ThermoCoach: Reducing Home Energy Consumption with Personalized Thermostat Recommendations

Conference Paper · November 2015

DOI: 10.1145/2821650.2821671

4 authors, including:


Rayoung Yang
University of Michigan


Mark W. Newman
University of Michigan

ThermoCoach: Reducing Home Energy Consumption with Personalized Thermostat Recommendations

Devika Pisharoty
U. Virginia

Rayoung Yang
U. Michigan

Mark W. Newman
U. Michigan

Kamin Whitehouse
U. Virginia

ABSTRACT

Thermostats have the potential for tremendous impact on global energy consumption, but unfortunately they are often not used effectively. In this paper, we present a new system called *ThermoCoach* that improves thermostat usability by giving personalized and actionable recommendations for thermostat use. The system senses human occupancy patterns in a home and emails the household suggested setpoint schedules that can be modified or activated with the click of a button. We performed a randomized controlled trial by deploying over 600 devices in 40 homes from 12 weeks to compare ThermoCoach with a manually programmable thermostat and the Nest learning thermostat. Results indicate that ThermoCoach saves 4.7% more energy than a manually programmable thermostat and 12.4% more energy than the Nest learning thermostat while significantly improving comfort over both approaches.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; D.2.8 [**Software Engineering**]: Metrics—*Complexity Measures, Performance Measures*; H.1.2 [**Information Systems**]: User/Machine Systems

Keywords

Software, Hardware, Infrastructure

1. INTRODUCTION

Thermostats are a simple technology that have the potential for tremendous impact on global energy consumption, but only if used effectively. Residential thermostats control nearly 10% of all energy consumed in the US [1] – approximately 4x the energy consumed by the entire US aviation industry. For decades, people have been advised that programmable thermostats would reduce their home's heating and cooling energy by 10-30% by relaxing the temperature setpoint when the occupants are away or asleep [4], and they are now installed in a third of all US homes [3]. Unfortunately, programmable thermostats are often not used correctly [5] and on average actually increase in energy usage [8]. As a result, the EPA suspended the Energy Star certification program for all programmable thermostats, effective December 31, 2009 [22].

Numerous studies have identified the Achilles' heel of the programmable thermostat to be usability [5, 15]. As a result, several "smart" thermostats have emerged that avoid the need for users to manually program a setback schedule. For example, *reactive thermostats* heat or cool the house in response to motion sensors that detect home occupancy, and are sold by several companies on the market today, including BayWeb and Telconet. More recently, the Nest Learning Thermostat introduced a learning algorithm to automatically create setback schedules based on the user's temperature adjustments. However, studies have found that these new thermostats have introduced new usability problems even as they solve old ones [24, 25]. To date, no technology has been demonstrated to solve the problems with the programmable thermostat.

In this paper, we present a new approach called *ThermoCoach* that improves thermostat usability by giving personalized and actionable recommendations for thermostat use. The system senses and models human occupancy patterns in a home and looks for a discrepancy between the occupancy patterns and the actual heating or cooling energy. Then, it emails the household three suggestions to configure their thermostat: a high comfort option, an energy saving option, and a balanced option. The user can select an option by clicking a button in the email and the recommended schedule is automatically programmed into the thermostat. Before selecting an option, the user is able to improve it based on knowledge about human needs that the system could not identify, such as the need for a warm house when waking up, to condition the house even when empty for plants or pets, or to have low humidity levels when going to sleep. Thus, ThermoCoach simplifies the act of programming a thermostat by only asking the user to select and/or refine a schedule, rather than to produce one from scratch. In contrast to existing smart thermostats, however, it leaves the user in complete control. We hypothesize that ThermoCoach will lead to better comfort and lower energy usage than state of the art approaches.

To evaluate, we performed a 12-week field study in which

BuildSys'15, November 4–5, 2015, Seoul, South Korea..

ACM 978-1-4503-3981-0/15/11 ...$15.00.
DOI: http://dx.doi.org/10.1145/2821650.2821671 .

we compare ThermoCoach with a manually programmable thermostat and the Nest learning thermostat that automatically creates a setpoint schedule for the home. We recruited over 100 people for the study and instrumented 40 homes with over 190 data collection endpoints, over 250 motion sensors, over 135 Bluetooth low energy(BLE) transmitter tags, and 40 Nest thermostats. We performed a randomized trial to compare the three approaches. The results indicate that ThermoCoach saves 4.7% more energy than a manually programmable thermostat and 12.4% more energy than the Nest learning thermostat while significantly improving comfort over both approaches.

2. BACKGROUND

When used correctly, programmable thermostats are one of the most cost effective ways of reducing a home's energy consumption: they cost approximately $50 and can save the average homeowner up to $180 per year [4]. Unfortunately, they are not always used correctly [5] and on average increase in energy usage [8]. There are several important reasons why programmable thermostats are not as effective as originally expected. First, studies have shown that programmable thermostats have poor usability [5, 15]. For example, some people use their thermostat as an on/off switch for the heater while others think it works like a gas pedal: the house will warm more quickly if the temperature setting is raised even higher [17]. This misuse can lead to inefficiency and overheating. Recently, "smart thermostats" have been introduced to simplify or eliminate programming altogether, but these approaches have introduced new usability problems even as they solve old ones [24, 25]. Prior research on making complex interactions intelligible has not addressed the types of long-lived, slowly evolving interactions that characterize human-thermostat interaction.

Even when thermostats are usable, users either cannot or do not always create an energy efficient setback schedule. A household's periods of occupancy change every day and, especially when a household contains multiple people with different schedules, people don't always know the household's occupancy patterns [9]. Getting the schedule incorrect has important consequences because consumers value their comfort more than energy savings [14] and the possibility of discomfort has been shown to discourage people from using setback temperatures [2]. Even if people know an occupancy schedule, their occupancy patterns can change over time and people do not want to continuously adjust their thermostats to optimize performance [11]. In any case, even knowing the occupancy patterns is not enough: studies show that consumers do not understand how their HVAC system works [18] nor how the setback schedules affect energy consumption [13]. Thus, many consumers simply do not have the knowledge or understanding needed to create schedules with the desired balance between comfort and energy usage.

Several recent technologies have tried to address the issue of user capability by eliminating the need to create a schedule. For example, reactive thermostats such as BayWeb and EcoBee use motion sensors or door sensors to control the heating and cooling based on occupancy. However, a recent study found that reactive thermostats save less energy than programmable thermostats in residential buildings, and in 4 out of 8 households actually increase energy usage by up to 10% [12]. This was caused by the long delay required to determine that a home is actually unoccupied, combined with the fact that they must use the highest capacity, lowest-efficiency heating stage to raise the temperature quickly when they detects that occupants suddenly arrive. Several studies of smart thermostats have demonstrated that energy savings and/or comfort can be improved by predicting future occupancy based on the past and present [10,12,20]. However, none of these systems have been demonstrated to address the usability issues that have long plagued thermostat operation in the wild. More recently, The Nest learning thermostat came on the market in 2012 to high acclaim and is now estimated to be installed in over 1 million homes. The Nest tries to automatically learn a setback schedule based on when the user sets back the temperature, but this approach only works if the user actually uses setbacks. Additionally, recent research found that the Nest's autonomous control eventually leads to user disengagement, resulting in inefficient thermostat management and wasted energy, and that frustrated users disable the thermostat's "smart" functionality when it does not perform in a predictable manner [23, 24]. In early work, researchers created the Self-Programming Thermostat [7] that automatically calculates optimal thermostat schedules but, instead of activating one of them autonomously, asks the user to select between a Pareto optimal set of schedules with different comfort/energy tradeoff. The ThermoCoach system described below is based on this same principle. This paper presents the first behavioral study of smart thermostat technology, demonstrating efficacy of the entire system including the human response to the technology.

3. THERMOCOACH OVERVIEW

ThermoCoach monitors occupancy patterns in the home using occupancy sensors and recommends setpoint schedules to the user. The recommended schedules are optimal in the sense that they minimize energy usage for a target comfort level. Three schedules are recommended to the users: high comfort, energy saver, and super energy saver. These three options are presented along with the household's current schedule. The user can "activate" one of the options and the schedule automatically gets programmed into their thermostat. The user can also edit a schedule before activating.

3.1 Hardware Instrumentation

ThermoCoach can operate with any set of occupancy sensors that can infer the *active*, *asleep*, and *away* states of the home, including GPS-enabled smartphones, wearable devices, home security systems, and so on. For this study, we chose to use Z-Wave Motion Sensors and Bluetooth Low Energy (BLE) tags. The hardware kit that was installed into each home included 1 wireless router, up to 4 data collection endpoints, 1 Nest thermostat, up to 6 motion sensors, and up to 4 BLE tags. The wireless router was attached to the home's existing router to avoid needing to ask participants for their WiFi encryption passwords. The data collection *endpoints* were created from a Raspberry PI model B running Debian Linux and a software platform called Piloteur that is designed to ensure reliable data collection from smarthome devices [6]. All hardware was pre-configured to wirelessly connect to each other and operate hands-free upon being powered up in the home. Thus, these hardware kits could easily be installed by a smart home enthusiast. The kits took about 90 minutes to assemble and 60 minutes to

Figure 1: The hardware kit that was installed into each home included up to 4 data collection endpoints, up to 4 BLE tags, up to 6 motion sensors, 1 wireless router, and 1 Nest thermostat.

install.

A 2nd Generation Nest Learning Thermostat was preconfigured in the lab with a unique identifier and password before installation in each home. Each participant was given access to the Nest thermostat's online interface and mobile app. Data from each thermostat was continuously logged approximately once per minute by a Piloteur script using a public RESTful API offered by Nest's servers, recording the current setpoint schedule, the current setpoint, the current state of the HVAC equipment, and states of the Nest-specific features. Because the thermostats could only be queried through Nest's servers, data was lost any time a home's network connection or a Nest's wireless connection was dropped. In some homes, a large amount of the thermostat data was lost because the thermostat had trouble maintaining WiFi connectivity. Two of the homes in the study found that their cooling coils become iced over after the Nest was installed. It was not clear if the Nest thermostat was responsible for the ice but nonetheless one participant asked for the thermostat to be removed and dropped out of the study due to this issue, resulting in only 39 participants completing the study.

To detect whether occupants were in the home or away from the home, we deployed wireless tags on the household members' keychains or the one item they always carry with them when they leave their home. Participants were asked to carry their tags with them when they left the house and to always store their tags in the same location when home. A Bluetooth Low Energy (BLE) tag created by StickNFind was given to every member of the household who might come and go from the house independently. Children who would be home only with another family member were not given tags. The wireless tags were detected by Piloteur endpoints with an IOGear Bluetooth 4.0 USB adapter to record the MAC address and signal strength of any tags within the range of the adapter. The transmitter/receiver pairs typically had a range of 7-8 meters with line of sight. Approximately three endpoints were installed in each home to mitigate data loss due to range limitations. One endpoint was deployed near each exterior door to detect occupants when they entered/left their homes, and one endpoint was installed where the homeowners indicated that they would keep their keys, e.g. on a kitchen counter, or on a bedroom night table.

We deployed Schlage S200HC motion sensors that use passive infrared (PIR) and communicate using the Z-Wave protocol. These wireless, battery-operated devices have a detection area of approximately 9 x 12 meters with a 120° detection angle and a detection range of up to 100 feet (30.5 meters) with line-of-sight. Data from the motion sensors was logged with an Aeon Labs S-2 Z-Stick connected to a Piloteur endpoint, which was installed in a central location in the home to ensure that it was within range of the motion sensors. Up to 4 motion sensors were installed in rooms of high activity such as the kitchen or living room and were placed to face the most active portion of the room, e.g. facing the range or sink in a kitchen or facing the couch in a sitting room. At least one sensor was placed in the transition zone between the active rooms and the bedrooms, such as a hallway or doorway to the bedrooms, to ensure detection of participants as they went to sleep. To avoid false activity detection at night, motion sensors were not installed inside bedrooms facing the bed because preliminary testing found that people do move substantially while sleeping.

3.2 Inferring Daily Occupancy Patterns

ThermoCoach uses occupancy sensors to determine one of three possible states of the home: `active`, `asleep` or `away`. The home is in an *active* state when at least one occupant is present in the home and not sleeping. The home is in the *away* state when all occupants have left the house and it is in the *asleep* state when all occupants of the home have turned into their bedrooms for the night. Based on long-term observations of the home state, ThermoCoach learns the home's *daily occupancy pattern*: the fraction of days that the home is in a given state at a given time of day. These values are stored in the variables $away_i$, $asleep_i$, and $active_i$. In this study, we represent occupancy patterns with 15-minute granularity because the Nest thermostat allows setpoint scheduling at the same granularity, so i falls in the range $[1, 96]$. Other implementations could extend the model to have finer resolution, to differentiate weekdays and weekends, or to infer different patterns for all 7 days of the week. These values can be derived from any occupancy sensors and, in the subsections below, we explain how we derive them from the BLE tags and the motion sensors that were installed for this study.

3.2.1 Inferring Away Patterns

We discarded all BLE data from any day in which the household was observed to be home for less than 10 hours, assuming it was caused by lost data or an anomalous overnight trip. The remaining days are defined to be the set $ValidDays$. We also discarded any BLE tags that were continuously detected in the home, assuming they were left in the home in error. This conservative approach to data cleaning results in fewer days of useful data but avoids having anomalous or erroneous data affect the schedule recommendations. For all remaining BLE data over all days $j \in ValidDays$, we defined the variable BLE_{ij} to be 1 if any tag was detected in the ith 15-minute interval on day j, and 0 otherwise. Then,

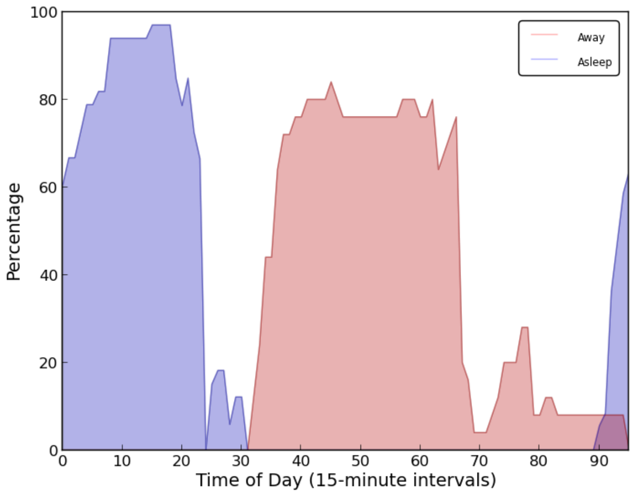

Figure 2: Daily occupancy patterns for one home show that the household typically woke up at 6am, left at 8am, returned at 5pm, and slept before midnight.

we define

$$away_i \quad = \quad \frac{\sum_{j \in ValidDays} BLE_{ij}}{|ValidDays|} \tag{1}$$

In other words, $away_i$ is the fraction of valid days in which at least one occupant was in the home during interval i.

3.2.2 *Inferring Asleep and Active Patterns*

In this study, we use motion sensors to differentiate between the asleep state and the active state. Similar to the BLE data, we defined the variable $Motion_{ij}$ to be 1 if any motion was detected in the ith 15-minute interval on day j, and 0 otherwise. However, motion sensors do not always detect people due to limited coverage or people sitting still. Therefore, we could not make assumptions about awake or asleep states based on the lack of motion sensor data. Instead, we identify a daily *sleep event* $sleep_{ij}$ to be 1 if $Motion_{ij}$ was the last motion detected before night j and 0 otherwise. Similarly, a daily *wake event* $wake_{ij}$ is 1 if $Motion_{ij}$ is the first motion detected after night j and 0 otherwise. The night is defined to be at 4am, so all sleep events occur before 4am and all wake events occur after 4am. A sensor was deployed in the transition area to the bedrooms in order to increase the chance of detecting people at least once as they enter or leave the bedroom the first and last time each night. Sleep and wake events were only identified if the home was occupied.

Detecting sleep was particularly challenging in homes with pets, which often sleep in the active rooms and could trigger motion sensors during the home's asleep state, which could result in unusually late sleep times or unusually early wake times. Homes with pets were identified before the study began and any nights with sleep or wake events within three hours of each other were attributed to pet motion and were discarded. Additionally, motion sensor sometimes failed to detect people or lost data due to communication or hardware failures, which could incorrectly result in earlier sleep times or later wake times. We therefore eliminated any sleep events before 9pm and any wake events after 1pm. As a result, some nights only had a sleep event detected while other nights only had a wake event detected. These conservative data cleaning procedures are designed to eliminate any dubious events to avoid impacting the schedule recommendations, even if that means that the recommendations must be generated based on a much smaller data set. For all remaining sleep events $m \in Sleep$ and wake events $n \in Wake$, we define

$$wakeSleepRatio_i = \frac{\sum_{k \trianglelefteq i,j} sleep_{kj}}{\sum_{k,j} sleep_{kj}} - \frac{\sum_{k \trianglelefteq i,j} wake_{kj}}{\sum_{k,j} wake_{kj}}$$

where $k \trianglelefteq i$ is used to denote that time k is earlier than time i with respect to a day that starts and ends at 4am rather than 12am. In other words, $11pm \trianglelefteq 1am$ is a correct statement while $3am \trianglelefteq 5am$ is not a correct statement. To summarize, $wakeSleepRatio_i$ indicates the ratio of days that the household is asleep rather than awake at time i, given that the household is occupied at time i. It is calculated to be the fraction of times the household was detected going to sleep before time i minus the fraction of times the household was detected waking up before time i. Using this value, we define the fraction of all days that the household is awake or asleep by subtracting out the fraction of days in which the household is away at time i:

$$asleep_i \quad = \quad (1 - away_i) * wakeSleepRatio_i \tag{2}$$
$$active_i \quad = \quad (1 - away_i) * (1 - wakeSleepRatio_i) \tag{3}$$

Figure 2 shows a daily occupancy pattern that is learned from one household. The dark blue region depicts the percentage of days the household is asleep during each 15 minute interval. The lighter red region depicts the percentage of days the household is away. In this home, the household typically woke up around 6am, left the house around 8am, returned home around 5pm, and went to sleep shortly before midnight.

3.3 Setpoint Schedule Analysis

A *setpoint schedule* is defined by a set of times $t_i \in \mathbf{t}$ at which the home's target temperature is changed to a setpoint value $T_i \in \mathbf{T}$. The setpoint temperatures will typically be one of three canonical values: T_{active}, T_{away}, or T_{asleep}, which represent the nominal desired temperatures when the home is active, away, and asleep, respectively. From these parameters, one can derive the sequence of target temperatures $setpoint_i(\mathbf{t}, \mathbf{T})$ for the schedule at every time i. Any setpoint schedule can be evaluated in terms of both energy savings and comfort as described below.

The expected energy savings of a schedule can be approximated as

$$E(\mathbf{t}, \mathbf{T}) = 1 - 0.06 \sum_i (setpoint_i(\mathbf{t}, \mathbf{T}) - min_i(setpoint_i(\mathbf{t}, \mathbf{T})))$$

This formula is derived from the rule of thumb that approximately 1% of energy can be saved for each 1°F setback over an 8-hour period [16]. This formula lets 1 be the normalized amount of energy to condition a home to its minimum temperature all the time. For any setback temperature above the minimum (or below it, in the case of heating), it estimates the energy that would be saved to be proportional to the number of degrees of setback. More accurate estimates can be achieved with more sophisticated equipment models, but even these predictions would be estimates due to unpredictable weather patterns and user overrides. The current approach is an approximation that may cause ThermoCoach

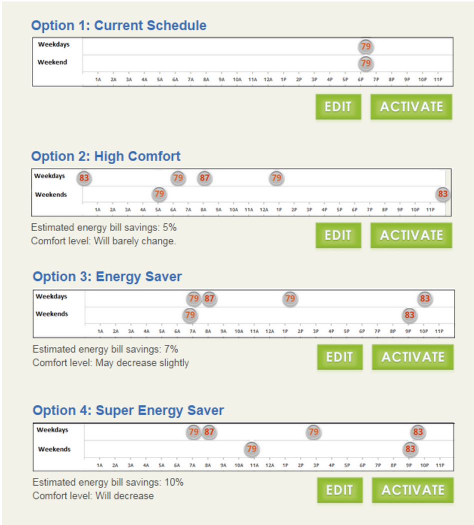

Figure 3: ThermoCoach emailed 4 options to each user: their current schedule, and high comfort schedule, an energy saver schedule, and a super energy saver schedule. Users could *Activate* a schedule by clicking a button and it would automatically be programmed into their thermostat.

Figure 4: If users click the *Edit* button in the ThermoCoach email, they are brought to a Web page that allows them to edit the schedule. While editing, users are informed by a graph of the times when their home is typically active (the gray shaded region right below the schedule).

to recommend a suboptimal schedule and so better energy predictions would only improve the energy savings produced by ThermoCoach. This is a topic for future work.

The expected effect of a schedule on comfort can be defined as

$$\begin{aligned} C(\mathbf{t},\mathbf{T}) \;=\; & \sum_i [(setpoint_i(\mathbf{t},\mathbf{T}) - T_{active}) * active_i \\ & + max(0, setpoint_i(\mathbf{t},\mathbf{T}) - T_{asleep}) * asleep_i \\ & + max(0, setpoint_i(\mathbf{t},\mathbf{T}) - T_{away}) * away_i] \end{aligned}$$

In other words, for every time period i, this equation penalizes the schedule when temperatures are too high for a given state, in proportion to the fraction of days the home is expected to be in that state at time i. No comfort penalty is assessed for setpoints that are too low for a given state. The comfort penalty C is measured in units of *miss time degrees*: the number of degrees of error in the target temperature times the number of minutes of error. This formulation is for the cooling season where setbacks are a higher temperature than the comfort temperature. During the heating season, it would penalize when setbacks are temperatures are too low for a given state.

3.4 Schedule Recommendations

ThermoCoach generates recommended setpoint schedules for each household based on the occupancy patterns inferred from the sensors. There is no perfect setpoint schedule for any home, however, because the occupancy patterns are not exactly the same every day. A home may be occupied one day at a given time and unoccupied the next day at the same time, but a setpoint schedule must condition the home to the same temperature at that time on all days. Thus, there is a fundamental tradeoff between energy usage and comfort: more energy efficient schedules produce lower comfort levels on average, and vice versa. For this reason, ThermoCoach does not recommend only a single schedule, it recommends several schedules that offer a different balance between comfort and energy.

In our current implementation, ThermoCoach creates a weekday/weekend schedule. A weekday schedule is defined by four times $[t_{wake}, t_{leave}, t_{arrive}, t_{sleep}] \in \mathbf{t}$ to model the times that a household first wakes up in the morning, leaves the house, returns to the house, and goes to sleep at night. These times correspond to four temperatures: $[T_{wake}, T_{leave}, T_{arrive}, T_{sleep}] \in \mathbf{T}$. The weekend schedule has only two events: $[t_{wake}, t_{sleep}] \in \mathbf{t}$. ThermoCoach creates a value for T_{wake} and T_{arrive} by searching the home's thermostat logs for the nominal temperature when the home is in the active state. It defines a maximum *setback* of 8°F, and sets $T_{leave} = T_{wake} + setback$ and $T_{sleep} = T_{wake} + 0.5 setback$. In other words, a full setback is used when the household is away and a half setback is used when it is asleep. This 4-state weekday/weekend model of occupancy does not match every household and models that capture more occupancy states could have greater energy saving potential. The Nest thermostat supported schedules with up to 96 states per day. However, we opted to use the conventional 4-state model so that the recommendations would be easy for participants to understand.

Given this definition of a schedule, ThermoCoach can solve the following minimization problem for any *comfort ratio* R:

$$\begin{aligned} \underset{\mathbf{t}}{\text{minimize}} \quad & E(\mathbf{t},\mathbf{T}) \\ \text{subject to} \quad & C(\mathbf{t},\mathbf{T}) \leq R * \sum_i active_i * setback \\ & t_{wake} \trianglelefteq t_{leave} \trianglelefteq t_{arrive} \trianglelefteq t_{sleep} \end{aligned}$$

In other words, ThermoCoach chooses the schedule times such that energy usage is minimized and the comfort penalty is bounded by R times the penalty of having a full setback for the entire expected active time per day. This optimization problem is solved by a brute force solver that compares all possible assignments of $\mathbf{t}$ and completes in less than 10 seconds.

ThermoCoach uses this minimization to define three schedules that are recommended to the user: 1) a *High Comfort* schedule with R=0.1, 2) an *Energy Saver* schedule with R=0.2, and 3) a *Super Energy Saver* schedule with R=0.3. These recommendations are emailed to the homeowners along with the schedule that is currently programmed into their thermostat and the household is asked to choose one of the four options. Each recommendation is annotated with its expected effect on energy savings and comfort levels. Figure 3 shows a sample email generated. Users can `Activate` a recommendation by clicking a link in the email and their thermostat is automatically programmed with that schedule. Otherwise, the users can choose to `Edit` a recommendation, at which point they are redirected to the ThermoCoach Web interface, shown in Figure 4. The Web interface displays the schedule together with a graph of the times when the home is in the active state (in gray below the schedule), as calculated from Equation 3. Informed by this graph, users can edit the setpoints by dragging them with a mouse to accommodate lifestyle, such as the need for a warm house when waking up, or a cool house when sleeping. Users can also switch between the 4 options. The user can accept and save a schedule from the Web interface and, similar to the email, the accepted schedule is automatically programmed into their schedule.

4. EVALUATION

We conducted a 12 week study in 39 homes to evaluate the energy saving potential of ThermoCoach, and to compare with both manually programmable thermostats and the Nest learning thermostat. The design and results of the study are described in the subsections below.

4.1 Recruiting and Deployment

To recruit participants, we distributed over 27,000 flyers through local newspapers and by manually placing doorhangs on doorknobs. The flyers invited people to participate in a study of energy saving technology but did not explain what technology was to be tested. To incentivize participation, the flyers indicated that a Nest learning thermostat would be installed in the home at no cost to the participants. Thus, the participant pool will likely have some bias towards people who are interested in energy or environment and possibly in technology or gadgets. People with interest in the study were instructed to volunteer by completing an online survey about their home and household. We screened for home owners with a detached home and 2 or more members of the household. Renters were not accepted due to the need for permission to install hardware in the home. Participants were asked whether their household included pets or children, the type of cooling equipment (single stage or multi-stage), and how many hours the house was typically unoccupied per day. However, we did not screen for these factors. 135 people volunteered for the study. After screening and drop outs, 40 households were enrolled in the study. During enrollment, all participants were informed through consent forms that the study may involve receive recommendations for energy savings, but they were not informed that the recommendations could involve setpoint schedules. 1 household withdrew from the study due to ice on the cooling coils after the Nest thermostat was installed and the other 39 households completed the study. All homes were located within 30 miles of each other and were subject to similar weather conditions throughout the study.

As the participants were being enrolled, we deployed over 190 data collection endpoints, over 250 motion sensors, over 135 Bluetooth low energy (BLE) transmitter tags, and 40 Nest thermostats in 40 homes. Every home had a unique layout and pattern of use and so, to ensure that all homes were instrumented in an unbiased and consistent fashion, we hired a third-party installation professional who was familiar with neither the study design nor the ThermoCoach system. The installer was given generic guidelines about how to install the sensors and did not receive any unique instructions for a specific home. The installer deployed a generic hardware kit into each home, verified hardware operation on-site using a smartphone, recorded the home's floor plan and the locations of the hardware deployment, and returned the deployment notes along with any unused hardware to the researchers.

4.2 Study Design

Before the study began, the 39 participating households were randomly divided into three groups: Group 1 would use manually programmable thermostats, Group 2 would use the Nest learning thermostat, and Group 3 would use ThermoCoach. Group 3 is the *treatment* group that will receive the experimental treatment: ThermoCoach recommendations. Groups 1 and 2 are the *control* groups against which the energy usage and comfort levels of Group 3 will be compared. All homes had the same hardware installed to avoid any bias caused by the presence of instrumentation, even though the occupancy data would only be used to make recommendations for households in Group 3.

To ensure that the same hardware HVAC control algorithms, and user interfaces were used in all the homes, a Nest thermostat was installed in all 39 homes. However, two features of the Nest were disabled for Groups 1 and 3 to emulate a more conventionally manually programmable thermostat: 1) the *Auto-Schedule* feature that automatically learns a setpoint schedule, and 2) the *Auto-away* feature that automatically activates a setback temperature when its occupancy does not detect human activity. The thermostats were configured at the time of install, participants were sent email instructions explaining which Nest features could be enabled, and the thermostats were monitored for any configuration changes. All other features of the Nest, including energy feedback emails, usage history, and the Web and mobile interfaces, were enabled for all participants. Table 1 summarizes the difference between each group. When the Nest was initially installed, the thermostat was not programmed with a setpoint schedule.

Data about baseline energy usage was collected on all participating households for 6 weeks, at which point we performed an *intervention event*. We sent an email to all 39 households with Nest's monthly energy report, which gives feedback about energy usage, comparison with other households, and occasionally energy saving tips. Additionally, households in Group 3 were emailed ThermoCoach schedule recommendations and were asked to respond to the recommendations within 48 hours. Once the participants selected an option, their thermostats were programmed with chosen schedule within 1 day.

Participants were reminded via email to carry their key fob sensors with them whenever they left home, both after

	Group 1	Group 2	Group 3
Eco-Feedback	√	√	√
Auto-Schedule	-	√	-
Auto-Away	-	√	-
ThermoCoach	-	-	√

Table 1: The 39 homes that completed the study were divided evenly in three groups for a randomized controlled trial. All groups received eco-feedback emails. Group 1 could only manually program their thermostat. Group 2 used Nest's learning algorithms. Group 3 received ThermoCoach suggestions.

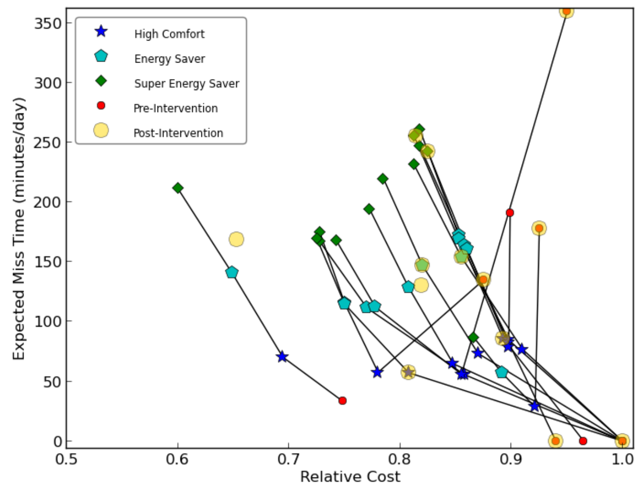

Figure 5: This graph shows the 4 ThermoCoach options for each home, connected by a line. The position of each option indicates its energy usage $E(\mathbf{t}, \mathbf{T})$ on the x-axis (lower is better) and comfort level on the y-axis (lower is better). The schedule that was selected by each home is shown as a yellow dot near the line. 8 of the 13 ThermoCoach homes saved energy by activating a recommended schedule.

the study started and again after the intervention. Otherwise, participants were told to interact naturally with their thermostat. Two weeks after all homes were instrumented, entry interviews were conducted with all 39 households to understand people's interaction with their thermostat, energy needs, and weekly schedules. Another interview was performed after the intervention event to understand how and why thermostat usage may have changed. Finally, a third and final exit interview was performed after the study was complete. Due to space limitations, the findings from these interviews are out of scope for this paper.

4.3 Recommendations and Adoption Rates

Figure 5 illustrates the energy usage and comfort levels of the 4 recommended schedules for each of the 13 homes in Group 3, along with the schedule chosen after intervention. The x-axis is the cost of a schedule relative to having a single setpoint all day long; lower values indicate lower energy usage. The y-axis indicates the average number of minutes a schedule would undercondition the home per day; higher values indicate lower comfort. The ThermoCoach recommended schedules are shown as a star, pentagon, and diamond, respectively. The home's pre-intervention schedule is shown as a red circle and the post-intervention schedule is shown as a yellow circle.

12 of the 13 households in Group 3 responded to the email within forty-eight hours. Of these, 8 households activated a schedule other than their pre-intervention schedule. All 8 of these households significantly improved both the comfort and energy profile of their schedule. Of these 8 homes, 3 choose the High Comfort recommendation or a variation of it; 3 chose the Energy Saver recommendation or a variation of it, and 2 chose the Super Energy Saver recommendation. Four of these households activated recommended schedules *as is*, two adjusted the setpoint times of a recommended schedule by one hour before activating it, and two reduced the setback temperature from 8°F to 2°F before activating it. Two of these 8 homes had already created setpoint schedules before intervention but they activated ThermoCoach recommendations that were more energy efficient. The other six homes that activated recommendations did not have any schedule prior to the intervention.

Of the 5 homes that kept their pre-intervention schedules, 2 simply did not select any option in response to the email. The other 3 all had manually created their pre-intervention schedule. These homes all cooled to a lower temperature at night than during the day, presumably because the bedrooms upstairs were more difficult to cool, making it difficult to sleep. ThermoCoach was not designed for this thermal zoning problem and therefore the recommendations were not workable for these families. This is an opportunity for improvement of ThermoCoach in the future.

5. IMPACT ANALYSIS

To estimate the impact of ThermoCoach on energy usage and comfort, we used a statistical *difference of differences* analysis, also called a panel regression. A panel analysis uses a control group to factor out the effect of uncontrolled factors such as weather. It also compares baseline performance of the control and treatment group to factor out unobservable time-invariant factors, such as a difference in the average furnace efficiency between groups. To perform the panel analysis, we modeled the outcome variable *cost* for any given home i on any given day t as a function of $Treatment_i$ and $Post_t$, which have the value of 1 if home i is in the treatment group (Group 3) and if day t is in the treatment period (post-intervention), respectively. Both variables have the value 0 otherwise. We used a linear model defined in the Guidelines on Measurement and Verification of Behavior-based Energy-efficiency Programs [21]:

$$\begin{aligned} ln(cost_{it}) = {} & \alpha_0 + \alpha_1 * Treatment_i + \alpha_2 * Post_t \\ & + \alpha_3 * Treatment_i x Post_t + \beta_1 * CDD_t \\ & + \beta_2 * CDD_t x Treatment_i + \beta_3 * CDD_t x Post_t \\ & + \beta_4 * CDD_t x Treatment_i x Post_t + v_i + u_i \end{aligned} \quad (4)$$

where $ln(cost_{it})$ is the natural log of the outcome variable in home i on day t and CDD_t (Cooling Degree Days) is the amount of time the outdoor temperature was above 70°F on day t and v_i is a time invariant fixed effect term and u_i is an independent random error term. The values for CDD for the study period were generated by BizEE Software using data from Weather Underground.

Treatment	Control	ATC
Group1	Group 2	+7.79%
Group2	Group 3	-12.39%
Group1	Group 3	-4.686%

Table 2: The table shows the impact (ATC) of the intervention on the energy usage value $E(\mathbf{t}, \mathbf{T})$ (Equation 3.3) for each group compared to a control group. ThermoCoach reduces energy by 4.7% and 12.4% compared to manual programming and Nest learning, respectively.

We use the model above to estimate the Average Treatment Impact (ATC): the average impact of the treatment on the outcome variable *cost*. To do this, we define a *control group* (either Group 1 or 2) and a *treatment group* (Group 3). We create an equation based on the model for every home in both groups on every day of the study, forming a system of equations that can be solved for the α and β coefficients. These coefficients estimate the impact of each term in the model on the outcome variable. The ATC is defined to the sum of all coefficients for the term $Treatment_i x Post_t$:

$$\widehat{ATC} = \hat{\alpha_3} + \hat{\beta_4} * CDD_t \tag{5}$$

This value represents the amount that $ln(cost_{it})$ changed on average after the intervention group for homes in the treatment group. Since small changes in the natural log approximate percentage changes in the original value, the ATC indicates the average percentage change of the outcome variable that is caused by the treatment.

The confidence intervals are then calculated as:

$$\widehat{ATC} = \widehat{ATC} \pm c * standardError(\widehat{ATC}) \tag{6}$$

where c is the $1-\alpha$ percentile of the Degrees of Freedom distribution with 7 independent variables in the model and 1678 data samples. If the confidence interval is strictly above or below 0, we conclude that the ATC is statistically significant for this α value.

Due to the extreme cost of having an electrician to install sub-metering equipment in all 40 homes, we approximate energy impact based on both the real-time temperature settings and the amount of time that the HVAC equipment was powered on, as described below. The energy impact indicated by both metrics are consistent with each other.

5.1 Energy Impact: Part I

To analyze energy impact, the setpoint temperatures at all times (including manual overrides) were extracted from the thermostat operational logs and were used to calculate the value $E(\mathbf{t}, \mathbf{T})$ (Equation 3.3) for each home on each day. This value measures the depth and frequency of setback temperatures, which has been considered a good predictor of average energy usage in the past [16]. We performed panel regression analysis with $E(\mathbf{t}, \mathbf{T})$ as the outcome variable using three different control/treatment pairings to achieve pairwise comparisons between Groups 1, 2 and 3. The results are summarized in Table2 and indicate that ThermoCoach reduced energy usage by 4.7% in comparison to manually programmable thermostats and 12.4% in comparison to the Nest programmable thermostat. These differences are statistically significant with $\alpha = 0.01$. This data supports the hypothesis that ThermoCoach recommendations save more energy than either energy feedback alone or autonomous control.

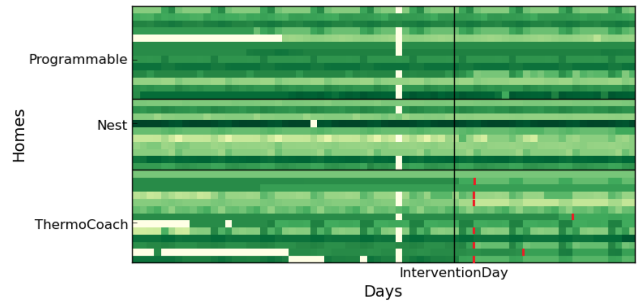

Figure 6: Darker colors indicate higher values of the energy usage $E(\mathbf{t}, \mathbf{T})$ (Equation 3.3) for a given home on a given day. The homes are sorted based on their group. The intervention day is shown as a black vertical line and the times that ThermoCoach recommendations were activated are shown as red lines. The graph indicates that 2 homes using manual programming saved energy, 1 home using Nest learning increased energy usage, and 7 homes using ThermoCoach saved energy after the intervention.

Figure 6 illustrates the average values of $E(\mathbf{t}, \mathbf{T})$ for each home throughout the study period, where each cell $[r, c]$ indicates the average value for household r on day c. Darker shades indicate higher energy usage and blank squares indicate data loss. The households are sorted along the y-axis as Groups 1, 2, and 3. The intervention day is illustrated as a black vertical line. For each home in Group 3, the day on which a recommendation was activated is indicated as a short, red vertical bar. A weekday-weekend pattern and increase in energy usage on weekends is evident in the data. The figure illustrates that ThermoCoach recommendations reduced energy costs significantly for seven out of eight homes that chose recommended schedules at time of *intervention*. One home chose a schedule that was very similar to their pre-intervention schedule and therefore exhibited little change. Energy savings is also apparent for two homes in Group 1. No change in any individual household is apparent in Group 2 for the duration of the study, except in one home that begins to use more energy immediately after receiving the energy feedback email.

5.2 Energy Impact: Part II

In addition to analyzing setback temperatures in Section 5.1, we also analyze energy in terms of *on-time*: the number of minutes per day the day the air-conditioning unit was actively cooling the home. Both setback temperatures and on-time are directly related to energy usage, but on-time is more affected by short-term weather patterns than setback temperatures. Thus, we expect setback temperatures to be a better indicator of long-term energy consumption. The on-time for each home was extracted from the thermostat logs and used as the outcome variable in a pairwise panel analysis with Groups 1, 2, and 3. The ATC values are summarized in Table 4 and are statistically significant with $\alpha = 0.05$. The results indicate that the ThermoCoach intervention decreased the average on-time by 6.2% in comparison to the manual thermostat and by 4.6% in comparison to the Nest thermostat. These results are consistent with the analysis

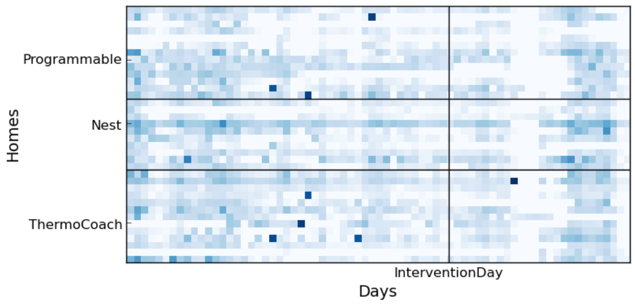

Figure 7: Darker colors indicate higher on-time for a given home on a given day. The data mirrors that for $E(\mathbf{t}, \mathbf{T})$ in Figure 6 while reflecting the effect of daily weather patterns.

in Section 5.1 and supports the hypothesis that ThermoCoach's recommended schedules reduce energy usage more than the other two approaches. Figure 7 shows the on-time for each home throughout the study. The per-home trends in this figure mirror those in Figure 6, and also reflect the effects of local weather patterns residential air conditioning demand.

5.3 Comfort Impact

We measure the human response to our schedule recommendations in terms of manual temperature overrides; comfortable schedules will have fewer manual overrides. Specifically, we use *average degrees changed* as a metric of schedule comfort: the average number of degrees by which the target temperature in a home differs from the scheduled temperature. This metric does not put a heavy weight on small temperature tweaks of 1-2 degrees and instead focuses on deep setbacks that occur even when the occupants are active in the home. We performed pairwise comparison of the average degrees changed in Groups 1, 2, and 3, and the results are presented in Table 3. The results show that the intervention reduced the average temperature change for Group 3 by over 45% compared to manually programmable thermostats or learning thermostats, indicating that people had less need to modify ThermoCoach's recommended schedules than the schedules they manually generated or that were generated by the Nest thermostat. The ATC values are all statistically significant with $alpha = 0.05$.

Figure 8 shows the average degree change per day when compared to the day's setpoint schedule. This data corroborates that very little change occurred in Groups 1 and 2, although one home in Group 2 did significantly reduce the manual overrides after intervention. This home use far less energy than most other homes pre-intervention and suddenly increased energy usage after the energy feedback email. This could be caused by the *deconstructive power* of an energy feedback email that tells a person they are performing above average [19]. The data also shows that several homes in Group 3 did increase the number of overrides after intervention, even though the average temperature change was reduced. This is likely to be caused by the setpoint schedule calling for a setback when the households are active. This affect may be counter balanced by the same deconstructive power of social norms that affected one house in Group 2, producing a decrease on average.

Treatment	**Control**	**ATC**
Group1	Group 2	-3.29%
Group2	Group 3	-45.67%
Group1	Group 3	-47.43%

Table 3: This table lists the impact (ATC) of the intervention on the size of manual overrides. Homes using ThermoCoach reduced manual overrides by over 45% compared to the other two groups after intervention.

Treatment	**Control**	**ATC**
Group1	Group 2	-1.5%
Group2	Group 3	-4.63%
Group1	Group 3	-6.15%

Table 4: This table lists the impact (ATC) of the intervention on on-time: the number of minutes that the air conditioning was on each day. ThermoCoach reduced ADC by over 4.6% compared to manual programming, and 6.2% compared to Nest learning.

6. LIMITATIONS OF THE STUDY

The conclusions of this study are limited by several factors. First and perhaps most importantly, we could not directly measure energy usage in the homes due to hardware costs, installation costs, and data collection challenges. In current work, we are exploring ways to combine the thermostat operation logs with sophisticated equipment models to better analyze energy usage. Second, this pilot study lasted for only one season over a period of three months, and in only one climate zone. A longer and larger study would help understand the seasonal weather effects and long-term human responses, including the possibility that energy savings decline because users revert to their old schedules or that they increase because repeated recommendations over time actually improve adoption rates of the schedule recommendations. Finally, participants may have been more likely to accept ThermoCoach recommendations due to the Hawthorne Effect, i.e. that they knew they were part of a study and may even have felt obliged to select an energy saving option. This effect would also be seen in the control groups.

At the beginning of the study, some Nest thermostats had degenerate schedules with only a single setpoint per day and our first participant survey revealed that some participants did not change the thermostat temperature because they thought that the settings were provided by researchers as part of the study, which was not the case. All participants were sent an email reminding them of what they could and could not do during the study. At time of intervention, the system that processed the recommendation selections malfunctioned and the selections were manually verified from participants of Group 3 over the phone. Eleven out of thirteen participants said they had already made their selection before the phone conversation and so it is unlikely that the adoption rate was biased by the phone call.

Our panel regressions compare pre- and post-intervention outcomes but the Nest thermostats in Group 2 were involved in a continuous learning process that does not necessarily coincide with the date of the intervention event. We aug-

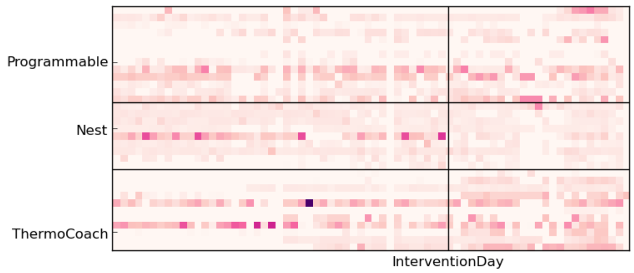

Figure 8: Darker colors indicate large manual overrides for a given home on a given day. Manual overrides changed very little in homes using manual programming, decreased in one home using Nest learning, and had mixed results in homes using ThermoCoach, with the average decreasing significantly.

mented our analysis by performing another regression analysis using only the first 2 weeks and the last 2 weeks of the study, based on the observation that the Nest did not change the schedules before the first 2 weeks or after the last 2 weeks. The new analysis verified that the learning time did not change the impact results or the statistical significance of the results. This is probably because Nest's Auto-schedule did not make major changes to the setpoint schedules in Group 2 at any point in the study, with setbacks of even 3-4°F or more being quite rare. Therefore, the time period used for learning did not affect the impact analysis.

7. CONCLUSIONS

In this paper, we deploy over 600 devices in 40 homes for a 3-month period to perform a randomized controlled trial to compare the ThermoCoach recommendation system with both manually programmable thermostats and the Nest Learning Thermostat. Our results indicate that ThermoCoach saves more energy and produces higher comfort than the other two systems. To our knowledge, this is the first smart thermostat technology that has been demonstrated to be effective while testing the both the technology and the human response to it. ThermoCoach's energy savings of 5-12% over existing technologies are statistically significant and are a first step towards addressing the usability problems that have plagued programmable thermostats. By addressing the limitations of our current implementation, such as addressing homes with two-story zoning issues, we are optimistic that future work can increase the energy savings enabled by the ThermoCoach approach.

8. REFERENCES

[1] Residential energy end-use splits by fuel type. table 2.1.6. U.S. Department of Energy, 2014.

[2] L. A-L, C.-K. A, and E. B. Efficient and inefficient aspects of residential energy behaviour: what are the policy instruments for change? *Energy Policy*, 2006.

[3] Energy Information Administration (EIA). *Residential Energy Consumption Survey, housing characteristics tables, table HC14.5: space heating usage.* Feb. 2011.

[4] Environmental. Environment Protection Agency (EPA). Programmable Thermostats web page. page page;. 2009.

[5] EPA. Summary of research findings from the programmable thermostat market. *Washington, DC: Office of Headquarters*, 2004.

[6] J. Feminella, D. Pisharoty, and K. Whitehouse. Piloteur: A lightweight platform for pilot studies of smart homes. In *BuildSys*. ACM, 2014.

[7] G. Gao and K. Whitehouse. The self-programming thermostat: optimizing setback schedules based on home occupancy patterns. In *BuildSys*, pages 67–72. ACM, 2009.

[8] S. H. *Programmable thermostats*. American Council for an Energy Efficient Economy, Washington, D.C, 2004.

[9] K. J. and Brush. Learning time-based presence probabilities. *Proceedings of Pervasive Computing*, 2011.

[10] C. Koehler, B. D. Ziebart, J. Mankoff, and A. K. Dey. Therml: occupancy prediction for thermostat control. In *UbiComp*, pages 103–112. ACM, 2013.

[11] L. L. A question of control: alternative patterns of room airconditioner use. *Energy and Buildings*, 1992.

[12] J. Lu, T. Sookoor, V. Srinivasan, G. Gao, B. Holben, J. Stankovic, E. Field, and K. Whitehouse. The smart thermostat: Using occupancy sensors to save energy in homes. In *SenSys*, pages 211–224, New York, NY, USA, 2010. ACM.

[13] N. M and P. S. Programmable thermostats that go berserk: taking a social perspective on space heating in Wisconsin. In *Proceedings of the 2000 ACEEE Summer Study on Energy Efficiency in Buildings*, 2000.

[14] P. M. Segmenting residential customers: energy and conservation behaviors. In *Proceedings of the 2008 ACEEE Summer Study on Energy Efficiency in Buildings*, pages 229–41, 2008.

[15] T. Peffer, M. Pritoni, A. Meier, C. Aragon, and D. Perry. How people use thermostats in homes: A review. *Building and Environment*, 46(12):2529–2541, 2011.

[16] A. Plourde. Programmable thermostats as means of generating energy savings: some pros and cons. *Canadian Building Energy End-Use Data and Analysis Centre, Technical Report CBEEDAC 2003-RP*, 1, 2003.

[17] V. R and Sinkkonen. A behavioural model of temperature controller usage and energy saving. *Personal Ubiquitous Computing*, 2005(9):250–9.

[18] K. Rathouse and B. Young. *Market Transformation Programme - Domestic Heating: Use of Controls.* UK Department of Environment, Food and Rural Affairs, 2004.

[19] P. W. Schultz, J. M. Nolan, R. B. Cialdini, N. J. Goldstein, and V. Griskevicius. The constructive, destructive, and reconstructive power of social norms. *Psychological science*, 18(5):429–434, 2007.

[20] J. Scott, A. Bernheim Brush, J. Krumm, B. Meyers, M. Hazas, S. Hodges, and N. Villar. Preheat: controlling home heating using occupancy prediction. In *UbiComp*, pages 281–290. ACM, 2011.

[21] S. Sergici and A. Faruqui. Measurement and verification principles for behavior-based efficiency programs. 2011.

[22] US Environmental Protection Agency. EnergyStar. Programmable Thermostats Suspension Memo., May 2009.

[23] R. Yang and M. W. Newman. Living with an intelligent thermostat: advanced control for heating and cooling systems. In *UbiComp*. ACM, 2012.

[24] R. Yang and M. W. Newman. Learning from a Learning Thermostat: Lessons for Intelligent Systems for the Home. In *UbiComp*, New York, NY, USA, 2013.

[25] R. Yang, M. W. Newman, and J. Forlizzi. Making Sustainability Sustainable: Challenges in the Design of Eco-interaction Technologies. In *CHI*, CHI '14, New York, NY, USA, 2014. ACM.