061)	(0.077)
Low-Type Coef. ($\hat{\gamma}_L^{thm}$)		-0.022	0.131	0.044	0.008	-0.015
		(0.145)	(0.103)	(0.070)	(0.056)	(0.050)
N	805	805	805	805	805	805
$N \times T$	398,243	398,243	398,243	398,243	398,243	398,243
R^2	0.650	0.650	0.650	0.650	0.650	0.650
F statistic	18.634	15.258	15.632	15.643	15.695	16.219
Panel B: Temporary Override Type Classification						
ITT ($\hat{\gamma}^{thm}$)	0.046					
	(0.044)					
High-Type Coef. ($\hat{\gamma}_H^{thm}$)		0.094	-0.034	-0.036	-0.050	-0.063
		(0.146)	(0.081)	(0.062)	(0.051)	(0.048)
Low-Type Coef. ($\hat{\gamma}_L^{thm}$)		-0.059	-0.047	-0.053	-0.007	0.449***
		(0.048)	(0.051)	(0.054)	(0.075)	(0.082)
N	805	805	805	805	805	805
$N \times T$	398,243	398,243	398,243	398,243	398,243	398,243
R^2	0.650	0.650	0.650	0.650	0.650	0.650
F statistic	18.634	15.314	15.247	15.262	15.376	20.493
Weather Controls	x	x	x	x	x	x
HH Fixed Effects	x	x	x	x	x	x
Day Effects	x	x	x	x	x	x

Note: Standard errors clustered at the household level in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. All estimates are based on a sample comprised of the Northern California experiment. The sample underlying the estimates in both panels is based on *daily* natural gas meter readings in thm. The coefficient estimate for the unknown type ($\hat{\gamma}_?^{thm}$) is 0.177 across all specifications, and it is statistically significant at the 1% level.

6 Conclusion

In recent years, citizens and lawmakers have become increasingly enthusiastic about adopting evidence-based policies and programs. Social scientists and engineers have delivered evidence of countless interventions that positively impact people's lives. And yet, most programs, when expanded, have not delivered the dramatic societal impacts promised. This is a common phenomenon known in the literature as "voltage drop" (List, 2022), but this type of predictable change is not accounted for in benefit–cost analysis. While the economics literature is beginning to provide insights into the features of ideas that make a policy predictably unscalable (Al-Ubaydli, List and Suskind, 2020), much remains unknown.

In this study, we use two framed field experiments to explore the scaling potential of smart technologies. Given that American households spend an average of over $2,200 on energy annually, and residential energy accounts for roughly 20% of the annual carbon dioxide pollution from energy production (EIA, 2018; 2019b), this exploration holds policy import. These high private and social costs have led to substantial interest in smart technologies that reduce energy use without reducing consumer utility by increasing efficiency. Given that the largest share of residential energy (almost 40%) goes to heating and cooling the home (EIA, 2019a), smart thermostats are an increasingly popular example of such a technology.

Smart thermostats allow individuals to program temperature setpoint schedules and adjust settings remotely via a smart phone application. While producers of these devices promise consumers substantial savings on their home heating and cooling bills, projected savings are often based on engineering simulations that fail to account for how people actually use their smart thermostats and therefore represent an upper bound on potential savings. Or they are based on studies that use non-experimental data and have methodological flaws that result in upwardly biased estimates of savings (see, e.g., Nest, 2019). Thus, the true marginal impact of smart thermostats on real world energy usage is uncertain.

Our work utilizes a framed field experiment to explore how smart technologies affect energy use—both through actual measurement and by investigating the mechanisms that prevent the realization of advertised energy savings. In our experiment, residential households are randomized into either a treatment group that receives a smart thermostat or a control group. In contrast to the commonly held prior that smart thermostats are an effective way to reduce residential energy use, we find little to no evidence that the installation of a smart thermosat reduces household energy consumption on average. This null result is robust to numerous specifications. We believe that the discord between the results of our field experiment and the extant belief stems from the source of the latter: engineering studies that do not adequately account for how individuals use their smart devices. We augment our experimental analysis with data on user interactions with their smart thermostat and find evidence that supports this belief.

There are many ways to extend our research. More than 90 percent of the households in our experiment faced a tiered energy pricing tariff. Although millions of Americans face tiered tariffs, they have been shown to be sub-optimal from a welfare point of view (Borenstein, 2012) and lead to the use of heuristics that dampen the response to changes in the marginal price of energy (Shin, 1985; Kahn and Wolak, 2013). Evidence from Jessoe and Rapson (2014), Harding and Lamarche (2016), Fabra et al. (2021), and Blonz et al. (2021) suggests that there are benefits from combining time-varying prices (time-of-use or real-time pricing) and smart tech. Future work should add to our understanding of the extent to which the returns to smart thermostats are enhanced when matched with smart economics - e.g., dynamic pricing plans that get energy prices right. A related avenue of inquiry would be to explore the impact of such technologies have on the price elasticity of energy demand (some preliminary evidence from Herter (2007) suggests that they do). If technology can enable people to better optimize their energy consumption, then price might become even more salient and therefore make people more marginal.

Viewed the lens of climate mitigation, our results provide little justification for the amount of subsidies directed towards smart thermostats; such technologies have no impact of energy use and associated greenhouse gas emissions. However, this does not mean that the technology has no social benefit and the subsidies cannot be justified. Perhaps the value of such technologies arise through adaptation and the ability of households to respond to the increased frequency/severity of extreme weather events that are projected to occur with climate change. A second avenue for future work is to explore this conjecture in greater detail. In this regard, we see promise in work designed to understand why smart thermostats are so popular amongst consumers given their costs and limited impact on energy use. Perhaps our focus on mitigation and energy savings has us thinking about the benefits of smart thermostats incorrectly. Rather than valuing smart thermostats for their expected savings, perhaps consumers value smart thermostats as a means to adapt to climate change and the increased severity of extreme weather events. This avenue speaks to the energy efficiency gap literature as outlined by Allcott and Greenstone (2012) by shifting attention to a broader set of characteristics than mitigation and expected energy savings.

A final avenue for extension would be to better understand when and how different smart technology features affect subsequent patterns of use. For instance, Harding and Lamarche (2016) and Blonz et al. (2021) estimate the effects of feature that automates temperature setting changes in response to time-of-use pricing. Given that various features of a smart thermostat may theoretically have opposing impact on energy use or simply facilitate a shift in energy use over the day, such a decomposition is a necessary next step in estimating the benefits of such technologies.

In summary, cooling and heating homes, powering transportation, and producing the wealth of goods and services enjoyed in modern economies are all heavily reliant on energy. Given that most of the world relies on non-renewable resources to produce energy and this reliance will not end any time soon (Covert, Greenstone and Knittel, 2016), one of the greatest policy challenges of this century is how to address the negative externalities associated with energy production. Without efforts to promote energy conservation and associated reductions in greenhouse gas emissions, future generations will face a lower quality of life due to a degraded environment.

Viewed through this lens, results from our paper provide a cautionary tale. Energy producers and policymakers alike are subsidizing smart technologies based on misleading information - estimates from engineering models of energy use. Had they instead complemented such an approach to evaluation with carefully designed field experiments, they would have realized that the estimates from engineering based models do not scale beyond the lab and could have reallocated some of the public funds spent subsidizing such technologies on more promising ways to promote energy conservation and associated reductions in greenhouse gas emissions.

The urgency of scaling up important ideas and enterprises impacts us every day, whether it's by protecting the health and safety of a community, improving the viability of a business, or enhancing the education and opportunities of a future generation. We hope that our paper represents a step towards ensuring that decision makers focus their energies on the use of science to identify the smartest, most scalable, policies possible.

References

Aarish, Carlyn, Matei Perussi, Andrew Rietz, and Dave Korn. 2015. "Evaluation of the 2013-2014 Programmable and Smart Thermostat Program." The Cadmus Group. 7

Abadie, Alberto. 2020. "Statistical Nonsignificance in Empirical Economics." *American Economic Review: Insights*, 2(2): 193–208. 6

Abaluck, Jason, Laura H. Kwong, Ashley Styczynski, Ashraful Haque, Md. Alamgir Kabir, Ellen Bates-Jefferys, Emily Crawford, Jade Benjamin-Chung, Shabib Raihan, Shadman Rahman, Salim Benhachmi, Neeti Zaman Bintee, Peter J. Winch, Maqsud Hossain, Hasan Mahmud Reza, Abdullah All Jaber, Shawkee Gulshan Momen, Aura Rahman, Faika Laz Banti, Tahrima Saiha Huq, Stephen P. Luby, and Ahmed Mushfiq Mobarak. 2022. "Impact of community masking on COVID-19: A cluster-randomized trial in Bangladesh." *Science*, 375(6577): eabi9069. 25

Abraham, Katherine G., Ron Haskins, Sherry Glied, Robert M. Groves, Robert Hahn, Hilary Hoynes, Jeffrey B. Liebman, Bruce D. Meyer, Paul Ohm, Nancy Potok, Kathleen Rice Moiser, Robert J. Shea, Latanya Sweeney, Kenneth R. Troske, and Kim R. Wallin. 2017. "The Promise of Evidence-Based Policymaking: Report of the Commission on Evidence-Based Policymaking." Commission on Evidence-Based Policymaking, Washington, DC. 5

Acemoglu, Daron, Philippe Aghion, Leonardo Bursztyn, and David Hemous. 2012. "The Environment and Directed Technical Change." *American Economic Review*, 102(1): 131–66. 2

Acemoglu, Daron, Ufuk Akcigit, Douglas Hanley, and William Kerr. 2016. "Transition to Clean Technology." *Journal of Political Economy*, 124(1): 52–104. 2

Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica*, 60(2): 323–351. 2

Agnew, Ken, and Mimi Goldberg. 2013. "Chapter 8: Whole Building Retrofit with Consumption Data Analysis Evaluation Protocol. The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures." National Renewable Energy Laboratory NREL/SR-7A30–53827. 7

Allcott, Hunt. 2011. "Social norms and energy conservation." *Journal of Public Economics*, 95(9): 1082 – 1095. Special Issue: The Role of Firms in Tax Systems. 3, 8

Allcott, Hunt. 2015. "Site selection bias in program evaluation." *Quarterly Journal of Economics*, 130(3): 1117–1165. 5

Allcott, Hunt, and Michael Greenstone. 2012. "Is There an Energy Efficiency Gap?" *Journal of Economic Perspectives*, 26(1): 3–28. 3, 36

Allcott, Hunt, and Todd Rogers. 2014. "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." *American Economic Review*, 104(10): 3003–37. 3

Alpízar, Francisco, María Bernedo, and Paul J. Ferraro. 2019. "Input Efficiency as a Solution to Externalities: Engineers vs Behavioral Scientists in a Randomized Controlled Trial." 4, 32

Al-Ubaydli, Omar, John A. List, and Dana L. Suskind. 2017. "What Can We Learn from Experiments? Understanding the Threats to the Scalability of Experimental Results." *American Economic Review*, 107(5): 282–86. 2

Al-Ubaydli, Omar, John A. List, and Dana Suskind. 2020. "2017 Klein Lecture: The Science of Using Science: Toward an Understanding of the Threats to Scalability." *International Economic Review*, 61(4): 1387–1409. 3, 5, 35

Al-Ubaydli, Omar, John A. List, Danielle LoRe, and Dana Suskind. 2017. "Scaling for Economists: Lessons from the Non-Adherence Problem in the Medical Literature." *Journal of Economic Perspectives*, 31(4): 125–44. 2, 5

Apex Analytics, LLC. 2014. "Energy Trust of Oregon Nest Learning Thermostat Heat Pump Control Pilot Evaluation." 7

Apex Analytics, LLC. 2016. "Energy Trust of Oregon Smart Thermostat Pilot Evaluation." 7, 10

Auffhammer, Maximilian, and Erin T Mansur. 2014. "Measuring climatic impacts on energy consumption: A review of the empirical literature." *Energy Economics*, 46: 522–530. 4

Ayres, Ian, Sophie Raseman, and Alice Shih. 2012. "Evidence from Two Large Field Experiments that Peer Comparison Feedback Can Reduce Residential Energy Usage." *The Journal of Law, Economics, and Organization*, 29(5): 992–1022. 3

Barrett, Scott. 2006. "Climate Treaties and "Breakthrough" Technologies." *American Economic Review*, 96(2): 22–25. 2

Battaglini, Marco, and Bård Harstad. 2016. "Participation and Duration of Environmental Agreements." *Journal of Political Economy*, 124(1): 160–204. 2

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-In-Differences Estimates?" *The Quarterly Journal of Economics*, 119(1): 249. 19

Blonz, Joshua, Karen Palmer, Casey J. Wichman, and Wietelman Derek C. 2021. "Smart Thermostats, Automation, and Time-Varying Prices." Resources for the Future (RFF) Working Paper 21-20. 35, 36

Bloomberg Finance L.P. 2019. "Sustainable Energy in America Factbook." The Business Council for Sustainable Energy. Smart thermostat facts on page 112. 6

Borenstein, Severin. 2012. "The redistributional impact of nonlinear electricity pricing." *American Economic Journal: Economic Policy*, 4(3): 56–90. 35

Brandon, Alec. 2019. "Social Pressure and Self-Selection in Experimental Policy Evaluation." 5

Brewer, Dylan, and Jim Crozier. 2022. "Who Heeds the Call to Conserve in an Energy Emergency? Evidence from SmartThermostat Data." 24

Broaddus, Phil, Barb Ryan, and Kelly Marrin. 2016. "PG&E Smart Thermostat Study: First Year Findings." 7

Broaddus, Phil, Barb Ryan, and Kelly Marrin. 2018. "PG&E Smart Thermostat Study: Second Year Findings." 7

California Municipal Utilities Association. 2017. "Savings Estimation Technical Reference Manual." ERS. 3

Cheng, Sierra, Erica J. McDonald, Matthew C. Cheung, Vanessa S. Arciero, Mahin Qureshi, Di Jiang, Doreen Ezeife, Mona Sabharwal, Alexandra Chambers, Dolly Han, Natasha Leighl, Kelley-Anne Sabarre, and Kelvin K.W. Chan. 2017. "Do the American Society of Clinical Oncology Value Framework and the European Society of Medical Oncology Magnitude of Clinical Benefit Scale Measure the Same Construct of Clinical Benefit?" *Journal of Clinical Oncology*, 35(24): 2764–2771. PMID: 28574778. 5

Christensen, Peter, Paul Francisco, Erica Myers, and Mateus Souza. 2021. "Decomposing the Wedge between Projected and Realized Returns in Energy Efficiency Programs." *The Review of Economics and Statistics*, 1–46. 4, 32

Clarke, Damian, and Kathya Tapia Schythe. 2020. "Implementing the Panel Event Study." Institute of Labor Economics (IZA) IZA Discussion Papers 13524, Bonn. 17

Cornelissen, Thomas, Christian Dustmann, Anna Raute, and Uta Schönberg. 2016. "From LATE to MTE: Alternative Methods for the Evaluation of Policy Interventions." *Labour Economics*, 41: 47 – 60. SOLE/EALE Conference Issue 2015. 19

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

Covert, Thomas, Michael Greenstone, and Christopher R. Knittel. 2016. "Will We Ever Stop Using Fossil Fuels?" *Journal of Economic Perspectives*, 30(1): 117–38. 36

Daken, Abigail A, Alan K Meier, and Douglas G Frazee. 2016. "Do Connected Thermostats Save Energy?" ACEEE Summer Study on Energy Efficiency in Buildings, eScholarship, University of California. 5, 9

Davis, Lucas W., Alan Fuchs, and Paul Gertler. 2014. "Cash for Coolers: Evaluating a Large-Scale Appliance Replacement Program in Mexico." *American Economic Journal: Economic Policy*, 6(4): 207–38. 4, 32

Davis, Lucas W., Sebastian Martinez, and Bibiana Taboada. 2020. "How effective is energy-efficient housing? Evidence from a field trial in Mexico." *Journal of Development Economics*, 143: 102390. 4, 32

Department of Energy. 2016. "Smart Grid Investment Grant Program Final Report." U.S. Department of Energy, Office of Electricity Delivery & Energy Reliability. 6

Department of Energy. 2017. "Guide for States: Establishing and Maintaining Technical Reference Manuals for Energy Efficiency Measures." SEE Action Network. 3

Department of Energy. 2020. "Thermostats." *https://www.energy.gov/energysaver/thermostats*, Accessed: July 21, 2020. 28

Deschênes, Olivier, and Michael Greenstone. 2011. "Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the US." *American Economic Journal: Applied Economics*, 3(4): 152–85. 4

Dolan, Paul, and Robert Metcalfe. 2015. "Neighbors, Knowledge, and Nuggets: Two Natural Field Experiments on the Role of Incentives on Energy Conservation." *Becker Friedman Institute for Research in Economics Working Paper*. 3

ecobee. 2019. "Savings from your ecobee." *https://www.ecobee.com/savings/*, Accessed: September 10, 2019. 3, 5

Energy Information Administration. 2018. "Energy Consumption and Expenditures Tables." U.S. Department of Energy, 2015 Residential Energy Consumption Survey. 35

Energy Information Administration. 2019*a*. "Annual Energy Outlook 2019 with Projections to 2050." U.S. Department of Energy, Office of Energy Analysis. 35

Energy Information Administration. 2019*b*. "Monthly Energy Review." U.S. Department of Energy, Office of Energy Statistics. 35

Environmental Protection Agency. 2019. "ENERGY STAR Summary of HVAC & Smart Thermostat Programs." U.S. Environmental Protection Agency. Accessed: July 24, 2020. 6

Fabra, Natalia, David Rapson, Mar Reguant, and Jingyuan Wang. 2021. "Estimating the Elasticity to Real-Time Pricing: Evidence from the Spanish Electricity Market." *AEA Papers and Proceedings*, 111: 425–29. 35

Ferraro, Paul J., and Michael K. Price. 2013. "Using Nonpecuniary Strategies to Influence Behavior: Evidence from a Large-Scale Field Experiment." *The Review of Economics and Statistics*, 95(1): 64–73. 3

Fowlie, Meredith, Michael Greenstone, and Catherine Wolfram. 2018. "Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program." *The Quarterly Journal of Economics*, 133(3): 1597–1644. 4, 25, 32

Ge, Qi, and Benjamin Ho. 2019. "Energy Use and Temperature Habituation: Evidence from High Frequency Thermostat Usage Data." *Economic Inquiry*, 57(2): 1196–1214. 12

Ghanem, Dalia, and Aaron Smith. 2021. "What Are the Benefits of High-Frequency Data for Fixed Effects Panel Models?" *Journal of the Association of Environmental and Resource Economists*, 8(2): 199–234. 7

Goeschl, Timo, and Grischa Perino. 2017. "The Climate Policy Hold-Up: Green Technologies, Intellectual Property Rights, and the Abatement Incentives of International Agreements." *The Scandinavian Journal of Economics*, 119(3): 709–732. 2

Gottfredson, Denise C, Thomas D Cook, Frances EM Gardner, Deborah Gorman-Smith, George W Howe, Irwin N Sandler, and Kathryn M Zafft. 2015. "Standards of Evidence for Efficacy, Effectiveness, and Scale-up Research in Prevention Science: Next Generation." *Prevention science*, 16(7): 893–926. 5

Hall, Bronwyn H. 2005. "Innovation and Diffusion." In *The Oxford Handbook of Innovation*. , ed. Jan Fagerberg, David C Mowery and Richard R. Nelson, Chapter 17, 459–485. Oxford University Press. 2

Harding, Matthew, and Carlos Lamarche. 2016. "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*, 35(4): 906–931. 35, 36

Harrison, Glenn W., and John A. List. 2004. "Field Experiments." *Journal of Economic Literature*, 42(4): 1009–1055. 9

Harstad, Bård. 2020. "Technology and Time Inconsistency." *Journal of Political Economy*, 128(7): 2653–2689. 2

Harstad, Bård. 2012. "Climate Contracts: A Game of Emissions, Investments, Negotiations, and Renegotiations." *The Review of Economic Studies*, 79(4): 1527–1557. 2

Heckman, James J., and Jeffrey A. Smith. 1995. "Assessing the Case for Social Experiments." *Journal of Economic Perspectives*, 9(2): 85–110. 57

Herter, Karen. 2007. "Residential implementation of critical-peak pricing of electricity." *Energy Policy*, 35(4): 2121 – 2130. 36

Ho, Benjamin. 2014. "Understanding Energy Efficiency Benefits from Smart Thermostats in Southern California." EnergyHub. 7

Hoel, Michael, and Aart De Zeeuw. 2010. "Can a Focus on Breakthrough Technologies Improve the Performance of International Environmental Agreements?" *Environmental and Resource Economics*, 47: 395–406. 2

Holladay, J. Scott, Michael K. Price, and Marianne Wanamaker. 2015. "The perverse impact of calling for energy conservation." *Journal of Economic Behavior and Organization*, 110: 1–18. 24

Houde, Sébastien, and Joseph E. Aldy. 2017. "Consumers' Response to State Energy Efficient Appliance Rebate Programs." *American Economic Journal: Economic Policy*, 9(4): 227–55. 4, 6, 32

Huchuk, Brent, William O'brien, and Scott Sanner. 2020. "Exploring smart thermostat users' schedule override behaviors and the energy consequences." *Science and Technology for the Built Environment*, 27(2): 195–210. 31

Hudson, Sally, Peter Hull, and Jack Liebersohn. 2017. "Interpreting Instrumented Difference-In-Differences." 57

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, 62(2): 467–475. 20, 57

Ito, Seiro. 2007. "A Practical Guide to the Program Evaluation Methods." *Health Service and Poverty: Making Health Services More Accessible to the Poor*, 143–203. 58

Jessoe, Katrina, and David Rapson. 2014. "Knowledge Is (Less) Power: Experimental Evidence from Residential Energy Use." *American Economic Review*, 104(4): 1417–38. 35

Joskow, Paul L. 2012. "Creating a Smarter U.S. Electricity Grid." *Journal of Economic Perspectives*, 26(1): 29–48. 6, 24

Kahn, Matthew, and Frank Wolak. 2013. "Using information to improve the effectiveness of nonlinear pricing: Evidence from a field experiment." 35

Kelsven, Phillip, Robert Weber, and Eva Urbatsch. 2016. "Nest Learning Thermostat Pilot Program Savings Assessment." Bonneville Power Administration & Franklin Public Utility District. 7

Kilbourne, Amy M, Mary S Neumann, Harold A Pincus, Mark S Bauer, and Ronald Stall. 2007. "Implementing evidence-based interventions in health care: application of the replicating effective programs framework." *Implementation Science*, 2(1): 1–10. 5

Kleibergen, Frank, and Richard Paap. 2006. "Generalized reduced rank tests using the singular value decomposition." *Journal of Econometrics*, 133(1): 97 – 126. 20

Kotchen, Matthew J. 2017. "Longer-Run Evidence on Whether Building Energy Codes Reduce Residential Energy Consumption." *Journal of the Association of Environmental and Resource Economists*, 4(1): 135–153. 8

Levinson, Arik. 2016. "How Much Energy Do Building Energy Codes Save? Evidence from California Houses." *American Economic Review*, 106(10): 2867–94. 4, 8, 32

Levitt, Steven D. 2008. "Evidence that Seat Belts Are as Effective as Child Safety Seats in Preventing Death for Children Aged Two and Up." *The Review of Economics and Statistics*, 90(1): 158–163. 25

List, John A. 2022. *The Voltage Effect: How to Make Good Ideas Great and Great Ideas Scale.* Currency. 2, 5, 25, 35

Muralidharan, Karthik, and Paul Niehaus. 2017. "Experimentation at Scale." *Journal of Economic Perspectives*, 31(4): 103–24. 2

Nest. 2019. "It saves energy. And here's the proof." *https://www.nest.com/thermostats/real-savings/*, Accessed: September 10, 2019. 3, 7, 35

Nest Labs. 2014. "Energy Savings from the Nest Learning Thermostat: The Impact of Nest Learning Thermostat." 7

Nest Labs. 2015. "Energy Savings from the Nest Learning Thermostat: Energy Bill Analysis Results." 7

Novan, Kevin, Aaron Smith, and Tianxia Zhou. 2022. "Residential Building Codes Do Save Energy: Evidence from Hourly Smart-Meter Data." *The Review of Economics and Statistics*, 104(3): 483–500. 8

Park, Toby, Pantelis Solomon, Michael Sanders, Marius Dietsch, Julie Faller, Daniel Gibbons, Elisabeth Costa, and Alex Tupper. 2017. "Evaluating the Nest Learning Thermostat: Four Field Experiments Evaluating the Energy Saving Potential of Nest's Smart Heating Control." The Behavioural Insights Team. 7

Peffer, Therese, Daniel Perry, Marco Pritoni, Cecilia Aragon, and Alan Meier. 2013. "Facilitating energy savings with programmable thermostats: evaluation and guidelines for the thermostat user interface." *Ergonomics*, 56(3): 463–479. PMID: 23005033. 3, 10, 26, 31

Peffer, Therese, Marco Pritoni, Alan Meier, Cecilia Aragon, and Daniel Perry. 2011. "How people use thermostats in homes: A review." *Building and Environment*, 46(12): 2529 – 2541. 10

Pritoni, Marco, Alan K. Meier, Cecilia Aragon, Daniel Perry, and Therese Peffer. 2015. "Energy efficiency and the misuse of programmable thermostats: The effectiveness of crowdsourcing for understanding household behavior." *Energy Research & Social Science*, 8: 190 – 197. 31

Robinson, Jen, Ram Narayanamurthy, Bienvenido Clarin, Christine Lee, and Pranshu Bansal. 2016. "National Study of Potential of Smart Thermostats for Energy Efficiency and Demand Response." ACEEE Summer Study on Energy Efficiency in Buildings. 7

Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy*, 98(5): S71–S102. 2

Roy, A. D. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers*, 3(2): 135–146. 19

Sachs, Olga, Verena Tiefenbeck, Caroline Duvier, Angela Qin, Kate Cheney, Craig Akers, and Kurt Roth. 2012. "Field evaluation of programmable Thermostats." National Renewable Energy Lab.(NREL), Golden, CO (United States). 31

Schellenberg, Josh, Alana Lemarchand, and Alexandra Wein. 2017. "Xcel Energy Colorado Smart Thermostat Pilot - Evaluation Report." Nexant, Inc. 7

Schmidheiny, Kurt, and Sebastian Siegloch. 2019. "On Event Study Designs and Distributed-Lag Models: Equivalence, Generalization and Practical Implications." Institute of Labor Economics (IZA) CESifo Working Paper 7481, Munich:Center for Economic Studies and ifo Institute (CESifo). 17

Shin, Jeong-Shik. 1985. "Perception of price when price information is costly: evidence from residential electricity demand." *Review of Economics and Statistics*, 591–598. 35

Shobe, William, Charles Holt, and Thaddeus Huetteman. 2014. "Elements of emission market design: An experimental analysis of California's market for greenhouse gas allowances." *Journal of Economic Behavior and Organization*, 107: 402–420. 13

Skinner, Jonathan, and Douglas Staiger. 2007. "Technology Adoption from Hybrid Corn to Beta-Blockers." *Hard-to-Measure Goods and Services: Essays in Honor of Zvi Griliches*, , ed. Ernst R. Berndt and Charles R. Hulten, 545–570. University of Chicago Press. 2

Stewart, James, and Jeremy Jackson. 2015. "Energy Savings from Honeywell Total Connect Comfort Thermostats-Estimates from 2nd National Impact Study." The Cadmus Group. 7

Supplee, Lauren H, and Aleta L Meyer. 2015. "The Intersection Between Prevention Science and Evidence-Based Policy: How the SPR Evidence Standards Support Human Services Prevention Programs." *Prevention Science*, 16(7): 938–942. 5

Supplee, Lauren H, and Allison Metz. 2015. "Opportunities and Challenges in Evidence-Based Social Policy." *Social Policy Report*, 28(4): 1–16. 5

Urban, Bryan, and Carla Gomez. 2013. "A Case for Thermostat User Models." International Building Performance Simulation Association. 5

Urban, Bryan, Diana Elliott, and Olga Sachs. 2012. "Towards Better Modeling of Residential Thermostats." Proceedings of SimBuild 1. 5

U.S. Chamber of Commerce Foundation. 2012. "Enterprising States: Policies that Produce." Praxis Stratgy Group. 2

Ward, Bryan, James Stewart, and Jeremy Jackson. 2014. "Energy Savings from Honeywell Total Connect Comfort Thermostats." The Cadmus Group. 7

Weiss, Michael J., Howard S. Bloom, and Thomas Brock. 2014. "A Conceptual Framework for Studying the Sources of Variation in Program Effects." *Journal of Policy Analysis and Management*, 33(3): 778–808. 5

Young, Alwyn. 1995. "The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience." *Quarterly Journal of Economics*, 110(3): 641–680. 2

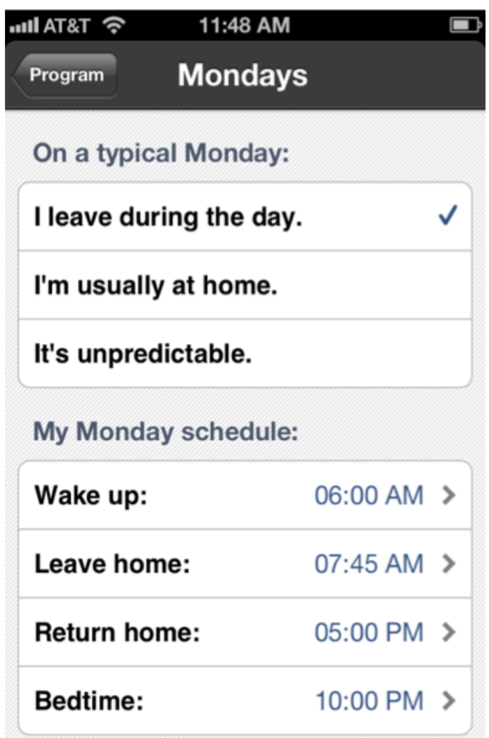
Zivin, Joshua Graff, and Kevin Novan. 2016. "Upgrading Efficiency and Behavior: Electricity Savings from Residential Weatherization Programs." *Energy Journal*, 37(4): 1–23. 4, 32

Appendices

A Smart Thermostat

Figure 12: Smart Thermostat Overview

(a) Interfaces: The left panel shows the web portal, the middle panel shows the smartphone app, and the right panel shows the thermostat.



(b) Permanent Setpoint Scheduling: Screenshot of the smartphone app scheduling interface.

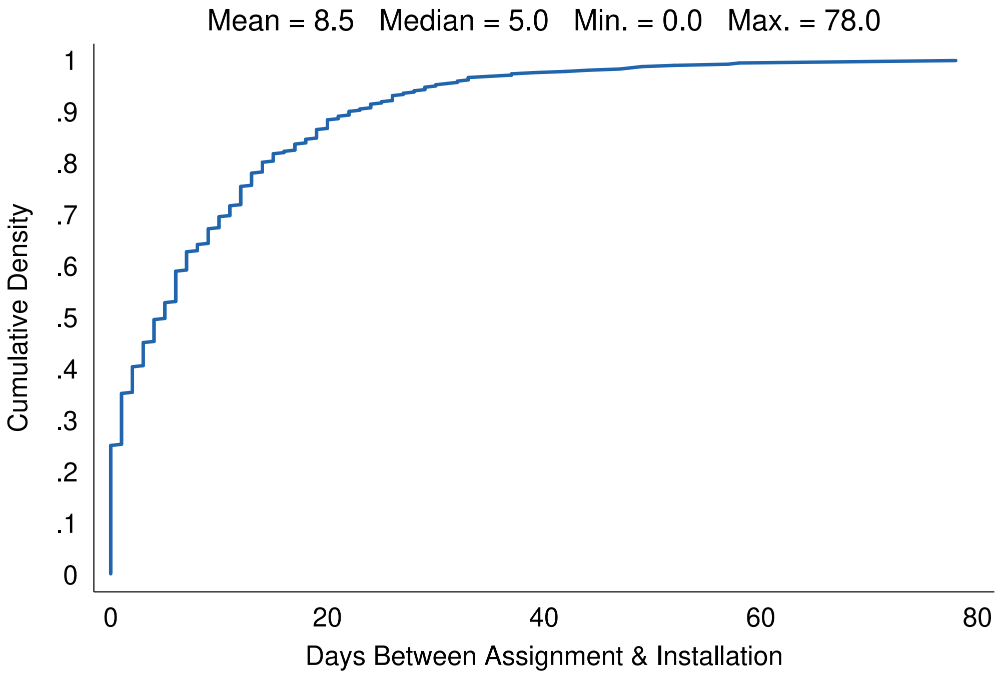
Figure 13: Smart Thermostat Features

(a) Setpoint Choice Messaging: Screenshots of smartphone app that shows the messaging associated with different temperature set points.

(b) Temporary Overrides: Screenshots of the smartphone app that facilitates changes to the temperature setpoint. The left panel shows the interface after the user indicates she is not home. The right panel shows the same interface when the user indicates she is at home.

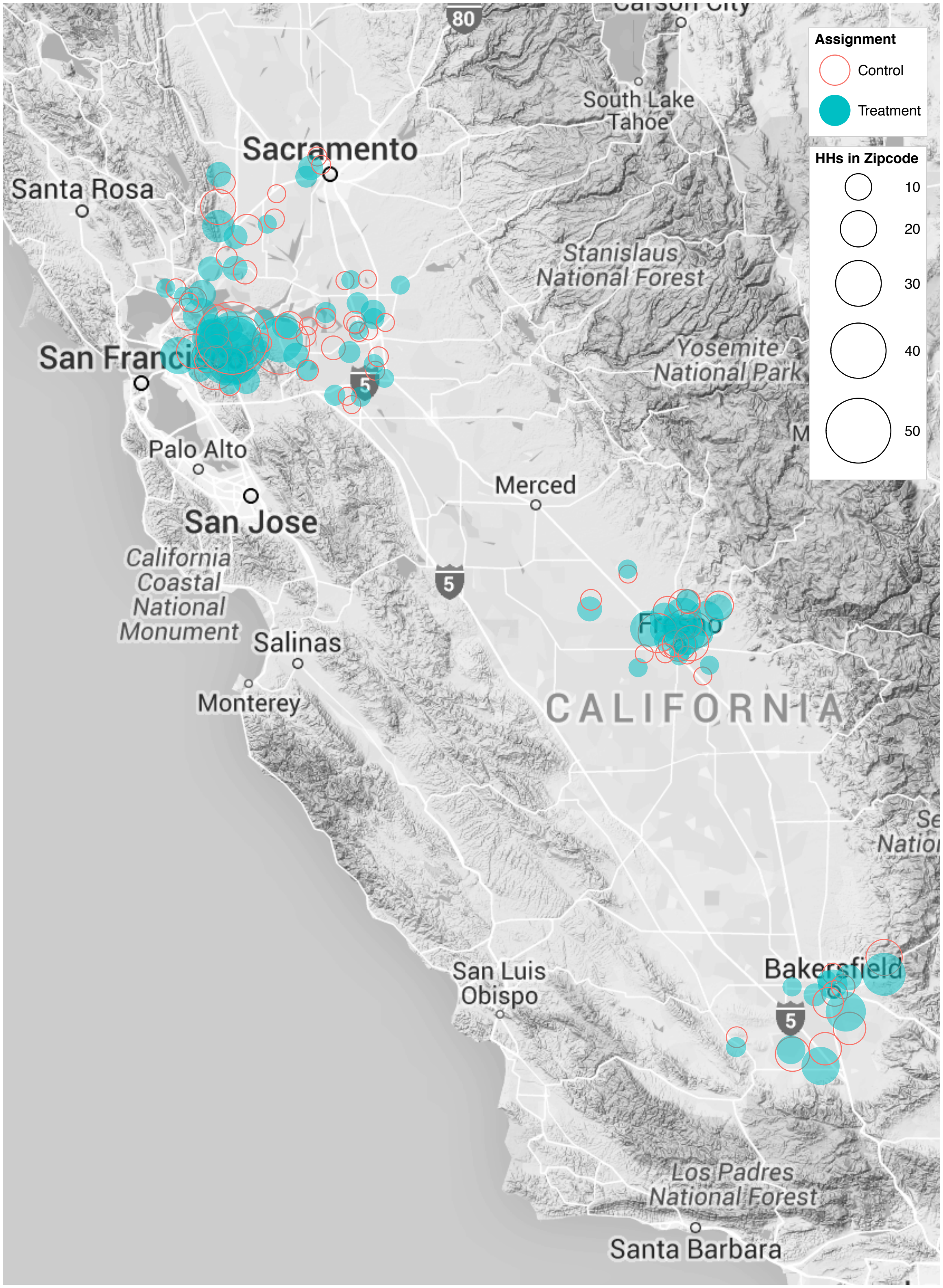
B Experimental Data

Figure 14: Conditional Distribution of Time from Assignment to Installation



Note: This figure plots the cumulative density of time from experimental assignment to the installation of the smart thermostat, conditional on eventual thermostat installation.

Figure 15: Locations of Treatment and Control Groups



Note: This figure displays a map of the locations of households in the two experiments by the Zip Code in which the household is located and the treatment or control status of the household.

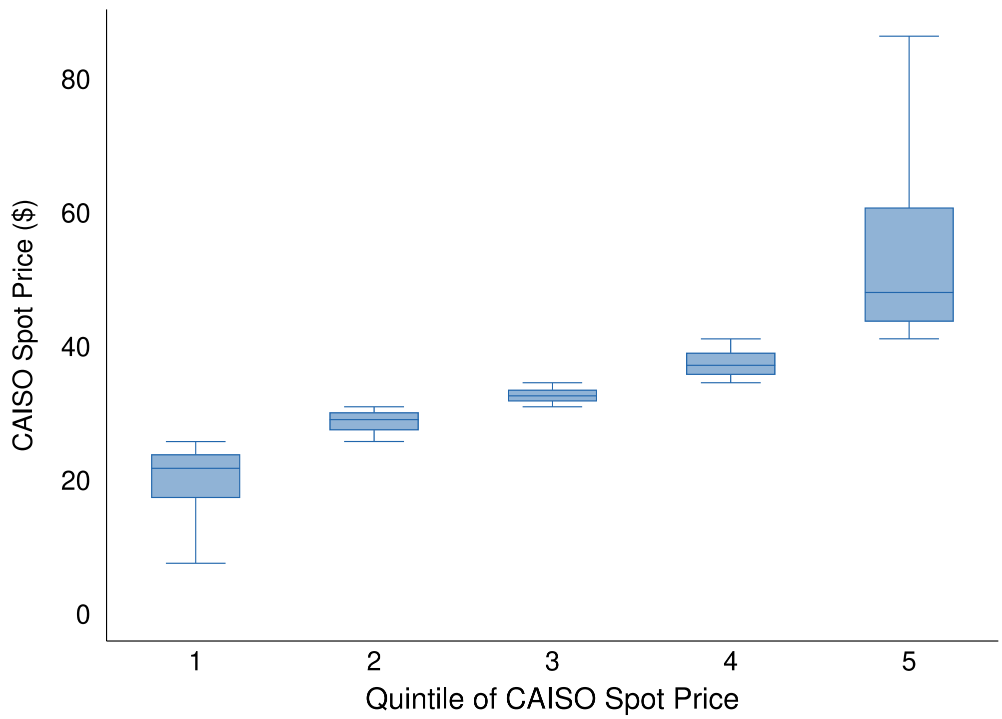
C External Data

Table 3: Daily Outdoor Weather Summary Statistics

Variable	Daily Measure	Mean	Std. Dev.	Between County Std. Dev.	Within County Std. Dev.	Min.	Max
Panel A: N. CA Experiment							
Temperature ($°F$)	Mean	61.54	11.53	1.23	11.48	33.29	91.25
	Minimum	49.19	9.88	0.44	9.88	21.00	76.00
	Maximum	75.78	14.35	1.70	14.27	43.00	108.00
Relative Humidity (%)	Mean	60.02	15.43	2.90	15.22	10.54	97.53
	Minimum	33.79	16.68	2.00	16.59	3.00	93.00
	Maximum	84.44	11.92	2.29	11.75	14.00	100.00
Heat Index ($°F$)	Mean	60.61	11.57	1.16	11.53	31.62	90.30
	Minimum	48.41	10.25	0.45	10.24	18.97	76.01
	Maximum	74.21	13.66	1.54	13.59	40.82	108.61
N				4			
$N \times T$				2,060			
Panel B: C. CA Experiment							
Temperature ($°F$)	Mean	66.58	14.36	2.56	14.20	32.63	96.04
	Minimum	54.20	12.91	3.66	12.56	19.00	85.00
	Maximum	79.85	16.05	1.17	16.02	45.00	110.00
Relative Humidity (%)	Mean	51.09	17.03	6.58	16.16	13.33	96.78
	Minimum	28.93	15.81	1.76	15.75	2.00	90.00
	Maximum	73.12	16.89	10.16	14.71	22.00	100.00
Heat Index ($°F$)	Mean	65.42	14.18	2.43	14.04	30.62	95.68
	Minimum	53.41	13.28	3.61	12.95	16.87	83.66
	Maximum	77.85	15.11	1.08	15.08	43.20	109.53
N				3			
$N \times T$				1,545			
Panel C: Both Experiments							
Temperature ($°F$)	Mean	63.70	13.06	3.20	12.71	32.63	96.04
	Minimum	51.34	11.55	3.43	11.10	19.00	85.00
	Maximum	77.52	15.23	2.58	15.05	43.00	110.00
Relative Humidity (%)	Mean	56.19	16.73	6.44	15.63	10.54	97.53
	Minimum	31.70	16.49	3.13	16.23	2.00	93.00
	Maximum	79.58	15.32	8.58	13.10	14.00	100.00
Heat Index ($°F$)	Mean	62.67	12.97	3.04	12.66	30.62	95.68
	Minimum	50.56	11.90	3.40	11.48	16.87	83.66
	Maximum	75.77	14.41	2.32	14.25	40.82	109.53
N				7			
$N \times T$				3,605			

Note: This table presents summary statistics of daily weather conditions in each of the two experiments and over both experiments.

Figure 16: Box and Whisker Plots of CAISO Spot Price by Quintile of Price

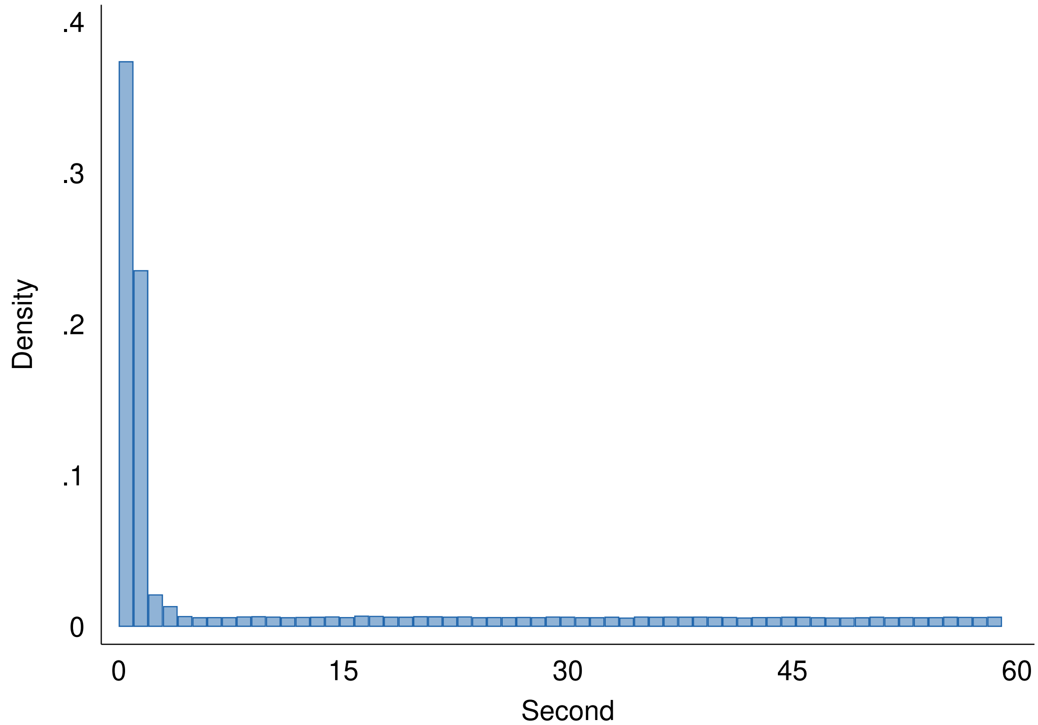


Note: This figure presents a box and whisker plot of spot prices from the California ISO (CAISO) by quintile of the spot prices. The horizontal lines in the shaded boxes represent the median, the ends of the boxes indicate the first and third quartiles, and the ends of the whiskers denote the upper and lower adjacent values.

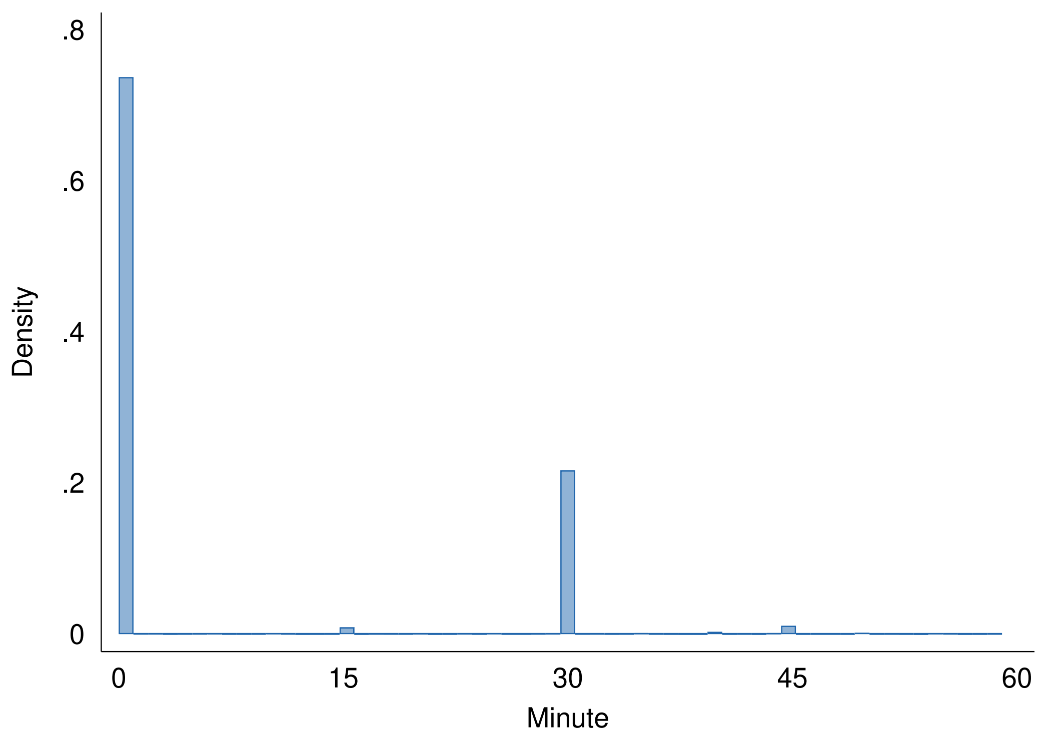
D Internal Data

The HVAC system events data does not label temperature changes as being the result of a permanent setpoint or temporary override. We infer this information based on the precise timing of when the change occurs. Appendix Figure 17 informs our approach to this classification. Panel (a) plots the density of the second of the minute at which temperature changes take place. The density is roughly uniform with a probability of about 0.70 across all seconds, save for a large increase in the probability of changes occurring at :00 through :02 (and to a lesser extent :03) seconds of the minute. Since we would expect temporary overrides to occur uniformly across seconds of the minute, we code temperature changes occurring at less than :03 seconds of the minute as permanent setpoints and all other temperature changes as temporary overrides. Panel (b) plots the density of permanent setpoints (as determined by our classification rule) by minute of the hour. Consistent with our priors, users schedule most setpoints on the hour or half hour (and to a lesser extent, at :15 and :45 minutes past the hour). This is both a finding and a confirmation of the validity of our approach to classifying setpoints and overrides.

Figure 17: Timing of HVAC System Events

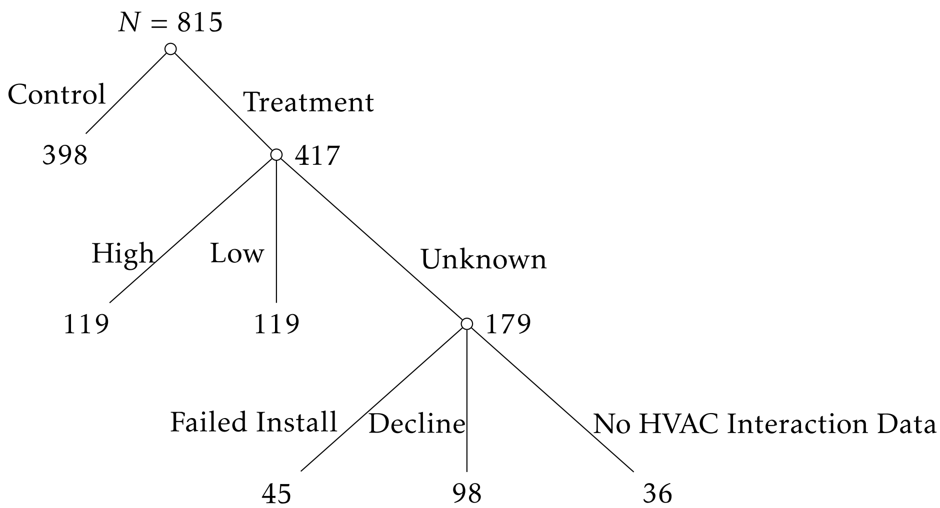


(a) Density of Temperature Changes by Second of the Minute



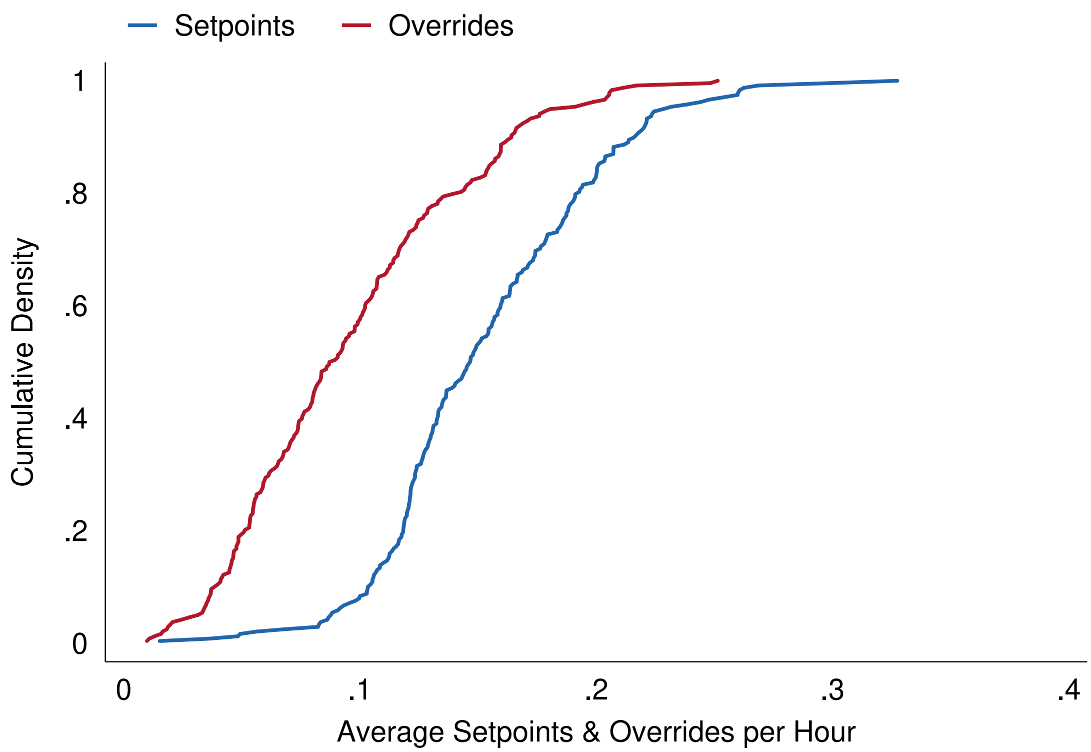
(b) Density of Permanent Setpoints by Minute of the Hour

Figure 18: Modified Sample Randomization with Energy-Efficiency Types (N. CA Experiment)



Note: This figure presents the number of households in treatment and control for each experiment, followed by counts of households classified as high-, low-, or unknown-efficiency types based on a definition that divides types at the median. Finally, installations of the smart thermostat, failed installations, and declined installations are also reported.

Figure 19: Distributions of Permanent Setpoints and Temporary Overrides



Note: This figure presents the empirical cumulative distribution of the average number of hourly setpoints and overrides by household.

Table 4: User Interactions Summary Statistics by Experiment

	Northern California			Central California			Both Experiments		
Variables	Mean	Std.Dev.	Obs.	Mean	Std.Dev.	Obs	Mean	Std.Dev.	Obs.
Ambient Temp.	69.08	5.20	320,601	66.88	4.36	25,598	68.92	5.18	346,199
Cooling Setpoints	78.91	4.10	54,164	78.22	4.52	3,033	78.87	4.12	57,197
Heating Setpoints	63.88	5.60	73,359	64.64	5.44	5,806	63.93	5.59	79,165
Cooling Overrides	77.51	3.87	14,414	77.56	4.99	1,208	77.51	3.97	15,622
Heating Overrides	67.37	4.14	41,497	68.16	4.58	6,538	67.48	4.21	48,035
N		238			134			372	
$N \times T$		357,392			28,742			386,134	

Note: This table presents summary statistics of the setpoints and overrides for the treatment group households for which setpoints and overrides were observable.

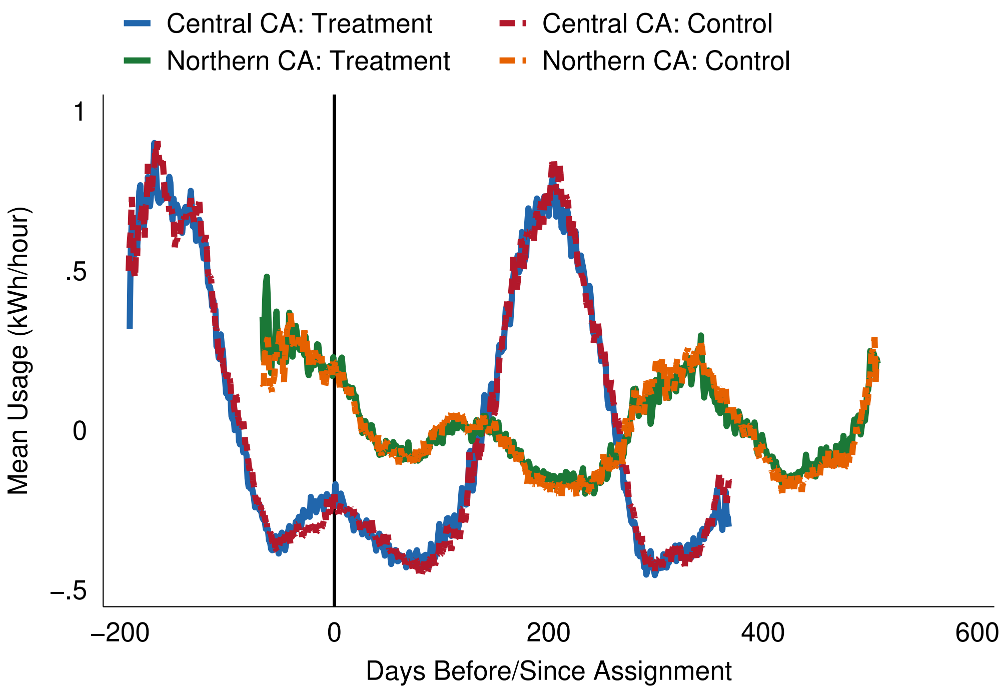
E Balance and Time-Trend Analysis

Table 5: Balance Table

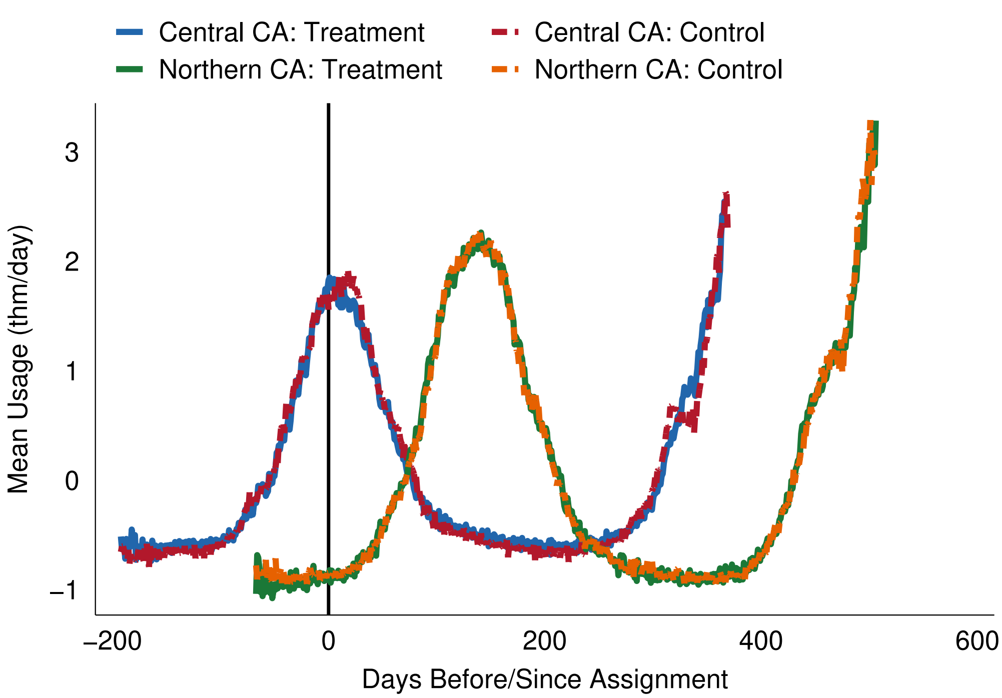
Variable	(1) Both Treatment Indicator	(2) N. CA Treatment Indicator	(3) C. CA Treatment Indicator
Household Characteristics			
Family in the Household Indicator	0.025	-0.028	0.083
	(0.053)	(0.071)	(0.080)
Pets in the Household Indicator	0.013	0.019	0.005
	(0.029)	(0.038)	(0.045)
HER Experiment Indicator	-0.021	-0.004	-0.046
	(0.031)	(0.040)	(0.048)
HER Recipient Indicator	-0.006	0.027	-0.062
	(0.039)	(0.049)	(0.063)
Home Characteristics			
Multi-Family Home Indicator	-0.017	-0.022	0.039
	(0.080)	(0.091)	(0.166)
Year Home Built (Year / 1,000)	0.23	-0.589	1.363
	(0.800)	(1.110)	(1.170)
Size of Home (Sq. Ft. / 10,000)	0.286	0.377	-0.061
	(0.246)	(0.324)	(0.433)
Pool Indicator	-0.006	0.037	-0.082
	(0.033)	(0.044)	(0.052)
Electric Heat Indicator	0.014	-0.068	0.126
	(0.094)	(0.125)	(0.140)
Pre-Period Energy Use			
Mean (kWh)	-0.036	-0.054	0.010
	(0.028)	(0.034)	(0.048)
Mean (thm)	-0.029	0.002	-0.054
	(0.032)	(0.050)	(0.040)
N	1,385	821	564
R^2	0.011	0.015	0.021
F	0.664	0.689	0.799

Notes: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The table reports linear probability model estimates of the probability of assignment to treatment. The *HER Experiment Indicator* variable is equal to one for households that participated in the Home Energy Report experiment, and the *HER Recipient Indicator* variable is equal to one for households that were assigned to the treatment group in that experiment. We interpolate missing values of continuous variables (year built, home size, and pre-period energy use). We also code as zero and include an indicator for missing values of binary variables (heating type) and mismeasured values of pre-period electricity means in Northern California that are based on less than two weeks of data (see Section 2.4). Models include indicators for month and county of recruitment, as well as the aforementioned indicators for missing/mismeasured values. All omitted coefficient estimates are statistically insignificant. The F-statistic tests the null hypothesis that all parameters are jointly equal to zero. We fail to reject the null in all three models.

Figure 20: Average Residual Energy Use by Experimental Status and Wave



(a) Electricity



(b) Natural Gas

Note: Subfigure (a) plots average hourly electricity use (kWh/hour) by day, and Subfigure (b) plots mean daily natural gas consumption (thm/day) by day. Means are calculated after projecting out household fixed effects. Only means based on 30 or more homes per day are included in the figures.

F Discussion of Identifying Assumptions

Our DDIV empirical specification identifies the ATT of a smart thermostat if our experimental instrument is relevant and valid, there are common or parallel trends, monotonicity holds, and there is one-sided experimental noncompliance.[57] We provide evidence that each of these assumptions is reasonable in this section. First, instrument relevance requires that assignment to treatment affects the probability that a household installs a smart thermostat. We report the first-stage F statistics with all of our results tables. As one would expect of a field experiment, we always easily reject the null of weak instruments.

Second, the instrument validity assumption in a DDIV model can be thought of as two separate conditions (Hudson, Hull and Liebersohn, 2017). The first is the traditional IV assumption that the instrument is exogenous and the only way assignment to the treatment group affects energy use is through the installation of a smart thermostat. The second is the assumption implicit in all DD analyses that post-period randomization does not affect the pre-period values of outcomes (energy use) or treatment (smart thermostat installation). Both assumptions are satisfied by the nature of our experiment: households are randomly assigned to a treatment or control group. Assignment occurs both (shortly) after the household first interacts with the experimenter and after the household's pre-period energy use decisions have been made. The analyses in Section 2.6 and 2.7 are consistent with an appropriate randomization process.

Third, the common or parallel trends assumption requires that the unobserved, counterfactual trend in energy use that would have been experienced by the treated group is parallel to the observable, untreated trend in the comparison group. In the context of our experiment, this means that the energy consumed by control group households is a good proxy for the energy homes who installed a smart thermostat would have used in a counterfactual world without a smart thermostat. While this assumption is fundamentally untestable because of the counterfactual outcomes problem, it is satisfied if there is appropriate randomization (Hudson, Hull and Liebersohn, 2017). Nonetheless, we provide additional support for this assumption by showing evidence of parallel pre-trends via the event studies in Section 2.7.

Finally, if there is two-sided noncompliance in an experiment, the estimates are confounded by substitution bias (Heckman and Smith, 1995). The standard in the literature is to relax the noncompliance assumption to one of monotonicity (or uniformity). In our case, this means that the experimental treatment makes all households in more (or less) likely to get a smart thermostat than they would have been otherwise. Under this alternative assumption, the DDIV specification recovers the LATE estimate of γ^j (Imbens and Angrist, 1994). This is an estimate of the average impact of a smart thermostat on the energy consumption of households that were induced to install one by our experiment.

[57] Alternatively, we can recover the ITT estimate of γ^j by replacing the S_i in Equation 1 with T_i. This is an estimate of the average effect of being randomized into the treatment group in our experiment. We estimate a baseline DDITT model in Section 5.5.

Our experimental environment allows us to make the stronger assumption that there is one-sided experimental noncompliance that allows us to identify the ATT of a smart thermostat. The assumption of one-sided noncompliance is tenuous to the extent that "the need for treatment under question is widely acknowledged and there is competition over implementation" (Ito, 2007). This is not the case in our context as smart thermostat technology was in its infancy at the time of our study. As noted in the main text, using data from the EIA's 2015 Residential Energy Consumption Survey (RECS), we find that only 4.09% of all households in the survey and 10.58% of households observationally similar to those in our study own a smart thermostat several years after our experiment.[58] Additionally, while we are unable to directly observe whether any households in the control group upgrade their thermostat, we never observe control households using a smart thermostat on Opower platform. Thus, the available evidence supports the validity of the one-sided noncompliance assumption in our experimental context.

[58]The RECS is not conducted annually, so we use data from the 2015 survey as it is the closest possible survey iteration subsequent to the time period observed in our data. The previous iteration of the survey in 2009 did not ask questions about smart devices. We define "observationally similar" households by restricting the RECS sample to homes that would pass Opower's initial eligibility screening to join the trial (to the extent possible given the measures available). Specifically, we condition on owner-occupied, single-family homes located in the Pacific Division (state of residence is not observed) that have a functioning central furnace or heat pump, central air conditioning, and an electrical connection. We are not able to condition on whether or not the household has a high-speed Internet connection or whether the occupants plan to move in the next year, as those questions are not part of the RECS survey.

G Main Results with Full Regression Diagnostics

Table 6: N. CA Experiment-ATT Estimates of the Effect of a Smart Thermostat on Energy Use

	(1)	(2)	(3)	(4)	(5)	(6)
	Power Use (kWh/hour or thm/day)					
Panel A: Electricity (kWh)						
ATT ($\hat{\gamma}^{kWh}$)	-0.055	-0.061	-0.016	-0.016	-0.016	-0.014
	(0.058)	(0.058)	(0.046)	(0.046)	(0.046)	(0.046)
Constant	1.294***	2.553***				
	(0.035)	(0.072)				
N	815	815	815	815	815	815
$N \times T$	9,729,849	9,729,849	9,729,849	9,729,849	9,729,849	9,729,849
F statistic	44.591	268.622	343.954	353.677	350.446	9.404
rk LM statistic	391.219	391.265	313.225	313.190	313.190	313.304
rk Wald F statistic	379.956	380.004	670.870	670.765	670.765	669.978
Panel B: Natural Gas (thm)						
ATT ($\hat{\gamma}^{thm}$)	-0.009	0.009	0.085	0.075	0.075	0.067
	(0.061)	(0.063)	(0.068)	(0.066)	(0.066)	(0.065)
Constant	0.523***	17.735***				
	(0.020)	(0.304)				
N	805	805	805	805	805	805
$N \times T$	398,243	398,243	398,243	398,243	398,243	398,243
F statistic	801.768	570.374	676.791	520.520	521.407	18.587
rk LM statistic	386.783	386.896	313.867	313.885	313.886	314.109
rk Wald F statistic	377.042	377.092	672.580	672.617	672.608	672.169
Weather Controls		x	x	x	x	x
HH Fixed Effects			x	x	x	x
Month-of-Year Effects				x	x	
Day-of-Week Effects					x	
Day-by-Hour or Day Effects						x

Note: Standard errors clustered at the household level in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. All estimates are based on a sample comprised of the Northern California experiment. The sample used to produce the estimates in Panel A is based on *hourly* electricity meter readings in kWh, while the sample underlying the estimates in Panel B is based on *daily* natural gas meter readings (thm). Thus, the day-by-hour effects noted in Column (6) are included in the electricity model (Panel A) only and are day effects in the natural gas model (Panel B). Based on the values of the rk LM and Wald F statistics, we reject the nulls of an under or weakly identified model across all specifications.

Table 7: C. CA Experiment-ATT Estimates of the Effect of a Smart Thermostat on Energy Use

	(1)	(2)	(3)	(4)	(5)	(6)
	Power Use (kWh/hour or thm/day)					
Panel A: Electricity (kWh)						
ATT ($\hat{\gamma}^{kWh}$)	0.009	0.005	0.002	0.002	0.002	0.003
	(0.029)	(0.028)	(0.025)	(0.025)	(0.025)	(0.025)
Constant	1.292***	3.182***				
	(0.030)	(0.091)				
N	564	564	564	564	564	564
$N \times T$	6,691,885	6,691,885	6,691,885	6,691,885	6,691,885	6,691,885
F statistic	49.321	413.519	541.873	393.169	389.966	10.470
rk LM statistic	394.996	395.009	384.992	384.985	384.985	385.006
rk Wald F statistic	677.494	677.450	1,352.535	1,352.619	1,352.618	1,350.492
Panel B: Natural Gas (thm)						
ATT ($\hat{\gamma}^{thm}$)	-0.003	0.004	-0.002	-0.001	-0.001	0.002
	(0.044)	(0.030)	(0.026)	(0.026)	(0.026)	(0.025)
Constant	1.101***	12.038***				
	(0.034)	(0.235)				
N	564	564	564	564	564	564
$N \times T$	279,061	279,061	279,061	279,061	279,061	279,061
F statistic	3.488	383.889	444.774	366.746	368.392	18.081
rk LM statistic	393.909	393.941	390.413	390.402	390.401	390.438
rk Wald F statistic	675.636	675.284	1,376.659	1,376.575	1,376.545	1,374.416
Weather Controls		x	x	x	x	x
HH Fixed Effects			x	x	x	x
Month-of-Year Effects				x	x	
Day-of-Week Effects					x	
Day-by-Hour or Day Effects						x

Note: Standard errors clustered at the household level in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. All estimates are based on a sample comprised of the Central California experiment. The sample used to produce the estimates in Panel A is based on *hourly* electricity meter readings in kWh, while the sample underlying the estimates in Panel B is based on *daily* natural gas meter readings (thm). Thus, the day-by-hour effects noted in Column (6) are included in the electricity model (Panel A) only and are day effects in the natural gas model (Panel B). Based on the values of the rk LM and Wald F statistics, we reject the nulls of an under or weakly identified model across all specifications.

Table 8: Both Experiments-ATT Estimates of the Effect of a Smart Thermostat on Energy Use (Single-Experiment Specification)

	(1)	(2)	(3)	(4)	(5)	(6)
	Power Use (kWh/hour or thm/day)					
Panel A: Electricity (kWh)						
ATT ($\hat{\gamma}^{kWh}$)	-0.031	-0.031	-0.003	-0.001	-0.001	-0.001
	(0.036)	(0.035)	(0.022)	(0.022)	(0.022)	(0.022)
Constant	1.293***	2.897***				
	(0.024)	(0.060)				
N	1,379	1,379	1,379	1,379	1,379	1,379
$N \times T$	16,421,734	16,421,734	16,421,734	16,421,734	16,421,734	16,421,734
F statistic	67.704	533.991	819.428	752.614	747.155	197.309
rk LM statistic	738.263	749.373	611.957	612.274	612.274	612.841
rk Wald F statistic	790.294	819.436	1,948.372	1,951.621	1,951.626	1,955.165
Panel B: Natural Gas (thm)						
ATT ($\hat{\gamma}^{thm}$)	0.062	0.061	0.024	0.021	0.021	0.022
	(0.060)	(0.049)	(0.027)	(0.026)	(0.026)	(0.026)
Constant	0.963***	15.295***				
	(0.028)	(0.222)				
N	1,369	1,369	1,369	1,369	1,369	1,369
$N \times T$	677,304	677,304	677,304	677,304	677,304	677,304
F statistic	126.946	742.877	998.183	837.432	838.243	82.848
rk LM statistic	733.785	744.066	618.740	619.142	619.144	619.783
rk Wald F statistic	790.386	817.151	1,976.088	1,980.001	1,979.994	1,983.556
Wave Indicator		x				
Weather Controls		x	x	x	x	x
HH Fixed Effects			x	x	x	x
Month-of-Year Effects				x	x	
Day-of-Week Effects					x	
Day-by-Hour or Day Effects						x

Note: Standard errors clustered at the household level in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. All estimates are based on a sample comprised of both experiments and a "single-experiment specification" that models the Northern and Central California samples as two waves of the same experiment. In practice, this means we include minimal controls for differences between the two: the estimates reported in Column (2) are based on a model that includes an indicator for the Northern California experiment. This indicator is perfectly co-linear with household fixed effects, so it is dropped from subsequent models. The sample used to produce the estimates in Panel A is based on *hourly* electricity meter readings in kWh, while the sample underlying the estimates in Panel B is based on *daily* natural gas meter readings (thm). Thus, the day-by-hour effects noted in Column (6) are included in the electricity model (Panel A) only and are day effects in the natural gas model (Panel B). Based on the values of the rk LM and Wald F statistics, we reject the nulls of an under or weakly identified model across all specifications.

Table 9: Both Experiments-ATT Estimates of the Effect of a Smart Thermostat on Energy Use (Multiple-Experiments Specification)

	(1)	(2)	(3)	(4)	(5)	(6)
	Power Use (kWh/hour or thm/day)					
Panel A: Electricity (kWh)						
ATT ($\hat{\gamma}^{kWh}$)	-0.008	-0.012	-0.003	-0.002	-0.002	-0.002
	(0.026)	(0.026)	(0.022)	(0.022)	(0.022)	(0.022)
Constant	1.288***	2.893***				
	(0.029)	(0.062)				
N	1,379	1,379	1,379	1,379	1,379	1,379
$N \times T$	16,421,734	16,421,734	16,421,734	16,421,734	16,421,734	16,421,734
F statistic	49.480	429.236	703.033	646.298	641.608	169.297
rk LM statistic	394.698	394.731	614.072	614.075	614.075	614.108
rk Wald F statistic	257.733	257.751	1,963.342	1,963.304	1,963.304	1,961.886
Panel B: Natural Gas (thm)						
ATT ($\hat{\gamma}^{thm}$)	-0.005	0.006	0.021	0.021	0.021	0.023
	(0.036)	(0.028)	(0.026)	(0.026)	(0.026)	(0.026)
Constant	1.100***	15.266***				
	(0.033)	(0.221)				
N	1,369	1,369	1,369	1,369	1,369	1,369
$N \times T$	677,304	677,304	677,304	677,304	677,304	677,304
F statistic	485.066	603.533	870.502	717.820	718.528	70.861
rk LM statistic	390.328	390.412	620.948	620.954	620.954	621.015
rk Wald F statistic	255.868	255.899	1,991.526	1,991.480	1,991.463	1,990.180
Wave Indicator & Interactions		x	x	x	x	x
Weather Controls		x	x	x	x	x
HH Fixed Effects			x	x	x	x
Month-of-Year Effects				x	x	
Day-of-Week Effects					x	
Day-by-Hour or Day Effects						x

Note: Standard errors clustered at the household level in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. All estimates are based on a sample comprised of both experiments. Since the "single-experiment specification" reported in Appendix Table 8 does not guarantee that the combined result is a convex combination of the Northern and Central California estimates reported in Appendix Tables 6 and 7, this "multiple-experiments specification" includes an indicator for the Northern California experiment and interacts it with the relevant DDIV control variables in each column. The sample used to produce the estimates in Panel A is based on *hourly* electricity meter readings in kWh, while the sample underlying the estimates in Panel B is based on *daily* natural gas meter readings (thm). Thus, the day-by-hour effects noted in Column (6) are included in the electricity model (Panel A) only and are day effects in the natural gas model (Panel B). Based on the values of the rk LM and Wald F statistics, we reject the nulls of an under or weakly identified model across all specifications.

H Heterogeneous Treatment Effects Estimates

H.1 Ambient Weather Estimates

Table 10: ATT Estimates of the Effect of a Smart Thermostat on Energy Use by Ambient Temperature Quintile

	(1)	(2)	(3)	(4)	(5)
	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
	Power Use (kWh/hour or thm/day)				
Panel A: Electricity (kWh)					
ATT ($\hat{\gamma}^{kWh}$)	-0.036	-0.033*	-0.024	-0.008	0.009
	(0.022)	(0.019)	(0.019)	(0.024)	(0.044)
N	1,376	1,379	1,379	1,379	1,378
$N \times T$	3,344,875	3,541,064	3,239,699	3,102,224	3,193,872
F statistic	1.522	1.610	14.479	16.745	24.597
rk LM statistic	368.074	652.296	681.451	600.120	545.434
rk Wald F statistic	1,379.306	1,920.331	1,966.982	1,879.175	1,769.185
Panel B: Natural Gas (thm)					
ATT ($\hat{\gamma}^{thm}$)	-0.054	-0.013	0.005	-0.008	0.010
	(0.064)	(0.038)	(0.023)	(0.018)	(0.015)
N	1,364	1,366	1,369	1,368	1,365
$N \times T$	145,525	147,327	120,200	138,512	125,737
F statistic	22.958	0.567	6.374	6.145	0.431
rk LM statistic	360.657	435.244	563.231	699.424	403.356
rk Wald F statistic	1,375.353	1,587.270	1,323.571	1,802.507	1,377.126
HH Fixed Effects	x	x	x	x	x
Month-of-Year Effects	x	x	x	x	x
Day-of-Week Effects	x	x	x	x	x

Note: Standard errors clustered at the household level in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. All estimates are based on a sample comprised of both experiments. The sample used to produce the estimates in Panel A is based on *hourly* electricity meter readings in kWh, and temperature quintiles are calculated from the distribution of *hourly* average ambient temperature readings. The sample underlying the estimates in Panel B is based on *daily* natural gas meter readings (thm), and temperature quintiles are calculated using the distribution of *daily* average ambient temperature readings. Based on the values of the rk *LM* and Wald *F* statistics, we reject the nulls of an under or weakly identified model across all specifications.

Table 11: ATT Estimates of the Effect of a Smart Thermostat on Energy Use by Ambient Humidity Quintile

	(1) Quintile 1	(2) Quintile 2	(3) Quintile 3	(4) Quintile 4	(5) Quintile 5
	Power Use (kWh/hour or thm/day)				
Panel A: Electricity (kWh)					
ATT ($\hat{\gamma}^{kWh}$)	0.050	-0.010	-0.020	-0.039**	-0.068***
	(0.048)	(0.024)	(0.019)	(0.018)	(0.021)
N	1,379	1,379	1,379	1,379	1,379
$N \times T$	3,309,175	3,331,392	3,263,471	3,256,818	3,260,878
F statistic	46.213	2.648	9.101	3.183	7.390
rk LM statistic	521.622	564.638	596.226	632.753	623.567
rk Wald F statistic	1,762.186	1,859.612	1,914.506	1,926.931	1,607.362
Panel B: Natural Gas (thm)					
ATT ($\hat{\gamma}^{thm}$)	0.004	-0.010	-0.007	0.041	-0.014
	(0.017)	(0.025)	(0.035)	(0.044)	(0.068)
N	1,367	1,369	1,369	1,369	1,367
$N \times T$	141,016	133,650	133,026	153,619	115,991
F statistic	0.930	0.188	0.076	26.762	53.950
rk LM statistic	380.444	564.518	644.051	612.586	545.259
rk Wald F statistic	1,356.189	1,740.682	1,907.282	1,530.025	1,288.742
HH Fixed Effects	x	x	x	x	x
Month-of-Year Effects	x	x	x	x	x
Day-of-Week Effects	x	x	x	x	x

Note: Standard errors clustered at the household level in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. All estimates are based on a sample comprised of both experiments. The sample used to produce the estimates in Panel A is based on *hourly* electricity meter readings in kWh, and humidity quintiles are calculated from the distribution of *hourly* average ambient relative humidity readings. The sample underlying the estimates in Panel B is based on *daily* natural gas meter readings (thm), and humidity quintiles are calculated using the distribution of *daily* average ambient relative humidity readings. Based on the values of the rk LM and Wald F statistics, we reject the nulls of an under or weakly identified model across all specifications.

Table 12: ATT Estimates of the Effect of a Smart Thermostat on Energy Use by Ambient Heat Index Quintile

	(1) Quintile 1	(2) Quintile 2	(3) Quintile 3	(4) Quintile 4	(5) Quintile 5
	Power Use (kWh/hour or thm/day)				
Panel A: Electricity (kWh)					
ATT ($\hat{\gamma}^{kWh}$)	-0.036	-0.030	-0.026	-0.009	0.009
	(0.022)	(0.019)	(0.019)	(0.024)	(0.043)
N	1,376	1,379	1,379	1,379	1,378
$N \times T$	3,295,672	3,273,443	3,296,244	3,273,250	3,283,125
F statistic	1.483	1.632	13.840	17.541	24.681
rk LM statistic	367.461	636.474	691.261	604.527	546.840
rk Wald F statistic	1,380.494	1,927.495	1,955.416	1,883.346	1,770.576
Panel B: Natural Gas (thm)					
ATT ($\hat{\gamma}^{thm}$)	-0.060	-0.004	-0.004	-0.003	0.009
	(0.066)	(0.044)	(0.024)	(0.018)	(0.015)
N	1,364	1,366	1,369	1,367	1,365
$N \times T$	135,502	136,288	134,989	135,319	135,202
F statistic	18.708	6.570	10.857	12.691	0.289
rk LM statistic	351.296	404.357	586.163	702.831	413.812
rk Wald F statistic	1,364.503	1,468.623	1,403.566	1,797.217	1,406.921
HH Fixed Effects	x	x	x	x	x
Month-of-Year Effects	x	x	x	x	x
Day-of-Week Effects	x	x	x	x	x

Note: Standard errors clustered at the household level in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. All estimates are based on a sample comprised of both experiments. The sample used to produce the estimates in Panel A is based on *hourly* electricity meter readings in kWh, and heat index quintiles are calculated from the distribution of *hourly* average ambient heat index readings. The heat index is calculated using temperature and humidity readings. See https://www.wpc.ncep.noaa.gov/html/heatindex_equation.shtml for the exact formula. The sample underlying the estimates in Panel B is based on *daily* natural gas meter readings (thm), and heat index quintiles are calculated using the distribution of *daily* average ambient heat index readings. Based on the values of the rk *LM* and Wald *F* statistics, we reject the nulls of an under or weakly identified model across all specifications.

H.2 Day of Week and Hour of the Day Estimates

Table 13: ATT Estimates of the Effect of a Smart Thermostat on Energy Use by Day of the Week and Weekday/Weekend

	(1) Sunday	(2) Monday	(3) Tuesday	(4) Wednesday	(5) Thursday	(6) Friday	(7) Saturday	(8) Weekday	(9) Weekend
	Power Use (kWh/hour or thm/day)								
Panel A: Electricity (kWh)									
ATT ($\hat{\gamma}^{kWh}$)	-0.006	-0.014	-0.007	-0.001	0.005	0.005	0.010	-0.002	0.002
	(0.024)	(0.023)	(0.024)	(0.023)	(0.023)	(0.023)	(0.023)	(0.022)	(0.023)
N	1,379	1,379	1,379	1,379	1,379	1,379	1,379	1,379	1,379
$N \times T$	2,338,599	2,331,710	2,331,777	2,331,619	2,362,409	2,362,700	2,362,920	11,720,215	4,701,519
F statistic	683.309	610.092	584.520	553.404	694.530	554.755	684.794	698.737	731.618
rk LM statistic	610.469	605.937	604.237	600.642	624.896	620.751	616.091	611.775	613.369
rk Wald F statistic	1,946.898	1,941.098	1,936.470	1,933.097	1,972.931	1,966.296	1,952.716	1,951.947	1,950.206
Panel B: Natural Gas (thm)									
ATT ($\hat{\gamma}^{thm}$)	0.024	0.005	0.026	0.024	0.018	0.017	0.033	0.018	0.028
	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)	(0.030)	(0.031)	(0.026)	(0.027)
N	1,369	1,369	1,369	1,369	1,369	1,369	1,369	1,369	1,369
$N \times T$	94,575	96,480	96,480	96,474	97,760	97,764	97,771	484,958	192,346
F statistic	648.558	645.340	638.998	520.106	661.750	590.074	592.727	813.967	738.649
rk LM statistic	622.274	610.989	609.909	606.782	630.704	626.905	623.224	617.583	622.893
rk Wald F statistic	1,981.323	1,965.567	1,959.042	1,960.855	1,998.501	1,993.507	1,984.444	1,977.988	1,983.797
Weather Controls	x	x	x	x	x	x	x	x	x
HH Fixed Effects	x	x	x	x	x	x	x	x	x
Month-of-Year Effects	x	x	x	x	x	x	x	x	x

Note: Standard errors clustered at the household level in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. All estimates are based on a sample comprised of both experiments. The sample used to produce the estimates in Panel A is based on *hourly* electricity meter readings in kWh, while the sample underlying the estimates in Panel B is based on *daily* natural gas meter readings (thm). Based on the values of the rk LM and Wald F statistics, we reject the nulls of an under or weakly identified model across all specifications.

Table 14: ATT Estimates of the Effect of a Smart Thermostat on Electricity Use by Hour of the Day

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	12:00	1:00	2:00	3:00	4:00	5:00	6:00	7:00	8:00	9:00	10:00	11:00
	Power Use (kWh/hour)											
Panel A: AM												
ATT ($\hat{\gamma}^{kWh}$)	-0.022	-0.013	-0.020	-0.030	-0.015	0.009	0.003	-0.003	0.005	-0.029	-0.042	-0.042
	(0.027)	(0.023)	(0.021)	(0.021)	(0.021)	(0.023)	(0.024)	(0.027)	(0.030)	(0.036)	(0.039)	(0.042)
N	1,379	1,379	1,379	1,379	1,379	1,379	1,379	1,379	1,379	1,379	1,379	1,379
$N \times T$	684,283	684,283	682,930	684,283	684,283	684,283	684,283	684,283	684,284	684,290	684,291	684,295
R^2	0.073	0.072	0.065	0.055	0.046	0.037	0.025	0.021	0.020	0.024	0.034	0.050
F statistic	271.841	252.500	218.651	202.622	183.757	173.456	155.631	176.782	179.429	206.557	254.895	306.567
rk LM statistic	614.268	614.232	614.352	614.224	614.218	614.173	614.185	614.191	614.150	614.088	613.917	613.624
rk Wald F statistic	1,956.218	1,955.971	1,956.107	1,956.049	1,956.053	1,955.797	1,955.944	1,955.917	1,955.730	1,955.705	1,955.316	1,955.010
Panel B: PM												
ATT ($\hat{\gamma}^{kWh}$)	-0.028	-0.005	0.015	0.021	0.046	0.076*	0.048	0.035	-0.004	-0.032	-0.022	-0.019
	(0.045)	(0.047)	(0.048)	(0.047)	(0.045)	(0.042)	(0.039)	(0.036)	(0.034)	(0.032)	(0.031)	(0.027)
N	1,379	1,379	1,379	1,379	1,379	1,379	1,379	1,379	1,379	1,379	1,379	1,379
$N \times T$	684,296	684,299	684,301	684,304	684,307	684,308	684,308	684,308	684,308	684,308	684,308	684,308
R^2	0.071	0.093	0.113	0.132	0.138	0.143	0.142	0.132	0.121	0.106	0.093	0.081
F statistic	384.005	452.878	526.584	614.748	663.529	700.821	748.612	680.700	617.432	548.780	409.439	309.301
rk LM statistic	612.715	611.981	611.317	610.723	610.202	609.945	609.791	609.666	609.584	609.590	609.615	609.612
rk Wald F statistic	1,952.823	1,950.652	1,949.532	1,948.435	1,947.214	1,946.442	1,946.577	1,946.179	1,946.103	1,946.453	1,947.011	1,947.112
Weather Controls	x	x	x	x	x	x	x	x	x	x	x	x
HH Fixed Effects	x	x	x	x	x	x	x	x	x	x	x	x
Month-of-Year Effects	x	x	x	x	x	x	x	x	x	x	x	x
Day-of-Week Effects	x	x	x	x	x	x	x	x	x	x	x	x

Note: Standard errors clustered at the household level in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. All estimates are based on a sample comprised of both experiments. The sample used to produce the estimates is based on *hourly* electricity meter readings in kWh. Panel A reports estimates from models that are conditional on the given hour in the AM. Panel B reports analogous estimates based on the given hour in the PM. Based on the values of the rk LM and Wald F statistics, we reject the nulls of an under or weakly identified model across all specifications.

Table 15: ATT Estimates of the Effect of a Smart Thermostat on Energy Use by Hour of the Day and Weekday/Weekend

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	12:00	1:00	2:00	3:00	4:00	5:00	6:00	7:00	8:00	9:00	10:00	11:00
	Power Use (kWh/hour)											
Panel A: Weekdays Only												
AM: ATT ($\hat{\gamma}^{kWh}$)	-0.027	-0.016	-0.025	-0.036*	-0.017	0.011	0.000	-0.010	0.003	-0.032	-0.040	-0.042
	(0.028)	(0.024)	(0.022)	(0.021)	(0.021)	(0.023)	(0.025)	(0.027)	(0.030)	(0.036)	(0.040)	(0.042)
PM: ATT ($\hat{\gamma}^{kWh}$)	-0.027	-0.009	0.010	0.017	0.051	0.085*	0.053	0.036	-0.007	-0.035	-0.022	-0.027
	(0.045)	(0.048)	(0.049)	(0.050)	(0.047)	(0.044)	(0.041)	(0.038)	(0.035)	(0.033)	(0.032)	(0.028)
N	1,379	1,379	1,379	1,379	1,379	1,379	1,379	1,379	1,379	1,379	1,379	1,379
$N \times T$	~488,000	~488,000	~488,000	~488,000	~488,000	~488,000	~488,000	~488,000	~488,000	~488,000	~488,000	~488,000
Panel B: Weekend Days Only												
AM: ATT ($\hat{\gamma}^{kWh}$)	-0.010	-0.006	-0.007	-0.017	-0.009	0.004	0.011	0.013	0.011	-0.023	-0.047	-0.041
	(0.028)	(0.024)	(0.022)	(0.022)	(0.021)	(0.023)	(0.024)	(0.027)	(0.032)	(0.038)	(0.042)	(0.044)
PM: ATT ($\hat{\gamma}^{kWh}$)	-0.032	0.005	0.027	0.031	0.034	0.053	0.037	0.032	0.002	-0.026	-0.023	0.000
	(0.048)	(0.050)	(0.050)	(0.048)	(0.045)	(0.043)	(0.040)	(0.037)	(0.036)	(0.034)	(0.031)	(0.029)
N	1,379	1,379	1,379	1,379	1,379	1,379	1,379	1,379	1,379	1,379	1,379	1,379
$N \times T$	~196,000	~196,000	~196,000	~196,000	~196,000	~196,000	~196,000	~196,000	~196,000	~196,000	~196,000	~196,000
Weather Controls	x	x	x	x	x	x	x	x	x	x	x	x
HH Fixed Effects	x	x	x	x	x	x	x	x	x	x	x	x
Month-of-Year Effects	x	x	x	x	x	x	x	x	x	x	x	x
Day-of-Week Effects	x	x	x	x	x	x	x	x	x	x	x	x

Note: Standard errors clustered at the household level in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. All estimates are based on a sample comprised of both experiments. The sample used to produce all estimates is based on *hourly* electricity meter readings in kWh. All panels report estimates of the coefficient of interest from separately estimated models that are conditional on readings corresponding to the given hour in the AM or PM. The results in Panel A are based on a sample comprised of weekdays only. Panel B reports analogous estimates based on weekend days only. Across all specifications, the min(rk LM statistic)=609.465 and the min(rk Wald F statistic)=1,942.422. Based on the values of the rk LM and Wald F statistics, we reject the nulls of an under or weakly identified model across all specifications. Full regression diagnostics are available from the authors by request.

H.3 Price and Peak-Alert Estimates

Table 16: ATT Estimates of the Effect of a Smart Thermostat on Energy Use by Price Quintile

	(1)	(2)	(3)	(4)	(5)
	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
	Power Use (kWh/hour)				
ATT ($\hat{\gamma}^{kwh}$)	0.027	-0.002	0.003	-0.008	-0.012
	(0.027)	(0.020)	(0.021)	(0.024)	(0.026)
N	1,379	1,379	1,379	1,379	1,379
$N \times T$	3,274,085	3,272,108	3,273,626	3,272,248	3,272,504
F statistic	694.623	780.587	786.169	717.666	631.038
rk LM statistic	713.548	557.231	514.273	466.016	585.747
rk Wald F statistic	1,827.181	1,849.143	1,798.583	1,651.202	1,938.916
Weather Controls	x	x	x	x	x
HH Fixed Effects	x	x	x	x	x
Month-of-Year Effects	x	x	x	x	x
Day-of-Week Effects	x	x	x	x	x

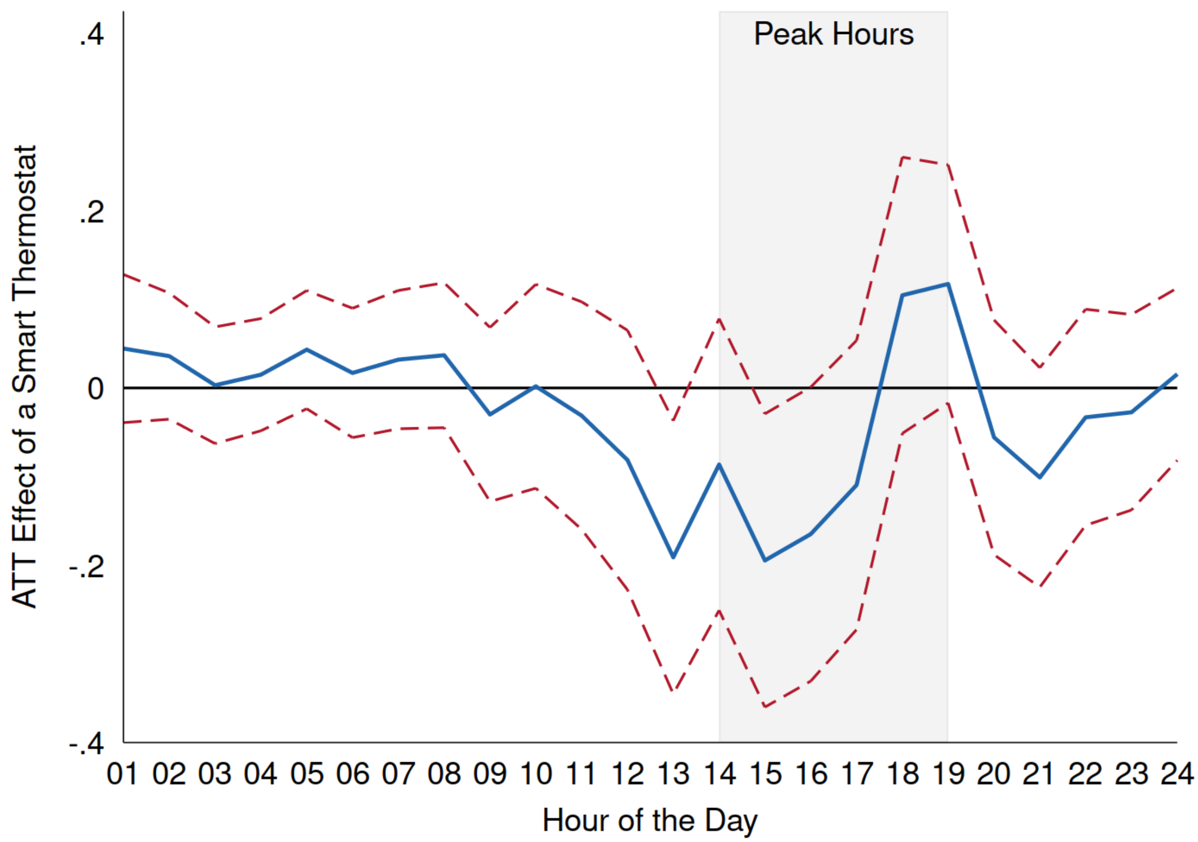
Note: Standard errors clustered at the household level in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. All estimates are based on a sample comprised of both experiments. The sample used to produce the estimates is based on hourly electricity meter readings in kWh. Based on the values of the rk LM and Wald F statistics, we reject the nulls of an under or weakly identified model across all specifications.

Table 17: ATT Estimates of the Effect of a Smart Thermostat on Energy Use on Peak-Alert Days

	(1) CAISO Alert Hours	(2) CAISO Alert Hours w/ Local Media Coverage	(3) PG&E Alert Days	(4) PG&E Alert Days: Off-Peak Hours	(5) PG&E Alert Days: Peak Hours
	Power Use (kWh/hour)				
ATT ($\hat{\gamma}^{kwh}$)	-0.045	-0.012	-0.071**	-0.038	-0.122*
	(0.052)	(0.069)	(0.033)	(0.029)	(0.065)
N	1,378	1,376	1,378	1,378	1,378
$N \times T$	227,005	158,614	475,920	356,940	118,980
F statistic	93.505	84.498	98.224	58.977	128.792
rk LM statistic	595.714	630.461	531.659	536.374	534.820
rk Wald F statistic	1,575.686	1,581.261	1,777.235	1,775.537	1,785.265
Weather Controls	x	x	x	x	x
HH Fixed Effects	x	x	x	x	x
Day-by-Hour Effects	x	x	x	x	x

Note: Standard errors clustered at the household level in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. All estimates are based on a sample comprised of both experiments. The sample used to produce the estimates is based on hourly electricity meter readings in kWh. Based on the values of the rk LM and Wald F statistics, we reject the nulls of an under or weakly identified model across all specifications.

Figure 21: ATT Estimates of the Effect of a Smart Thermostat on Electricity Use on PG&E Peak Alert Days by Hour of the Day



Note: This figure displays hourly estimates of the ATT effect of a smart thermostat on electricity use (kWh/hour) during PG&E Peak Alert days. All estimates are from a DDIV model estimated on a sample comprised of households from both experiments during the given hour on days when an alert was issued.

I Recruitment and Enrollment

I.1 Subject Eligibility

Appendix Table 18 summarizes the eligibility requirements for participation in the experiment. Participants had to own their residence and have central air conditioning with a single thermostat. They also had to have a smart phone and high-speed Internet. Finally, individuals who were planning to move in the near future were excluded from the experiment.

Table 18: Subject Eligibility Summary

	Eligible	Not Eligible
Rent or own?	Own	Rent
Home Type	House or Condo	Apartment or Other
Phone	iPhone or Android	Blackberry or Other
# of Thermostats	1	≥ 2
A/C	Central Air	Box Unit, Fans, Other
Heating	Air Vents	Baseboard or Other
High-speed Internet?	Yes	No
Plan to move in next year?	No	Yes

Note: This table presents the eligibility requirements for participating in the experiments we analyze.

I.2 Trial Recruitment and Enrollment Guide

FOR ONLINE PUBLICATION: ONLINE APPENDIX B

UTILITY Smart Thermostat Trial

Recruitment and Enrollment Guide

November 19, 2012

Introduction

Experimental Trial Information

UTILITY is running an experimental thermostat trial with Opower and Honeywell, offering eligible customers a free remote-controlled thermostat solution (a thermostat controlled by a smartphone and web application). The goal of the experiment is to test the energy savings and customer experience of the thermostat solution. Customers gain a thermostat and app that helps them save energy, by creating a customized, energy efficient schedule that fits their lifestyle.

For this trial, 1 in 2 qualifying customers will receive the thermostat solution. Customers who meet the eligibility qualifications must complete the online enrollment process to determine if they will receive a thermostat or not. At the end of the online enrollment process the system will randomly flip a coin to determine which customer will receive the remote-controlled thermostat and which will not. All customers who enroll for a chance to participate are benefiting the trial (even those who do not receive a thermostat), and it is important that all qualified customers complete the full enrollment process.

Customers should be encouraged to enroll for a chance to receive this exciting solution, which allows them to control their thermostat on-the-go. UTILITY, Opower, and Honeywell are grateful for the time each customer takes to enroll online for a chance to participate, and all customers should be thanked for their time regardless of the outcome.

Customers should be encouraged to answer all qualification and enrollment questions honestly. If a customer provides inaccurate information during enrollment it negatively impacts the trial and the customer will ultimately be turned down for the trial.

Talking Points for Recruitment Events

Initial Communication

Initial communication should be a call to action, provide quick benefits (FREE remote-controlled thermostat), provide a fun atmosphere and garner attention.

- Do you own an iPhone or an Android? If so, would you be interested in a free thermostat controlled by your smartphone?
- How would you like to gain better control of your energy use at home? You can control your thermostat at home from right here! Want to know how?
- Sign-up for a free remote-controlled thermostat, a $500 dollar value and take control of your energy consumption and improve the comfort of your home.
- I know you're in a hurry but this opportunity will allow you to take control of your energy use and you'll always come home to a house at the perfect temperature.
- Save energy while you're away and stay comfortable while you're at home, all by using your smartphone or the web.
- How would you like to control your heating/cooling by your iPhone or Android and

through the internet from anywhere in the world?

After Initial Communication

After initial communication, you should be focused on getting the customer more excited about the offering by providing key information and benefits unique to the opportunity.

- We are conducting a trial on behalf of UTILITY that allows you to interact with your heating & cooling system using your smartphone or the web. That means you can control your home's comfort at your fingertips from wherever you are. All you need is your smartphone of the web. Are you ready to take control?
- Did you know that a typical family spends almost half (49%) of its energy cost on heating and cooling? (*Source:* Energy Star)-- How would you like to have the opportunity to be selected for a special trial UTILITY is conducting to provide a limited number of customers a thermostat controlled by your smartphone? That's right you can control the comfort of your home at anytime or any place using your smartphone or the web.
- How would you like to be one of the lucky UTILITY customers who receives a free thermostat controlled on-the-go from your smartphone or the web? This is over a $500 value completely free with professional installation and a 1-year warrantee. UTILITY is conducting this trial to allow customers a unique way to reduce energy use and save money. The process for signing up only takes a few minutes of your time. Let's see if you qualify.
- Check out this free thermostat controlled by your smartphone. You'll have complete control over your comfort, and you can see how your temperature settings stack up against other participants in the trial.

Overcoming Initial Objections

Objection: "I don't have time"

- You'll never come home to a cold house again and sign-up only takes a few minutes.

Objection: "I still don't have time"

- Okay; here's how you can see if you qualify and sign-up from home (postcard)

Objection: "I don't want to give out my personal information"

- You're information is completely confidential and will be only used to determine if you qualify for the free thermostat.

Objection: "I'm not interested"

- Here is a free pen, compliments of UTILITY. Have a great day!

Initial Eligibility Screening

	Eligible	Not eligible

Do you rent or own your home?	Own	Rent
What kind of home do you own?	Single family, Townhome, Condo	- Apartment - Other
What kind of phone do you have?	- iPhone - Android	- Blackberry - Other
How many thermostats do you have in your home?	One (1)	Two (2) or more
How do you cool your home?	Central air	- Window box unit - Fans - Other
What is the main way you heat your home?	Air vents	- Baseboard - Other - None
Are your heating and air conditioning systems functional and have you used them the last 6 months?	Yes	No
Do you have high-speed internet access (Cable, DSL, satellite, Broadband)?	Yes	No
Do you have an available *ethernet* port on your internet router?	Yes	No
Do you plan to move to a new home in the next 12 months?	**No**	Yes
Will other adults in your household object to enrolling in this program?	**No**	Yes

Customer Does NOT Pass Initial Eligibility Screening

- Thank you for your interest, but unfortunately you don't meet the eligibility requirements for this trial. However, UTILITY is developing a number of residential energy efficiency programs that you may qualify for. Please fill out this post card in to enable them to contact you in the future for other offerings. Thank you and please accept this free pen, compliments of UTILITY. We appreciate your time!

- If you do know someone else who may be interested, please let them know about this free trial and they can sign-up right away. (Staffer hands the customer a post card.)

Customer Passes Initial Screening

- Great! You've pre-qualified to participate in the selection process, which only takes a few minutes. Would you like to learn how the thermostat and app works? (demo)

- Let's get you signed-up and see if you are selected to join the UTILITY Smart Thermostat Trial, with a free remote- controlled thermostat and professional installation. The sign-up process just takes a few minutes and we can help you complete it here.

- You'll need your UTILITY account number for enrollment. You can use my phone to retrieve your utility account number from UTILITY. You will also be asked to provide the last four digits of the Social Security Number of the UTILITY account holder—this may be you or a housemate. Staffer provides customer phone & contact number (1-888-743-0011).

Customer is Selected to Join the Trial

Encourage customers to take the first available appointment. Explain that technicians are only in the area for a limited amount of time.

- Congratulations! You've been selected to participate in the UTILITY Smart Thermostat Trial. A customer service representative will contact you with further information about your free installation. You will receive an email reminder with the date and time of your installation appointment, but you may want to write it down now, so you don't forget.
- Tell your friends and family to see if they are eligible and sign-up online! (postcard)
- Here is a free lens cleaner or smartphone holder for your smartphone, compliments of UTILITY. We appreciate your time!
- You will be contacted within a few days to confirm your eligibility and appointment time. (Honeywell CSR will conduct a follow-up call to confirm appointment time & answer any additional questions)

Customer is NOT Selected for the Trial

Thank you for your interest in the Smart Thermostat Trial. Unfortunately, this is currently a trial so participation cannot be granted for everyone.

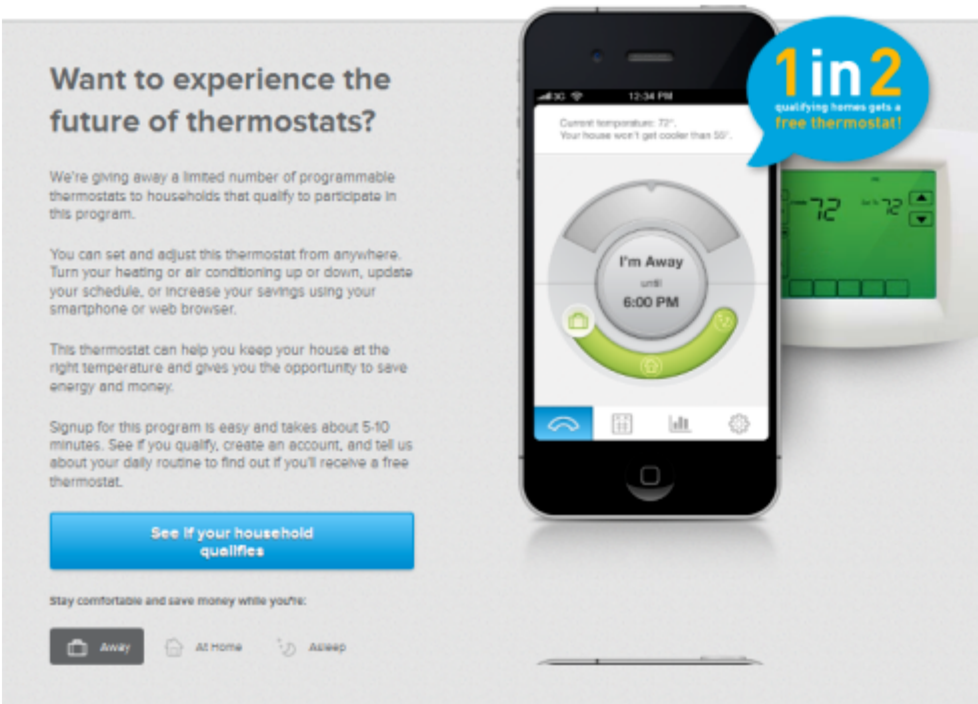
- In the event the trial is extended, would you like to leave your contact information, which will only be used to contact you regarding other opportunities to participate in UTILITY residential trials or programs?
- Please accept this free pen, compliments of UTILITY. Have a great day.
- Tell your friends and family to see if they are eligible and sign-up! (postcard).
- Here is a free lens cleaner or smartphone holder for your smartphone, compliments of UTILITY. We appreciate your time!

How Online Enrollment Works

If a customer passes the initial qualification screening, direct them to the Opower Web application to enroll online. Eligible customers have a 1 in 2 chance of being selected to receive a thermostat.

Enroll online at: https://thermostat.opower.com/

The customer begins by clicking "See if your household qualifies."

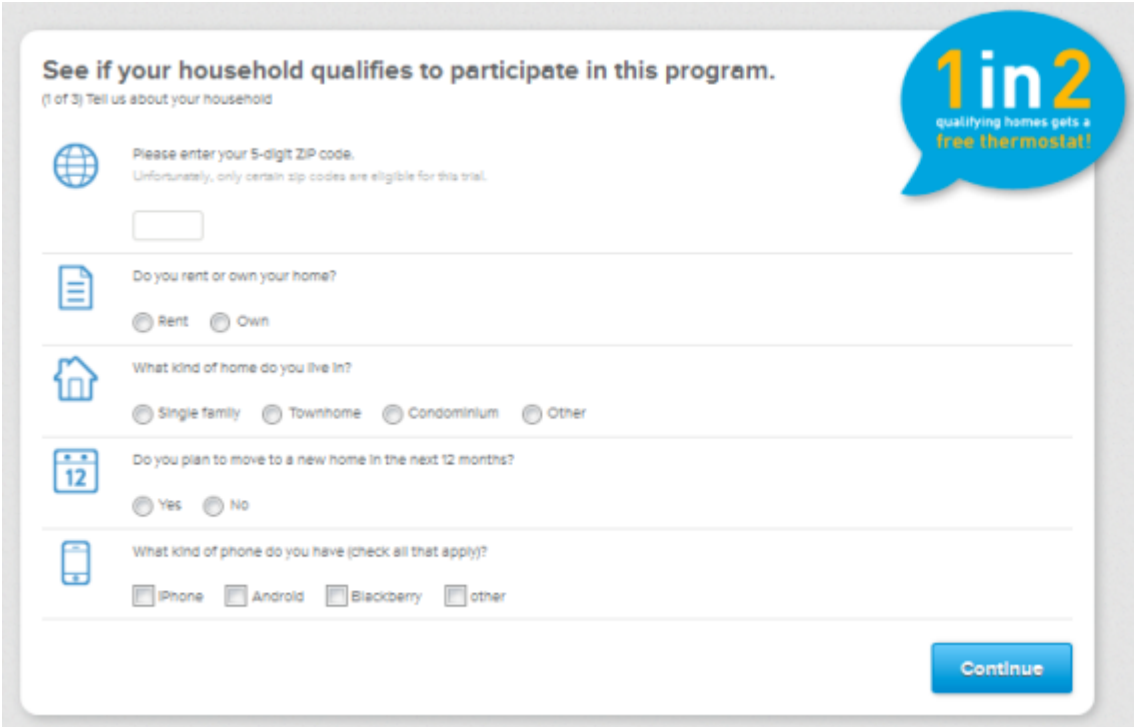


Verifying if the Household Qualifies

In order to verify that they can participate in the program, customers must answer a series of questions about their home.

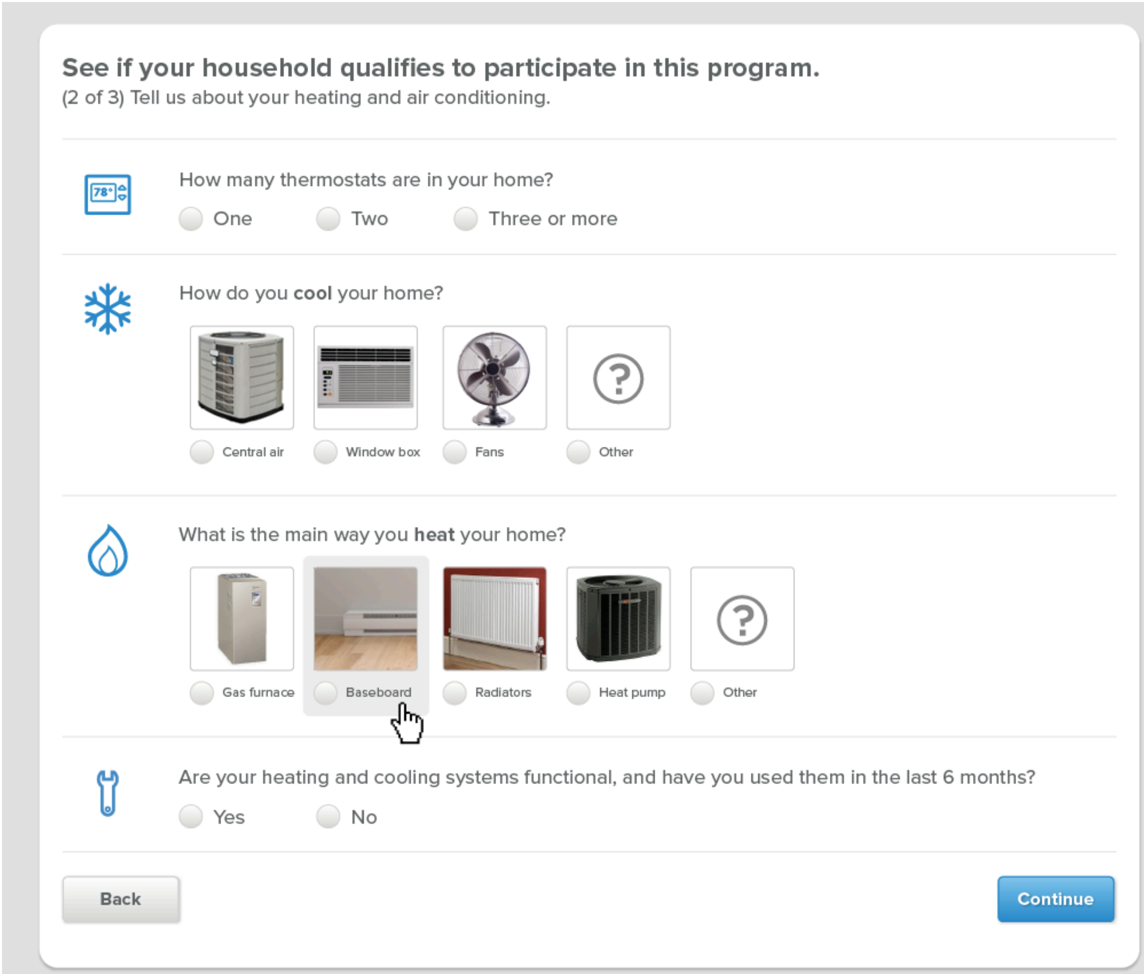
On the first verification screen, they are asked to provide the following information:

- Zip code: Qualified zip codes are those within the greater Fresno and Bakersfield areas, see list provided by Honeywell.
- Whether they rent or own: Customers must own their own home.
- What kind of home they live in: Customers can select any option except "other."
- Whether they plan on moving in the next year: Customers must plan on remaining in the same home.
- What kind of phone they have: Customers must have an iPhone or Android phone if the utility program requires a smartphone.



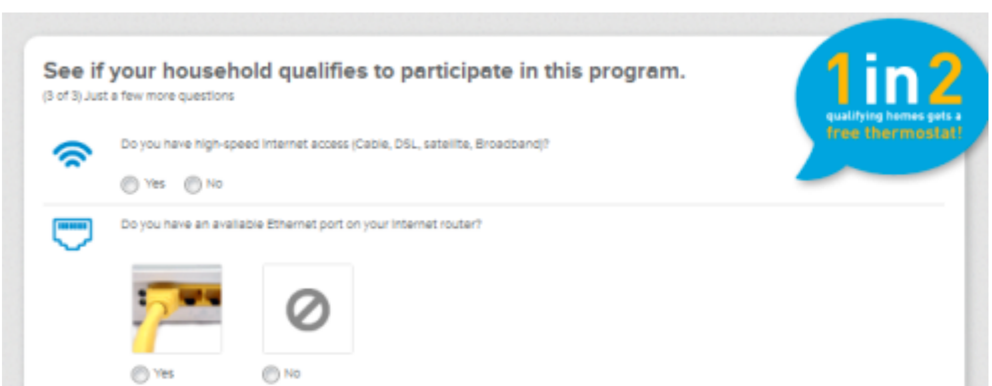
If a customer qualifies based on the answers to the questions above, they are asked to provide the following additional information:

- Number of thermostats: Customers can have only one thermostat.
- Primary cooling system: Customers must have central air.
- Main way they heat their home: Customers must have a gas furnace.
- If their air conditioning and heat are currently working: Customers must have an operational air conditioner and heater that they have used in the last 6 months.



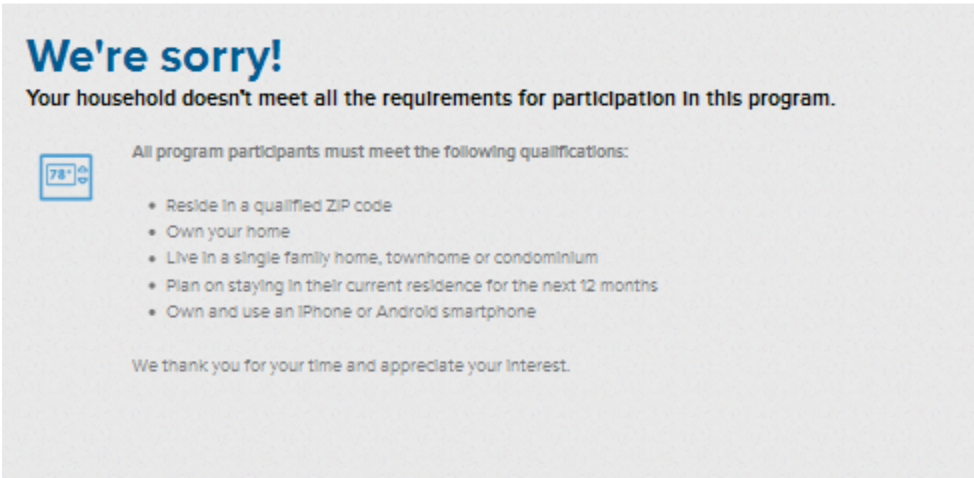
Finally the customer is asked, if they:

- Have high-speed Internet access: Customers must have high-speed access.
- Have an available Ethernet port on their router: Customers must have an available port.
- Are in agreement with the terms and conditions of the program: Customers must agree to the terms. Terms vary by utility.



When they complete the final verification screen, they are told if they are eligible to receive an account. They must meet all of the qualifications to be considered for the program.

If a customer answers any of the qualification questions with a response that makes them ineligible, they are excluded from the program.

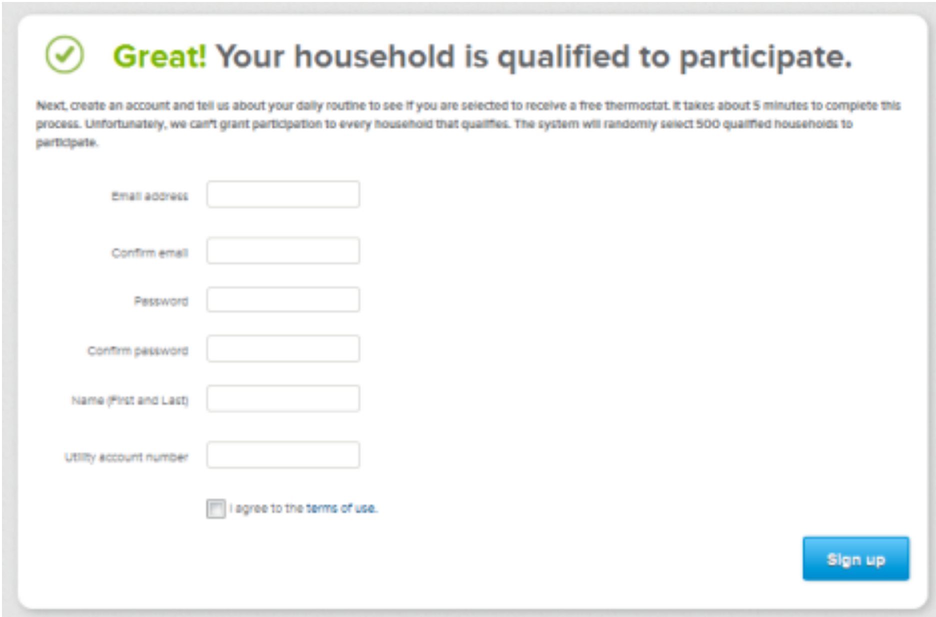


Creating an Account

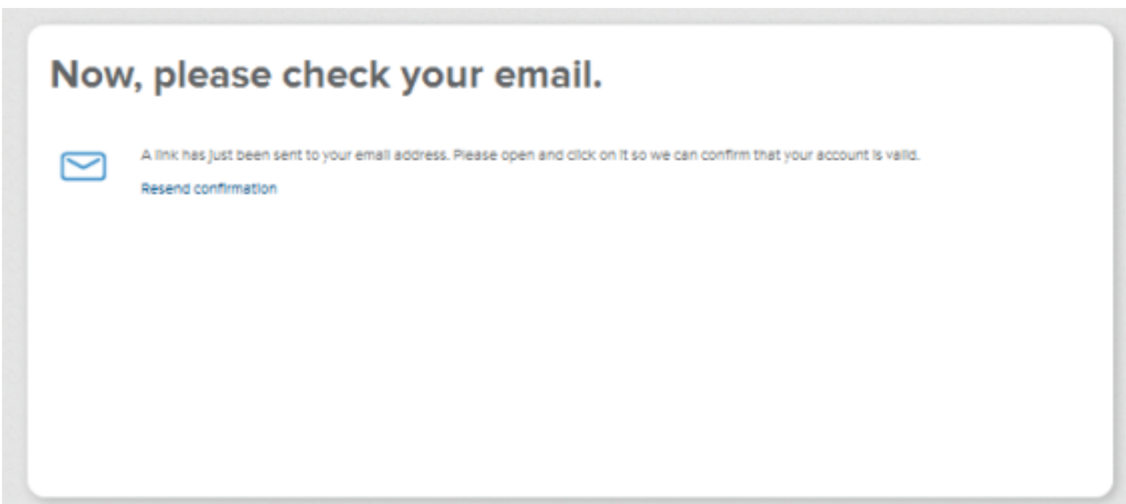
Customers who are eligible for the program are required to enter the following information to create an account:

- The email address they will use to access the Web application. Basic validation is performed to verify that the email address is well-formed.
- A unique password. The password must be at least eight characters long. Passwords must not be or contain the customer's name or email address.
- Customers enter the same password again and are prompted to correct the password if it is not identical in the two password fields.
- The full name of the utility account holder exactly as it appears on the utility bill. The customer enrolling in the program must enter the name of the utility account holder as it appears on the utility bill, even if they are not the account holder.
- The utility account number exactly as it appears on the utility bill. This includes spaces or any other characters included in the data.

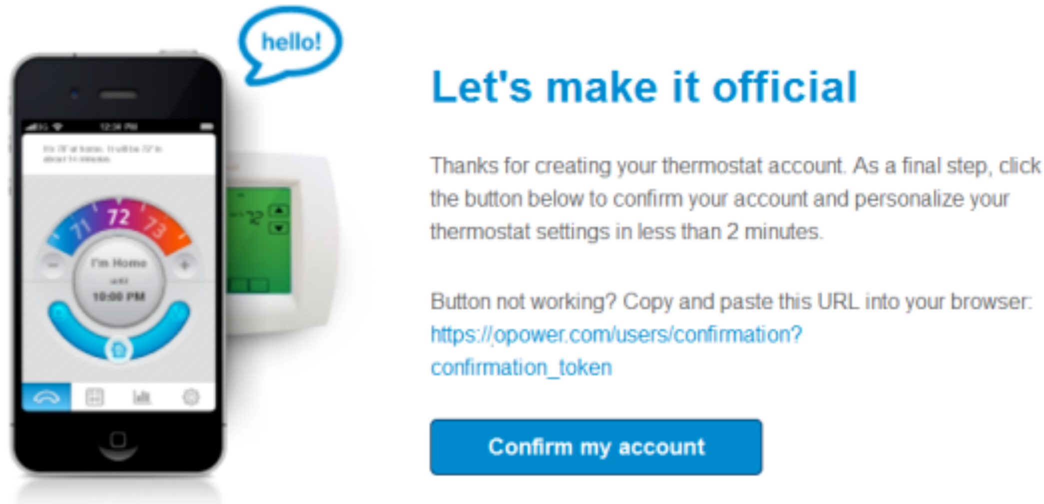
Customers are prompted to agree to the Opower Terms of Use.



Customers submit their account information, and then a new page prompts the customer to check their email.



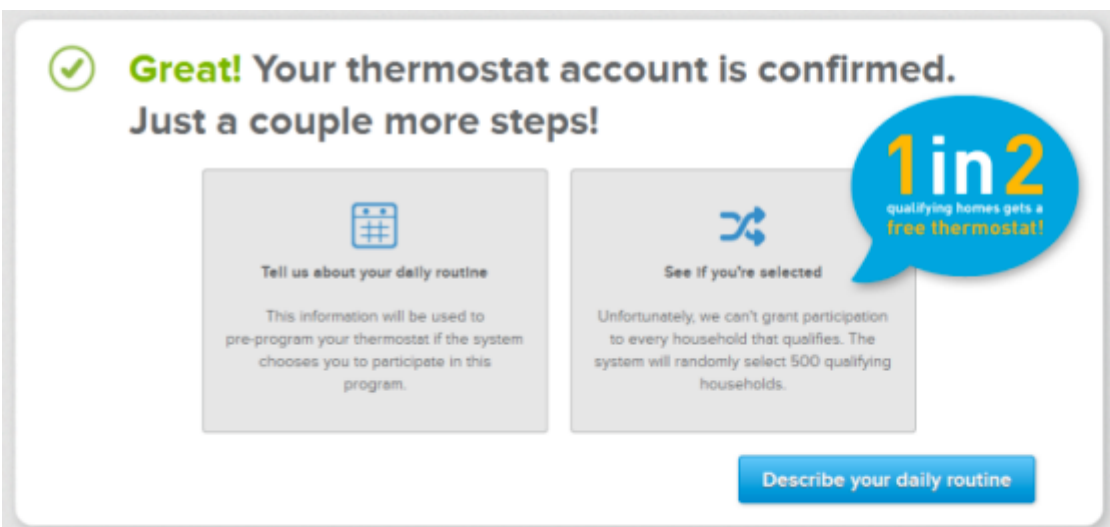
Customers should receive an email message at the address they specified. If the customer does not receive the email, they have the option to "Resend confirmation" in the Web application. The email is titled "Your Thermostat," and it will arrive from an @opower.com email address. The customer may need to check their junk/spam folder for the email.



The customer must click "Confirm my account" to complete their registration and verify their email address. If nothing happens when the button is clicked, the customer can copy and paste the customer-specific URL provided in the email to their Internet browser to confirm the account.

Thermostat Registration

Once the customer has confirmed their account, they are provided with more information about the program and asked to describe their daily routine.



Qualifying Questions

The customer begins to program their thermostat by providing the following information:

- Whether multiple people live in their home. Opower tailors the language in the application to the number of people in the household.

- Whether they have pets. If the customer has pets, the default away temperature of the home is adjusted to a safe temperature for household pets. For homes with pets, the default away temperature is 82 instead of 85 for cooling and 60 instead of 55 for heating.

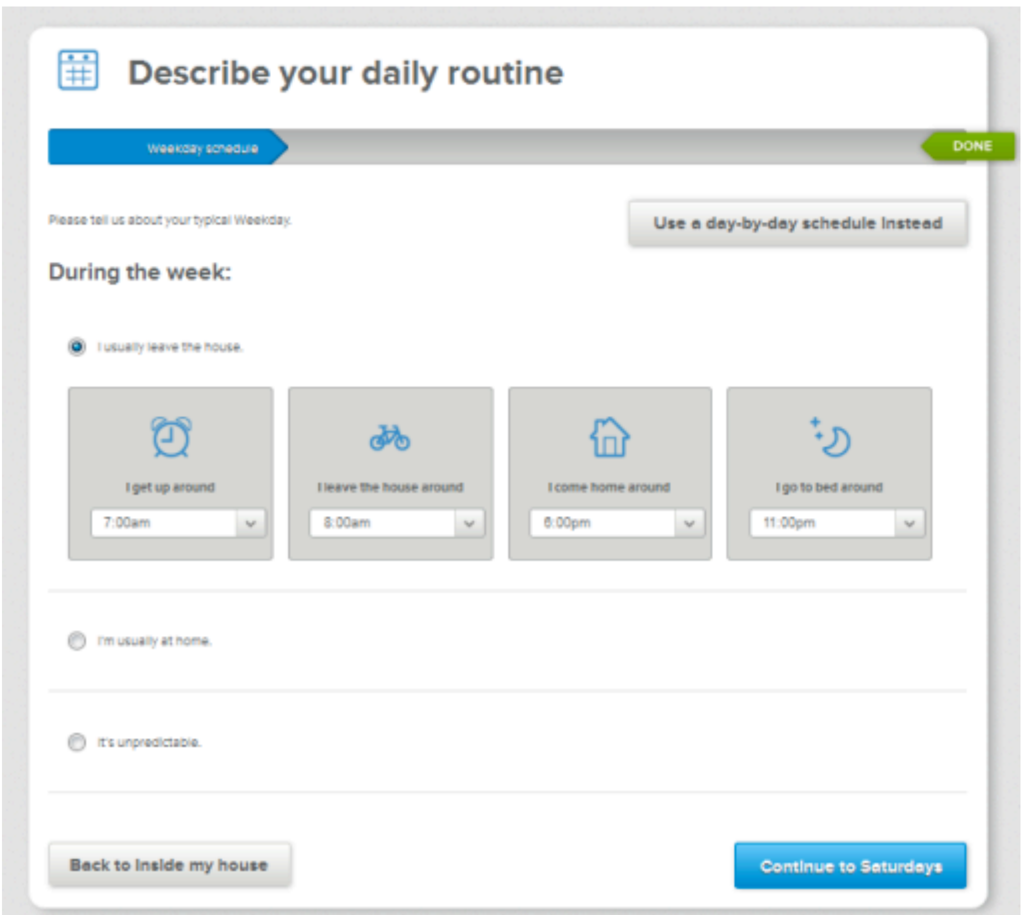
- Their mobile phone number. Customers are sent a text message to this number with a link to the Opower mobile application..

Setting an Initial Schedule

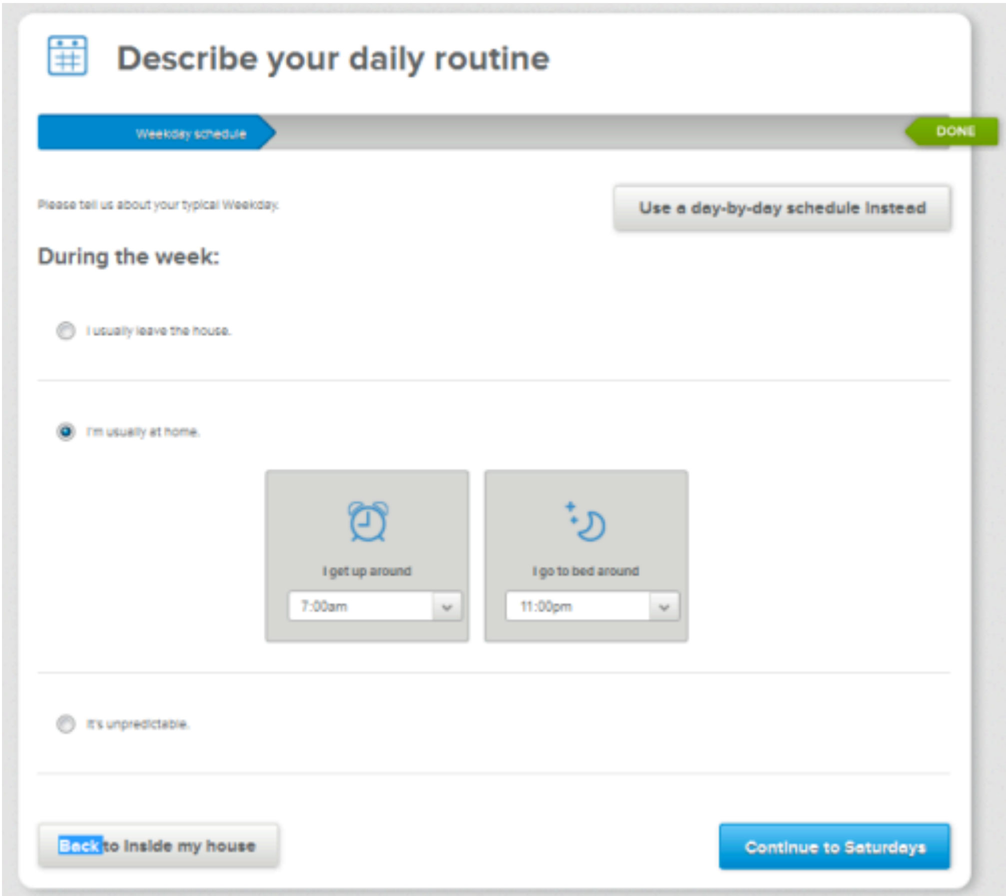
After completing the qualification questions, the customer is prompted to create a personalized schedule. By default, customers set a schedule for all weekdays and then Saturday and Sunday.

For all weekdays, Saturday, and Sunday, the customer has the following options:

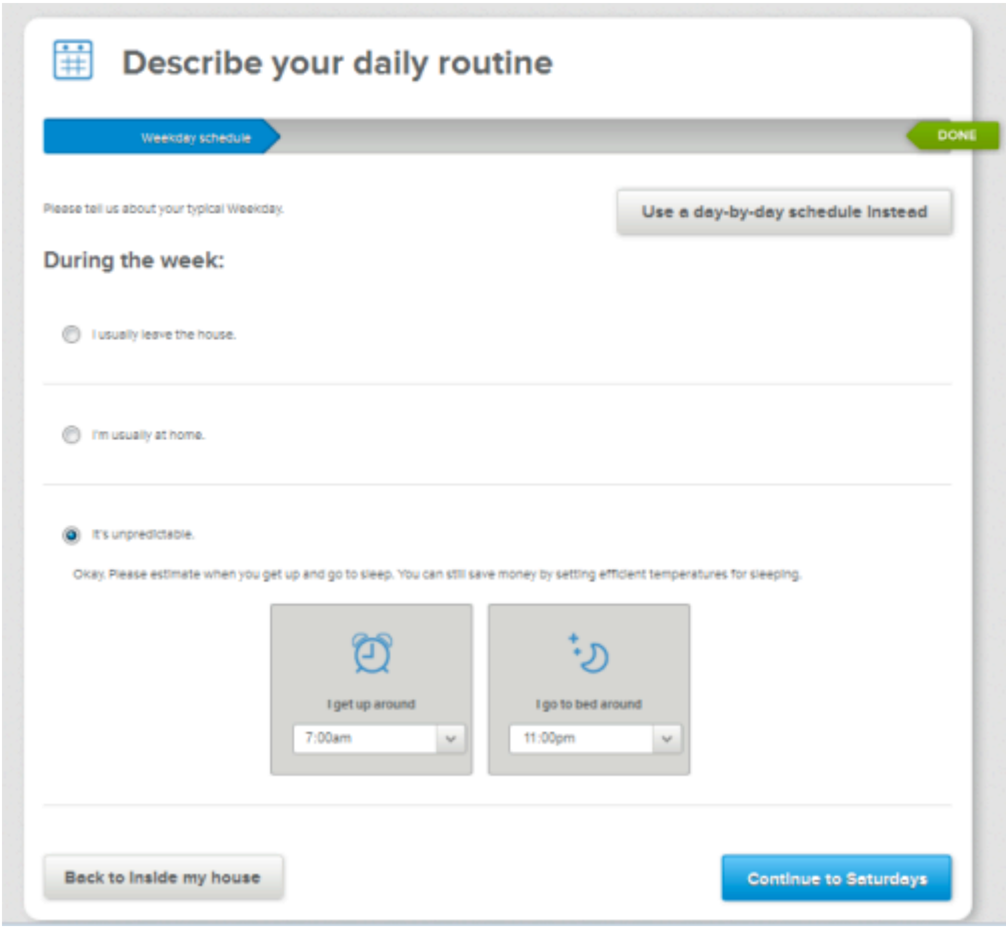
- They can set a schedule for when they typically wake, leave the home, return home, and go to sleep.



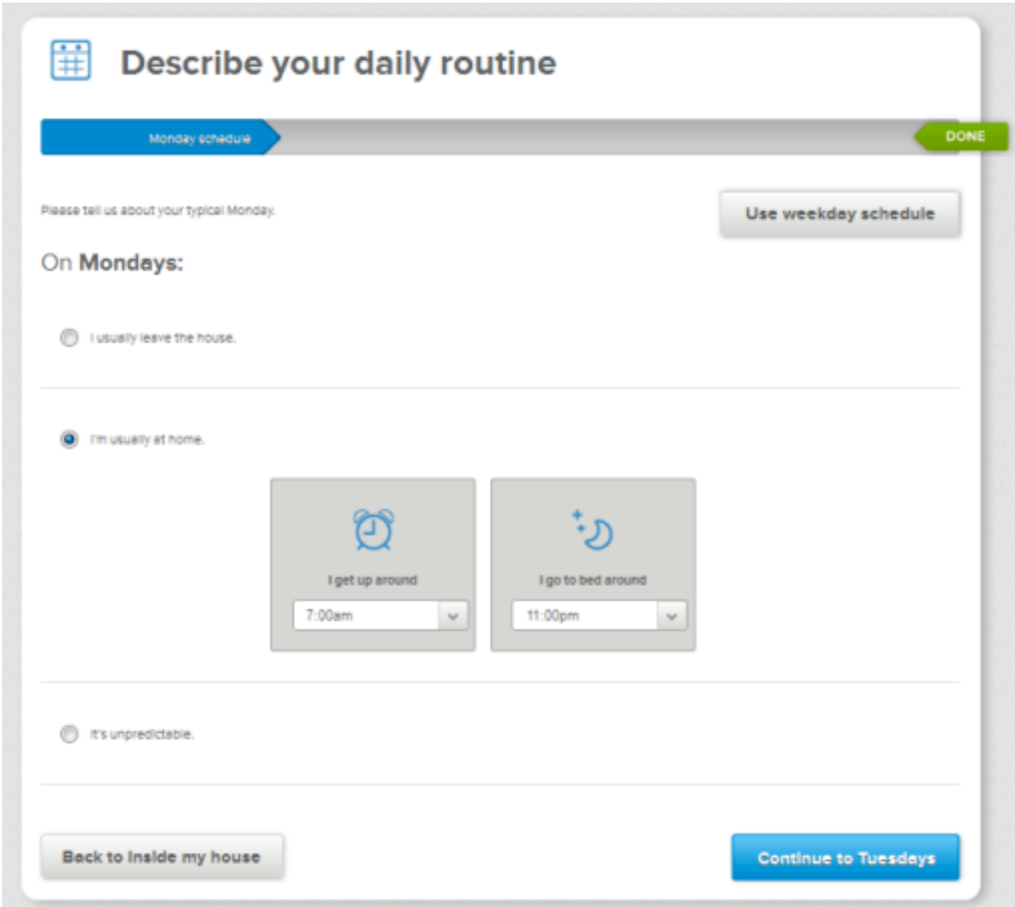
- They can indicate they are home all day and set the time for when they usually wake and go to sleep.



- They can indicate their schedule is unpredictable. In this case, they are still asked when they typically wake and go to sleep.



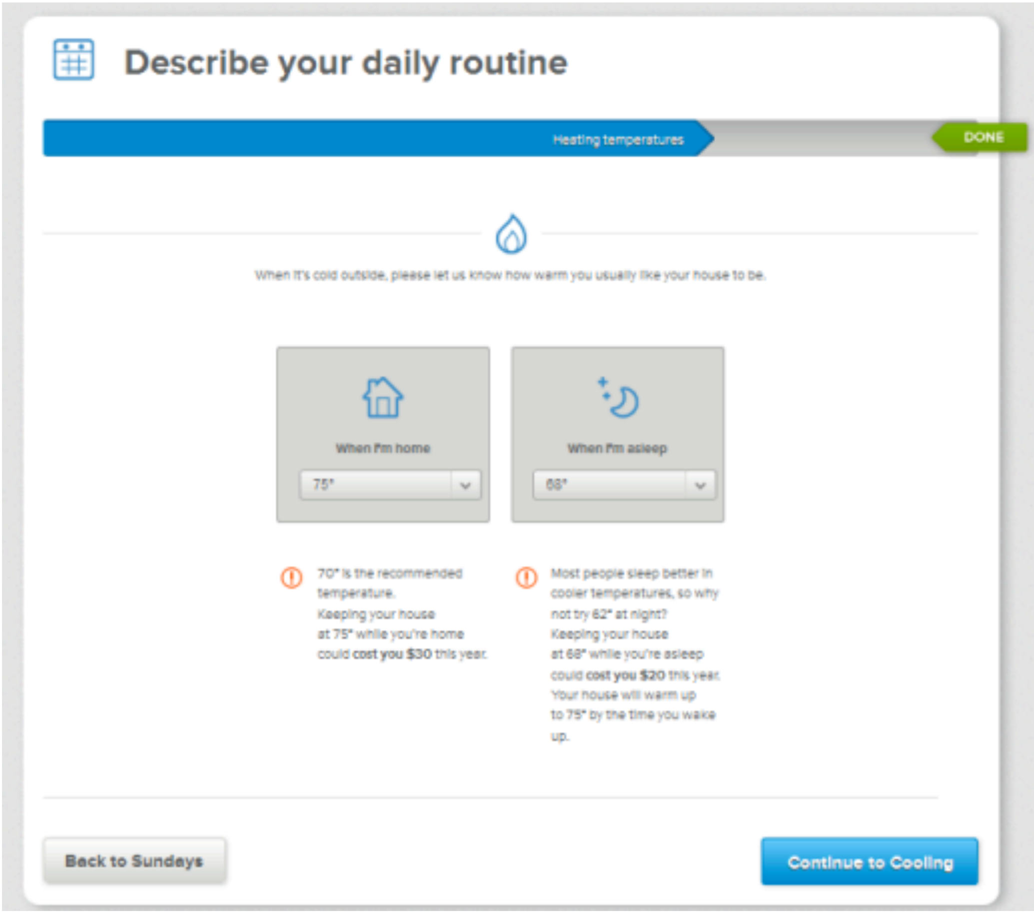
Instead of setting the same schedule for all weekdays, a customer can also create a day-by-day schedule for each weekday separately. The same schedule options are available on a daily basis.



Setting Initial Temperatures

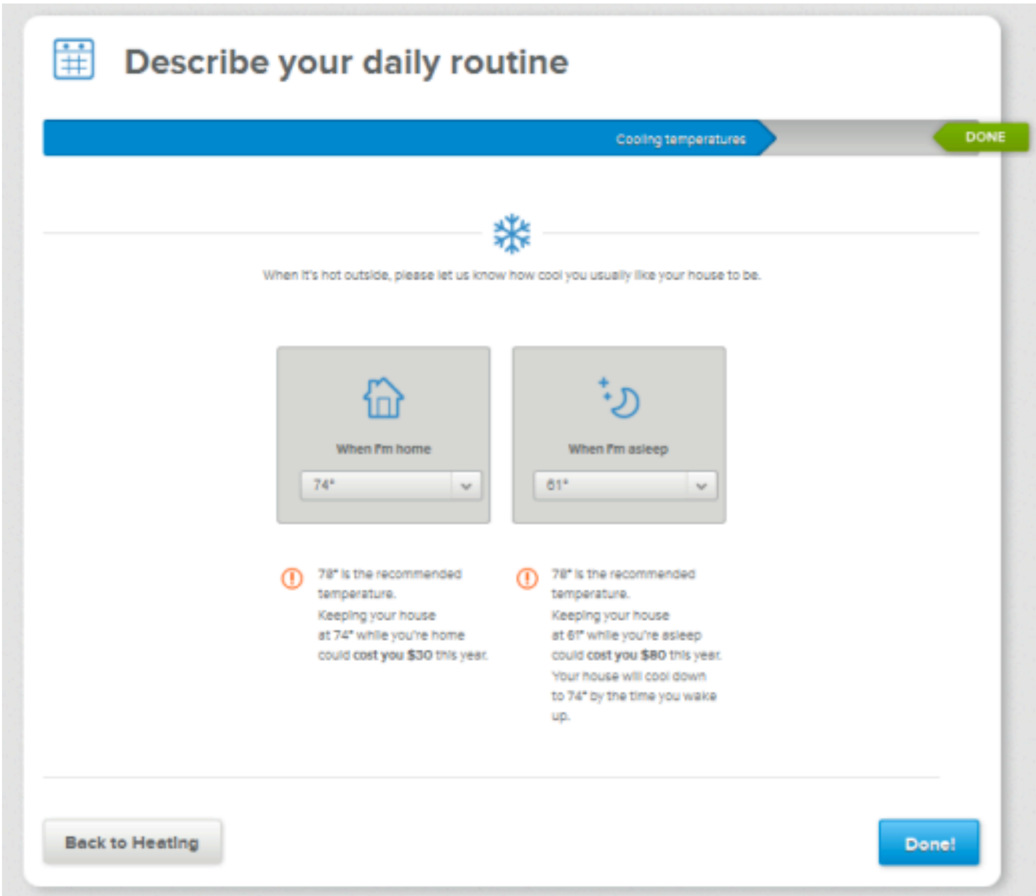
Customers are prompted to set their home and sleep temperatures for heating and cooling. The default temperatures for these settings are based on the suggested Energy Star settings (ENERGY STAR® Program Requirements for Residential Climate Controls, Version 1.0 Partner Commitments, DRAFT 2).

On the heating page, customers are asked how warm they would like their home to be when they are home and asleep.



If the home temperature is greater than the recommended setting (less efficient), an insight appears to tell them how much money they will spend during the winter keeping the home at this higher temperature. If the away temperature is higher than the recommended setting, they are prompted to try setting the temperature lower since the house will warm up to a comfortable setting before they wake up.

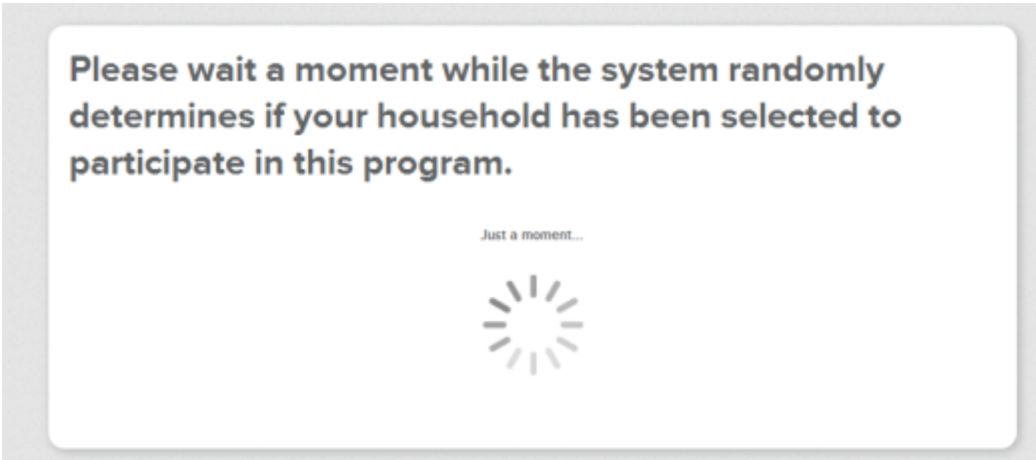
On the cooling page, customers are asked how cool they would like their home to be when they are home and asleep.



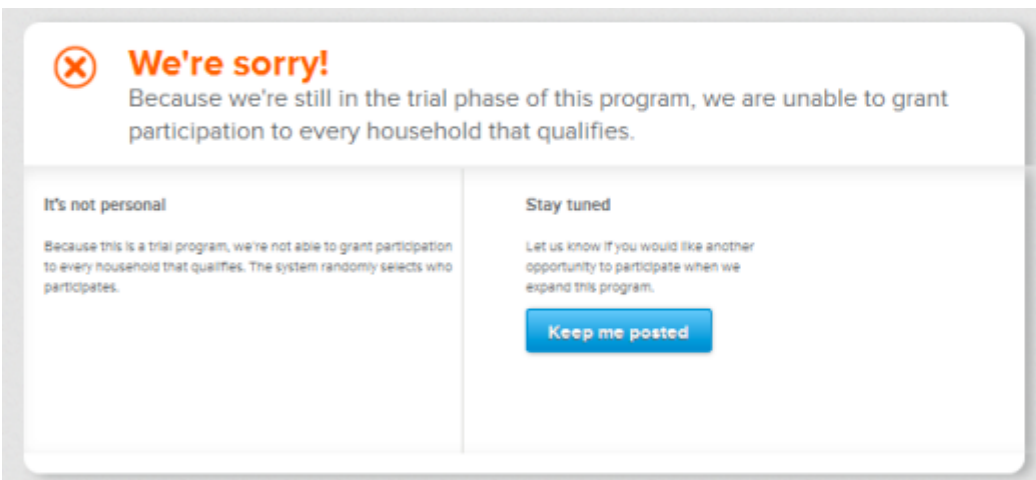
If the home temperature is less than the recommended setting (less efficient), an insight appears to tell them how much money they'll spend during the summer keeping the home at this lower temperature. If the away temperature is lower than the recommended setting, they are prompted to try setting the temperature higher since the house will cool down to a comfortable setting before they wake up.

Installation

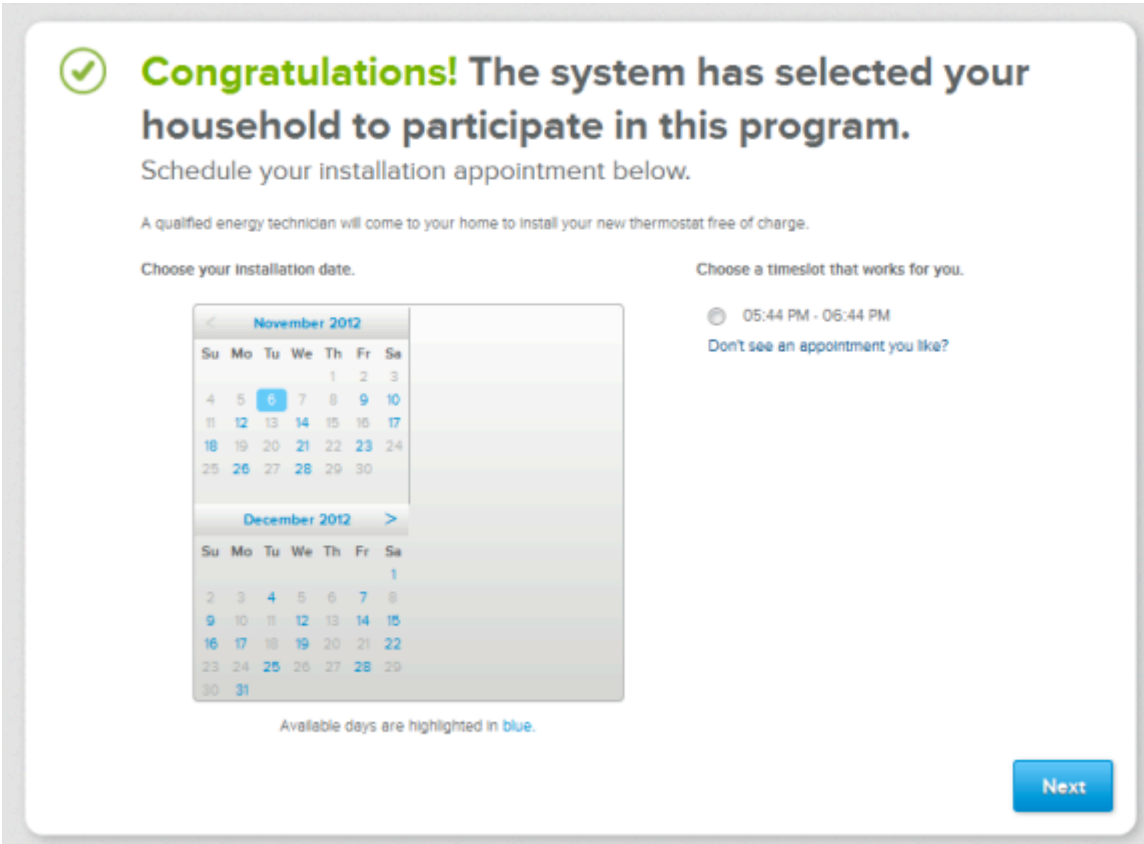
After submitting their temperature settings, the customer is randomly selected to be part of the test or control group.



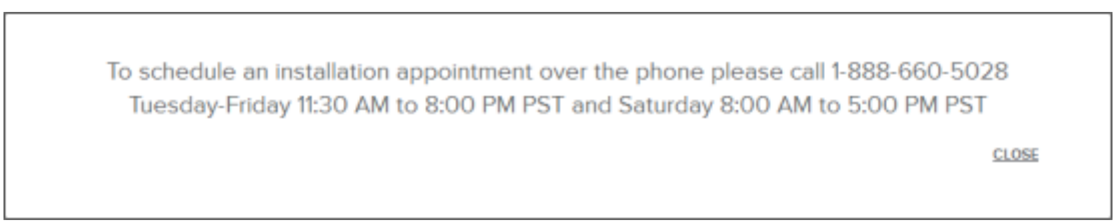
If they are part of the control group, they will not receive a thermostat. Customers in the control group may opt to sign up for a waiting list and may receive a thermostat if the program is expanded.



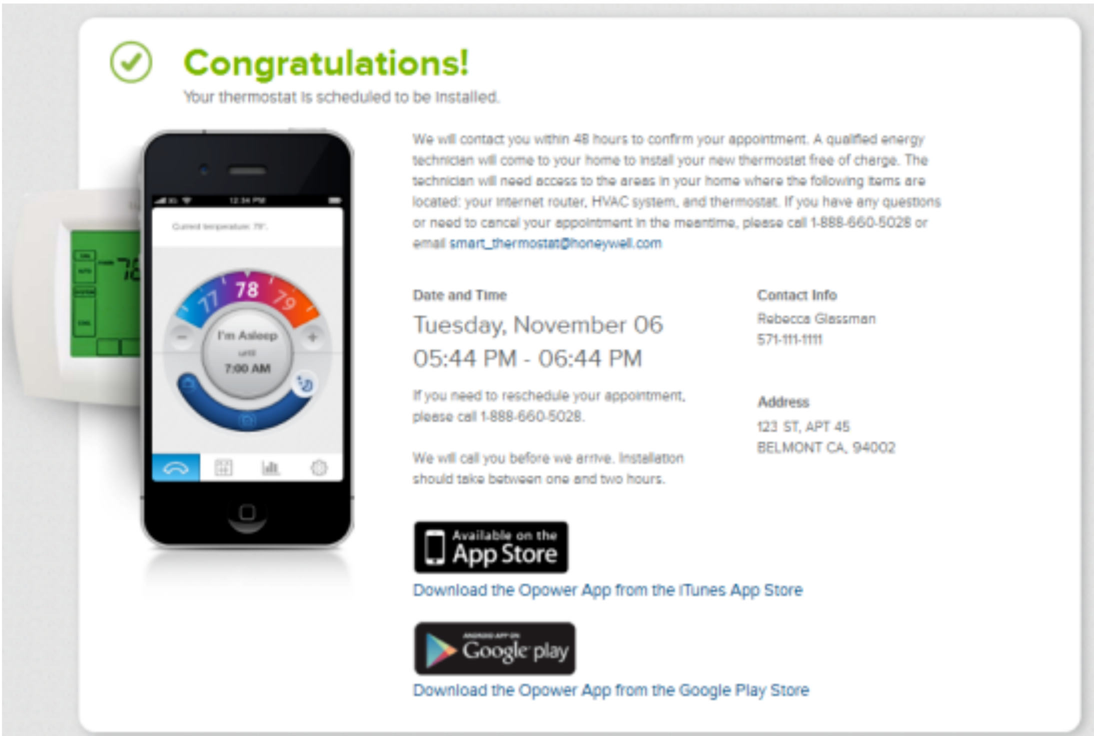
If they are randomly selected into the test group, they will receive a thermostat and become part of the program. Customers participating in the test group can schedule an appointment to have their thermostat installed.



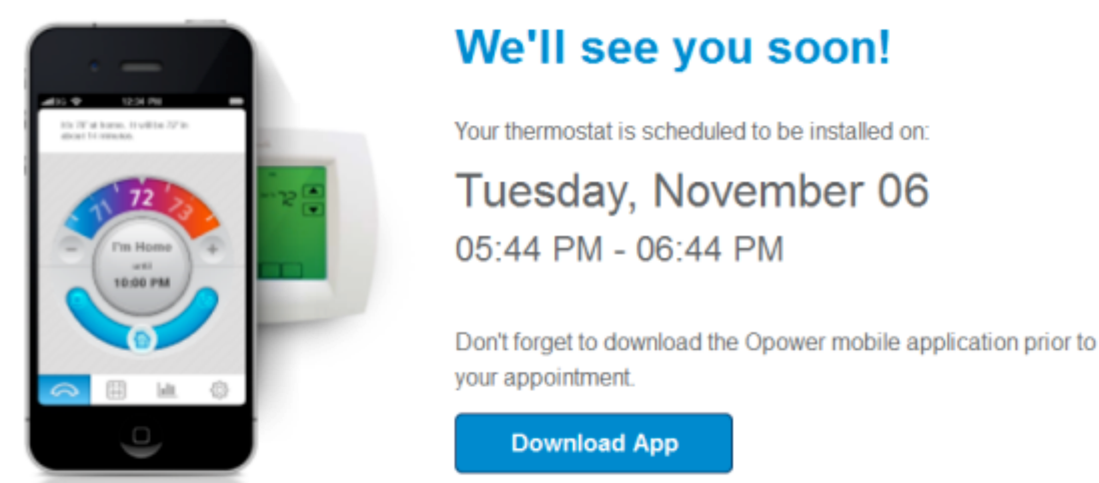
If none of the times available on the screen are convenient for the customer, they can click "Don't see an appointment you like?" to see a phone number they can call to schedule the appointment (1-888-660-5028).



Once they have selected the date and time for their appointment, they will see a confirmation screen. This includes information on how to reschedule the appointment and where to download the mobile application.

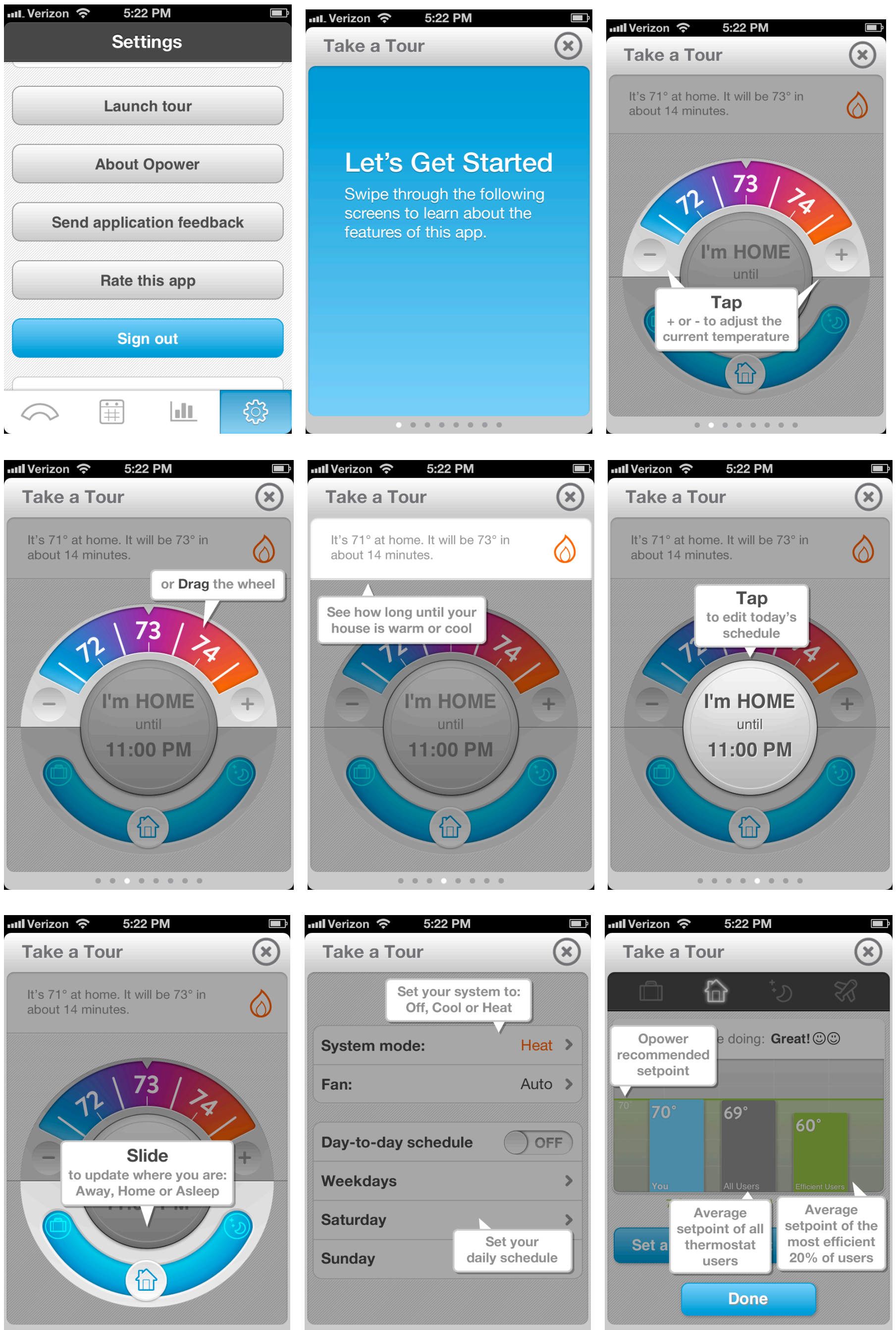


The customer will also receive an email confirmation for their appointment and a reminder to install the mobile application in advance of the appointment.



Mobile Application Tour

The mobile application tour can be launched at anytime, using the Opower mobile app on the iPod Touches, and later on the customer's smartphone. Click on the Settings tab, click "Launch tour," slide through the tour pages, and click "Done" to exit. The tour provides an overview of some of the main application functions and customer messages.



Answering Customer FAQs

This section will help you answer customer questions about the program, mobile and Web applications, and thermostat. A full set of customer FAQs can be found at https://thermostat.opower.com/faq.

What is this thermostat program?

Opower and Honeywell have partnered to create a smart thermostat solution, which allows utility customers to program and monitor heating and cooling energy usage, not just from the thermostat itself, but also via Internet-connected devices like smartphones. This solution also gives you the ability to create optimal thermostat schedules that fit your lifestyle and provides customized recommendations to help you trim your energy bills.

How can I save?

A programmable thermostat can help reduce your heating and cooling costs. You can save all year long if you ensure your thermostat is set at the optimum program settings that match your lifestyle. You can manipulate your temperature setting and conserve energy, even while you are away, through the use of the Internet or your smartphone. Setting your programmable thermostat to the highest comfortable temperature in the summer and lowest comfortable temperature in the winter can help you reduce your energy bill.

What are the estimated savings based on?

The estimated costs and savings calculations are based on average heating and air conditioning usage and utility billing rates in your area. These are only estimations and are not a guarantee of savings from your utility company.

What other benefits does this program provide?

This thermostat program also benefits the community by helping to educate customers about energy use and energy efficiency goals. The energy customers save will not only help the environment, but also help reduce the need for new power plants and the occurrence of power outages.

Are there any safety or privacy concerns I should be aware of related to this thermostat program?

The Honeywell VisionPro thermostat used for this program was rigorously tested prior to being installed in customers' homes. These devices go through numerous quality control checks by multiple parties, to ensure they meet a high level of customer safety, reliability, and satisfaction.

It is also our top priority to protect our customers' information. We apply the same privacy protection standards to all data collected by the company from customers. We treat each customer's personal information and data as confidential, consistent with all regulatory requirements, including those established by the Public Utilities Commission. Therefore, be assured that your information is kept private.

Can I get this device for my other properties and/or business?

The smart thermostat program is only available for residential use at this time. Only a single thermostat is available for each program participant.

How many devices can I access the applications from?

Only a single wall-mounted thermostat is available for each program participant. You can install and access the mobile application from as many smartphones as you would like, but the application must be registered with the same username and password. Similarly, you can use the

Web application from any supported web browser on any computer. If more than one member of your household uses the application at the same time, the changes are preserved for the last person who saves their changes.

Can people see if I am home or not?

No. We apply the same privacy protection to this data as other all other data collected by the company for customers. The only way someone can see your status and schedule is if you give them your login credentials to the web or smartphone application.

If I work from home or have a severe illness for which I have special temperature needs, can I still benefit from this program?

You will always have control of your thermostat, so you can set safe and comfortable temperatures that are suitable for your lifestyle. An easy way to save energy is to lower your heating temperatures and raise your cooling temperatures when you are away. Depending on your personal needs, you may also be able to use more efficient temperatures while you are asleep.

How safe is the program? Can anyone hack into the system?

It is our top priority to protect our customers' information. Our system employs industry-standard defense mechanisms against brute-force attacks, code injection, and other malicious activity. We apply the same privacy protection standards to all data collected by the company from customers. We treat each customer's personal information and data as confidential, consistent with all regulatory requirements, including those established by the Public Utilities Commission. Therefore, be assured that your information is kept private.

What smartphones support the mobile application?

The mobile application is currently supported on the Apple iPhone 3GS or later, running IOS 4.3 or later, and Android phones running 2.2 or above. To locate your operating system on your iPhone, open the *Settings* app, click on "About," and see what "Version" your iPhone is running (needs to be 4.3 or above). To locate your operating system on your Android, open the *Settings* app, click on "About phone," and see what "Android version" your phone is running (needs to be 2.2 or above).

How do I make a one-time change to my schedule?

You can use the "Thermostat" page of the mobile application or the " My Thermostat" page of the Web application to manually change your temperature, change your current state (away, home, asleep), or set a new time to come home, wake, or go to sleep. On the thermostat on the wall, you can also manually change your temperature.

How can I change my email address and/or password?

Open the Web application, and then select "My account" to change your password or email address.

I now have three ways to change my thermostat. How are they different?

You can use your thermostat to manually change temperatures, turn on and off your heating and AC, and control your fan. The Web application has the same functionality as the thermostat and also allows you to register for an account, set a vacation schedule, and change your account settings, primary schedule, default temperatures, state (home, away, asleep), and schedule for today. The mobile application has all of the functionality of the thermostat and Web application, plus it allows you to compare your temperature settings, set a passcode, and set and receive notifications.

Which browsers are supported for the Web application?

The current major release and previous major release of the four desktop browsers with the largest market share are supported. Currently, this means Internet Explorer, Safari, Mozilla Firefox, and Google Chrome are supported.

Will my house really be comfortable enough when I get home?

Yes. You just set the time you will return home and your thermostat does the rest. Your home will be heated or cooled for you before you return home after being away or on vacation. Your smart thermostat learns the amount of time it takes to heat or cool your house before you arrive, based on the actual temperature in your home and past usage.

Can I enroll in the program using my smartphone?

You can only enroll in the program using the Web application. If you are selected for the program, you will receive information about how to install the mobile application.