Building and Environment 221 (2022) 109252

Contents lists available at ScienceDirect

Building and Environment

journal homepage: www.elsevier.com/locate/buildenv

MySmartE – An eco-feedback and gaming platform to promote energy conserving thermostat-adjustment behaviors in multi-unit residential buildings

Huijeong Kim [a,b], Sangwoo Ham [a,b], Marlen Promann [c,d], Hemanth Devarapalli [a,b], Geetanjali Bihani [e], Tatiana Ringenberg [e], Vanessa Kwarteng [f], Ilias Bilionis [b,f], James E. Braun [a,b,f], Julia Taylor Rayz [e], Leigh Raymond [g], Torsten Reimer [h], Panagiota Karava [a,b,*]

[a] Lyles School of Civil Engineering Purdue University, West Lafayette, IN, USA
[b] Center for High Performance Buildings, Ray W. Herrick Laboratories, Purdue University, West Lafayette, IN, USA
[c] Department of Art and Design, The Patti & Rusty Rueff School of Design, Art, and Performance, Purdue University, West Lafayette, IN, USA
[d] Department of Computer Graphics Technology, Purdue Polytechnic Institute, Purdue University, West Lafayette, IN, USA
[e] Department of Computer and Information Technology, Purdue University, West Lafayette, IN, USA
[f] School of Mechanical Engineering, Purdue University, West Lafayette, IN, USA
[g] Department of Political Science and Purdue Climate Change Research Center, Purdue University, West Lafayette, IN, USA
[h] Brian Lamb School of Communication, Purdue University, 100N University Street, West Lafayette, IN, 47907, USA

A R T I C L E I N F O

Keywords:
Eco-feedback
Home energy management
Occupant behavior
Residential buildings
Social energy game
User interface

A B S T R A C T

In this paper, we present a first-time cloud-based eco-feedback and gaming platform that aims to promote energy conserving thermostat-adjustment behaviors in multi-unit residential buildings. To achieve this goal, we introduce a new modeling approach for personalized eco-feedback design integrated with a collaborative social game to assist residents enhance their thermostat use while promoting community-level energy savings. Our modeling framework is integrated into a cloud-based application, MySmartE, with visual (wall-mounted tablet) and voice (Alexa) user interfaces to facilitate behavioral changes in a user-centric approach. The platform is deployed in a multi-unit residential community in Fort Wayne, IN, and data from the field study are used to investigate: (i) how occupants' thermostat behaviors changed after using the MySmartE app; (ii) how users interacted with the app during the game; and (iii) how was users' experience with the developed platform. Despite the heterogeneous characteristics of households, the results from the field study showed the positive effect of the intervention in the thermostat-adjustment behaviors, which resulted in an increase in the indoor temperature during the cooling season compared to the baseline. Findings from the user interaction analysis and post-experiment interviews also revealed the significant potential to nudge households' energy conservation behaviors with the developed platform along with the challenges that should be tackled to derive long-term behavior changes.

1. Introduction

Residential energy consumption is about 20% of the total energy use in the U.S and it is significantly affected by residents' energy-related behavior [1,2]. For example, it has been shown that up to 30% of heating and cooling (HT) energy savings are possible if residents adopt efficient setpoint schedules [3]. This may result in significant cost savings for low-income households spending a larger portion of their income (i.e., about 16%) on home energy costs than average-income households (i.e., 4%) [4]. By developing energy-saving habits, low-income residents can reduce their HC energy bill which is about 40% of their total energy cost [4].

To address opportunities for reducing energy consumption in low-income households, eco-feedback intervention has been examined to motivate residents' energy conservation behavior [5,6]. Considering that energy-related behaviors often derive from unconscious routines and habits [7,8], various types of eco-feedback have been developed to help users understand the relationship between their everyday behavior and energy consumption [9]. Improved energy literacy [5,10–13], however, is not sufficient to transform users' behaviors into sustainable

* Corresponding author. Center for High Performance Buildings, Ray W. Herrick Laboratories, Purdue University, West Lafayette, IN, USA.
E-mail address: pkarava@purdue.edu (P. Karava).

https://doi.org/10.1016/j.buildenv.2022.109252
Received 28 February 2022; Received in revised form 12 May 2022; Accepted 30 May 2022
Available online 9 June 2022

habits as the perceived feedback does not always lead to an actual behavior change [14–16].

To transform users' knowledge into actions, one can refer to technology-mediated persuasion and the Persuasive System Design Model, in particular, that highlight the importance of designing a product considering users' behavior change processes [17]. According to the Fogg Behavior Model [18–20], three key factors are required to successfully induce behavioral changes: 1) motivation; 2) trigger; and 3) ability. With these, the likelihood of behavior changes increases when a fully motivated user encounters a behavior cue that triggers the desired action that is easy to execute.

To design a system that enables such a process, Wendel proposed a four step-iterative product development cycle that accounts for behavior change processes: 1) understanding behavior change strategies; 2) defining target users, actions to change, and the outcome; 3) designing a user interface based on user scenarios, and 4) assessing the outcome and refining the product [21]. For eco-feedback design, the literature review of prior work can help address the first two steps. For Step 1, one can review the list of intervention strategies that have been implemented in previous residential eco-feedback research. Then Step 2 can be addressed by reviewing the residents' general behaviors with programmable thermostats from the existing literature. Once the target behaviors are defined, one can design a field study to implement the developed system in a real-world setting addressing the last two steps. For Step 3, an interface can be designed accordingly to account for users' primary tasks and potential behavior scenarios. Finally, agile development with multiple tests and refinements should be followed to ensure a flawless user experience [22]. After implementing the designed system, Step 4 can be carried out by analyzing its impact on occupants' behavioral change.

1.1. Behavior change strategies applied in residential eco-feedback

To reduce home energy use, information-based intervention strategies and social feedback are adopted in existing eco-feedback research supported by evidence from field experiments [5,10,23–28].

Information-based strategies are widely used to promote user awareness and pro-environmental behaviors. Such strategies enable 1) self-evaluation with a detailed view of current status, 2) self-tracking with access to personal historical data, and 3) estimating the potential behavioral impacts. Various types of information, such as energy consumption or savings, have been provided to enhance users' energy literacy with diverse metrics (e.g., kWh, $, CO_2) [25,29] and time resolutions [15–21]. Such metrics are used to help users understand the gap between their actual and desired behaviors and eventually contribute to increasing the rate of recommendation acceptance [15,19–44]. Historic data allow individuals to keep track of their improvements and conduct a self-comparison [30]. The potential outcomes or impacts from the pro-environmental behaviors can be delivered using various message framing techniques in the context of monetary or environmental units [25].

Normative feedback is widely used for nudging users' behavior changes [31]. It allows residents to compare their energy usage with that of their peers living in similar buildings. Such feedback is widely presented with intuitive graphics (e.g., bar charts) to allow easy comparison. Such socially contextualized feedback design triggers social pressure among peers to nudge residents to save energy [5,29,32,33]. One can strengthen the effect of these descriptive norms, and reduce the risk of "boomerang effects" where high performers regress to average levels of conservation, by including injunctive norms indicating social praise for above performance [34]. Such information influences users to take similar actions, in particular when behavior is shown by similar others and when uncertainty is high. As an additional condition, the descriptive norm should suggest that a vast majority of relevant others are displaying the respective behavior [21].

These norms can also play a key role in social games where various gamification strategies (e.g., competition, cooperation, mixture of both) are used for increasing user engagement [28,33,35–43]. For example, competitive social games often provide leaderboards to compare ones' performance against their competitors [44]. Such comparisons motivate users to reduce more energy than their peers to increase their chance of winning [42]; however, such an approach in housing needs careful consideration to avoid penalizing households with different demographics or with underlying health conditions having an impact on energy consumption. In collaborative social games, the norms are presented to illustrate the groups' collective efforts to achieve a common goal. For instance, Grevet et al. visualized the community's goal achievement by providing a multi-dimensional view of the individuals' achievements [41]. Such cooperative goal structures play a key role in promoting positive user attitudes [45]. The social proof about neighbors' improved actions can increase social conformity and contribute to increasing the cooperation rate in goal-oriented collaborative social games [33,41,46]. The impact of collective behavioral changes can be significant when implemented in affordable housing communities where energy bills are subsidized [47,48].

1.2. Residential programmable thermostat

The installations of households' programmable thermostats are highly encouraged by many energy-efficiency programs and product vendors for effective HC energy management. Growing rapidly, residential programmable thermostat users are expected to reach 40% of the US population by 2021 [48]. With a thermostat's programmable features, users can effectively set up or set-back thermostat setpoints during unoccupied periods using thermostat schedules and significantly reduce their home energy usage. However, despite its benefits and increased adoption, the range of savings varies significantly depending, among others, on users' knowledge and usability of programmable features [49]. For example, an online survey conducted by Pritoni et al. [50] reported that about 40% of programmable thermostat users did not understand how to program thermostat schedules and about 33% maintained a permanent hold mode without using its schedule features. This aligns with the field evaluation report from Cadmus that about 32% of the programmable thermostat users had a single setpoint throughout a season [51]. Similarly, a field experiment conducted by the U.S. Department of Energy with 64 affordable housing apartments showed that only 3.2% and 1.6% of programmable thermostat users applied nighttime and daytime setbacks, respectively [47]. Such discrepancies between the ideal and actual usage often arise due to users' lack of understanding about the energy-conserving options in the product [19], misconceptions about how thermostats work, or the fear of technology failure. For instance, some residents often irrationally adjust their setpoint by assuming the space heating speed is correlated with the level of the thermostat setpoint [52], while some users prefer to keep their thermostat as is to avoid potential misuse [49]. Other underlying factors related to users' lifestyles, health conditions, or properties of the building envelope that limit the usability of programmable thermostats may affect the users' thermostat misuse as well [53]. These consequently lead users to bypass the smart programming features and downgrade their thermostats into traditional manual thermostats [52].

1.3. Aims of the study

In this paper, we present a first-time cloud-based eco-feedback and gaming platform (MySmartE app) that aims to promote energy conserving thermostat adjustment behaviors in multi-unit residential buildings. Our study introduces a new methodology for personalized eco-feedback design integrated with a collaborative social game to assist residents enhance their thermostat use while promoting community-level energy-savings. Our feedback mechanism is based on 1) a new energy conservation behavior score that mathematically quantifies users' efforts in reducing HC energy usage on a 0–100 scale; 2) an

[61] K. Gopalakrishnan, B. Hedayatnia, Q. Chen, A. Gottardi, S. Kwatra, A. Venkatesh, R. Gabriel, D. Hakkani-t, A.A. Ai, Topical-Chat : towards Knowledge-Grounded Open-Domain Conversations Ur, (n.d.).
[62] G. Bihani, J.T. Rayz, Fuzzy classification of multi-intent utterances, Lect. Notes Networks Syst. 258 (2022) 37–51, https://doi.org/10.1007/978-3-030-82099-2_4.
[63] W. Minker, S. Bennacef, SPEECH AND HUMAN-MACHINE DIALOG, Springer Sci. Bus. Media, 2004.
[64] Serverless architectures with AWS Lambda overview and best practices, Feb. 04, (n.d.), https://aws.amazon.com/lambda/, 2022, https://d1.awsstatic.com/whitepapers/serverless-architectures-with-aws-lambda.pdf.
[65] J.T. Rayz, S. Tazhibayeva, P. Karava, Back to Classics: Controlling Smart Thermostats with Natural Language… with Personalization, Springer International Publishing, 2020, https://doi.org/10.1007/978-3-030-40814-5_23.
[66] J. Saldana, The Coding Manual for Qualitative Researchers, SAGE Publications Ltd, 2021.
[67] H. Allcott, T. Rogers, The short-run and long-run effects of behavioral interventions: experiment evidence from energy conservation, Am. Econ. Rev. 104 (2012) 3003–3037, https://doi.org/10.1257/aer.104.10.3003.
[68] H.E. Moore, J. Boldero, Designing interventions that last: a classification of environmental behaviors in relation to the activities, costs, and effort involved for adoption and maintenance, Front. Psychol. 8 (2017), https://doi.org/10.3389/fpsyg.2017.01874.
[69] P. Candio, K.B. Pouwels, D. Meads, A.J. Hill, L. Bojke, C. Williams, Modelling decay in effectiveness for evaluation of behaviour change interventions: a tutorial for public health economists, Eur. J. Health Econ. (2021), https://doi.org/10.1007/s10198-021-01417-7.

algorithm for personalized action recommendations along with their potential impacts. To help users easily turn their perceived feedback into actions, we developed a cloud-based app, MySmartE, that integrates the functionality of a smart thermostat. All features are implemented in visual (wall-mounted tablet) and voice (Alexa) user interfaces that facilitate behavioral changes in a user-centric approach. Based on these features, the platform is designed to help residents easily apply energy-saving tips to their setpoint schedule, maintain pro-environmental behaviors, and eventually form sustainable energy-saving habits. To evaluate the platform, we conducted a field study in a multi-residential community aiming to address the following research questions:

RQ1: How do occupant behaviors change after using the MySmartE app? Does it improve households' thermostat settings and setpoint selections compared to the baseline period?
RQ2: How do users interact with MySmartE app ?
RQ3: How was users' experience with the developed platform?

In Section 2, we present an overview of our research test-bed and field study design. In Section 3, we discuss major thermostat adjustment behavior trends of residents observed in the absence of any behavioral intervention (baseline period). Section 4 includes the eco-feedback and social game design methodology. Section 5 presents the overall design methodology of the software platform (MySmartE) with the visual and voice eco-feedback devices. The field deployment of the MySmartE (intervention period) and the data analysis methodology are presented in Section 6. In Section 7, we present results on residents' behavior, interactions with the app, and user experience with the developed platform. Finally, conclusions of the study and its limitations are provided in Section 8.

2. Research test-bed and field study overview

2.1. Research test-bed

Our test-bed is an affordable housing community located in Fort Wayne, Indiana (Fig. 1). The community constitutes a pair of two-story multi-family buildings occupied by low-income residents whose utility bills are subsidized. Each building consists of 22 apartment units (14, 2-bed and 8, 3-bed units) located on two floors. The 1st floor includes 10 residential units and a leasing office while the 2nd floor has 12 units, with all the corners occupied by 3 bedroom units.

Each apartment is equipped with a WiFi-enabled smart thermostat (i.e., Ecobee4[1]) that controls the unit's dedicated electric heat pump and air-handling unit. Through the thermostat, we measure the system settings (e.g., heat, cool, off), temperature setpoint (i.e., heating and cooling), indoor temperature, relative humidity, and occupancy. Sub-circuit smart power meters (i.e., GreenEye Monitor[2]) are also installed on each unit to measure disaggregated energy usage (heat pump, air handler, plug load, water heater, etc.). All sensor data are collected via each vendor's web-based application program interface (API) in 5-min intervals. Hourly weather information such as outdoor air temperature and relative humidity is downloaded from National Oceanic and Atmospheric Administration (NOAA[3]), which provides public weather station data located at Fort Wayne International Airport [54]. Eco-feedback devices including a wall-mounted tablet and Amazon, Alexa are installed in each household during the intervention period discussed in detail in the next section.

2.2. Field study overview

The field study includes two phases: 1) a baseline period (from December 1st 2019 to December 31st, 2020) and 2) an intervention period (from January 1st to August 1st, 2021). The baseline period refers to the time frame without any behavioral intervention, while the intervention period starts after the installation of our eco-feedback devices. During the baseline period, we collected smart thermostat and power meter data to understand baseline user behaviors (Section 3) and guide the design of the feedback mechanism (Section 4) and its integration into the software platform (Section 5). During the intervention period, we implemented MySmartE and collected user interaction data in addition to the smart thermostat data (Sections 6 and 7).

3. Baseline user behavior

In this section, we use smart thermostat data collected during the baseline period of the field study to identify major adjustment behavior trends observed during typical heating and cooling months (i.e., heating months: December, January, February; cooling months: June, July, August). The available baseline data for thermostat settings from 26 households are visualized in Fig. 2. Based on users' baseline thermostat usage, we observe 1) low adoption rate of smart programming features and 2) frequent use of constant setpoint temperatures. Such observations aligned with reports from the literature that address general usability issues of smart thermostats [49].

Specifically, the majority of households relied only on 'hold' and 'off' thermostat settings and did not use any schedule features. This is illustrated in the heatmaps of Fig. 2 (a) and (b). These heatmaps show the percentage of each thermostat setting used by different households. During the heating season, about 70% of households manually adjusted temperatures in hold mode (e.g., lowering or increasing setpoints) instead of programming thermostat setpoints based on their schedule for more than 90% of the time. A similar trend was observed in 70% of the households during the cooling months with the increased use of the "off" mode.

Also, the majority of households used constant setpoint temperatures throughout both seasons. For instance, the interquartile of the heating setpoints selected by 70% of households are distributed within the narrow band of 2 °C (Fig. 2 (a), right). Similarly, 80% of the households that mainly used hold/off mode (i.e., above 90% during cooling months) selected their cooling setpoints within the interval of 2 °C where about 50% of them were distributed within 1 °C.

4. Personalized eco-feedback and social energy game

In this section, we build on the analysis presented in Section 3, to introduce our methodology for personalized eco-feedback design to assist users effectively set-back or set-up thermostat setpoints using programmable features. We focus on two types of feedback information: 1) energy scores to inform users about their current status based on their thermostat setpoint selections (Section 4.1) and 2) personalized action recommendations with their potential impacts (Section 4.2). A collaborative social game is then built upon the custom eco-feedback elements to promote community-level energy efficiency (Section 4.3).

4.1. Personal energy score

We define a personal energy score that mathematically quantifies users' efforts in reducing HC energy usage on a 0–100 scale. This score is presented through the app every day to inform users about their progress each calendar week (i.e., from Monday to Sunday). For the calculation, we use a household's time-series thermostat data collected each week.

To derive the energy score, we first define an energy conservation behavior (ECB) metric to evaluate the household's thermostat setpoint observed in each timestamp. The metric is also scaled between 0 and 100

[1] https://www.ecobee.com.
[2] https://www.brultech.com/greeneye.
[3] https://www.ncei.noaa.gov/access/weather-climate-tools.

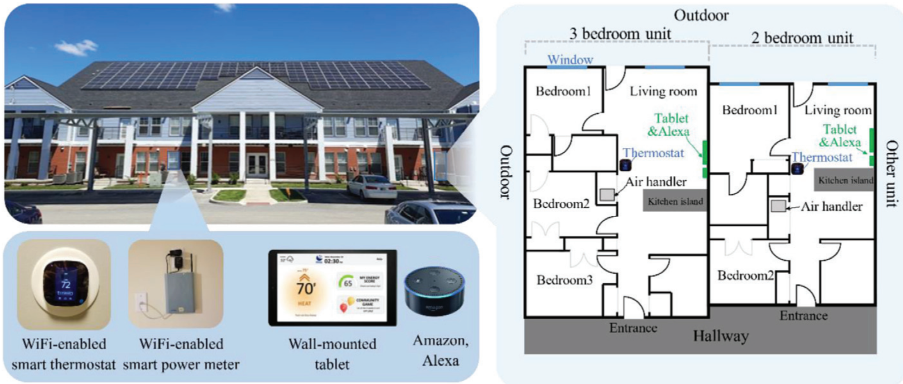
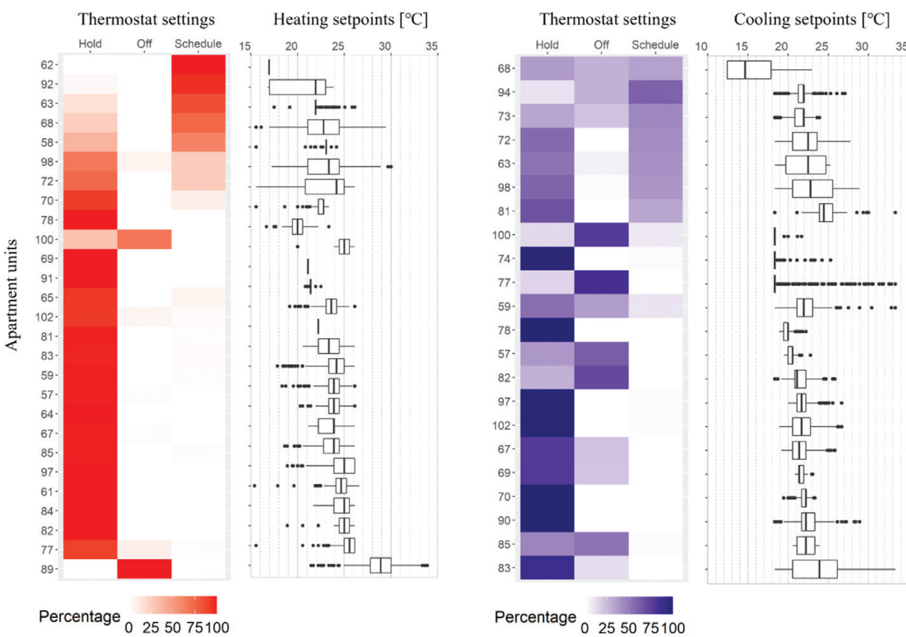


Fig. 1. Overview of the research test-bed. (a) Left top: Exterior view of the building; (b) Left bottom: smart sensors and eco-feedback devices; (c) Right: Floor layouts of 2bed-/3bedroom units.



(a) Thermostat settings during heating season (12/1/2019 - 2/28/2020)

(b) Thermostat settings during cooling season (6/1 - 8/31/2020)

Fig. 2. Thermostat settings during the baseline period.

[4] J.F. Eisenberg, Weatherization Assistance Program Technical Memorandum Background Data and Statistics on Low-Income Energy Use and Burdens, 2014.

[5] J. Froehlich, L. Findlater, J. Landay, C. Science, The Design of Eco-Feedback Technology, 2010.

[6] M. Costanzo, D. Archer, E. Aronson, T. Pettigrew, Energy conservation behavior. The difficult path from information to action, Am. Psychol. 41 (1986) 521–528, https://doi.org/10.1037/0003-066X.41.5.521.

[7] M. Baborska-Narożny, F. Stevenson, Service controls interfaces in housing: usability and engagement tool development, Build. Res. Inf. 47 (2019) 290–304, https://doi.org/10.1080/09613218.2018.1501535.

[8] M. Baborska-Narożny, F. Stevenson, M. Grudzińska, Overheating in retrofitted flats: occupant practices, learning and interventions, Build. Res. Inf. 45 (2017) 40–59, https://doi.org/10.1080/09613218.2016.1226671.

[9] T. Hargreaves, M. Nye, J. Burgess, Keeping energy visible? Exploring how householders interact with feedback from smart energy monitors in the longer term, Energy Pol. 52 (2013) 126–134, https://doi.org/10.1016/j.enpol.2012.03.027.

[10] R. Sussman, M. Chikumbo, Behavior Change Programs: Status and Impact, 2016.

[11] K. Ehrhardt-Martinez, K.A. Donnelly, J.A. "Skip" Latiner, Advanced Metering Initiatives and Residential Feedback Programs: A Meta-Review for Household Electricity-Saving Opportunities, 2010.

[12] R.K. Jain, J.E. Taylor, G. Peschiera, Assessing eco-feedback interface usage and design to drive energy efficiency in buildings, Energy Build. 48 (2012) 8–17, https://doi.org/10.1016/j.enbuild.2011.12.033.

[13] A. Mäkivierikko, Designing energy feedback for a local social network, Energy Build. 199 (2019) 88–101, https://doi.org/10.1016/j.enbuild.2019.06.038.

[14] E. Shove, Beyond the ABC: climate change policy and theories of social change, Environ. Plann. 42 (2010) 1273–1285, https://doi.org/10.1068/a42282.

[15] N. Murtagh, S. Badi, Y. Shi, S. Wei, W. Yu, Living with air-conditioning: experiences in dubai, chongqing and London, Build. Cities 3 (2022) 10–27, https://doi.org/10.5334/bc.147.

[16] R. Hitchings, Understanding air-conditioned lives: qualitative insights from Doha, Build. Cities 3 (2022) 28–41, https://doi.org/10.5334/bc.155.

[17] H. Oinas-Kukkonen, M. Harjumaa, Persuasive systems design: key issues, process model, and system features, Commun. Assoc. Inf. Syst. 24 (2009) 485–500, https://doi.org/10.17705/1cais.02428.

[18] B. Fogg, A behavior model for persuasive design, ACM Int. Conf. Proceeding Ser. 350 (2009), https://doi.org/10.1145/1541948.1541999.

[19] J. Pierce, D.J. Schiano, E. Paulos, Home, habits, and energy: examining domestic interactions and energy consumption, Conf. Hum. Factors Comput. Syst. - Proc. 3 (2010) 1985–1994, https://doi.org/10.1145/1753326.1753627.

[20] J.Y. Mao, K. Vredenburg, P.W. Smith, T. Carey, The state of user-centered design practice, Commun. ACM 48 (2005) 105–109, https://doi.org/10.1145/1047671.1047677.

[21] Stephen Wendel, Designing for Behavior Change, 2013, https://doi.org/10.1017/CBO9781107415324.004.

[22] P. Abrahamsson, O. Salo, J. Ronkainen, J. Warsta, Agile Software Development Methods: Review and Analysis, CoRR. Abs/1709, 2017, 0.

[23] G. Ma, J. Lin, N. Li, Longitudinal assessment of the behavior-changing effect of app-based eco-feedback in residential buildings, Energy Build. 159 (2018) 486–494, https://doi.org/10.1016/j.enbuild.2017.11.019.

[24] J. Froehlich, S. Patel, J.A. Landay, L. Findlater, M. Ostergren, S. Ramanathan, J. Peterson, I. Wragg, E. Larson, F. Fu, M. Bai, The design and evaluation of prototype eco-feedback displays for fixture-level water usage data, Proc. 2012 ACM Annu. Conf. Hum. Factors Comput. Syst. - CHI '12. (2012) 2367, https://doi.org/10.1145/2207676.2208397.

[25] O.I. Asensio, M.A. Delmas, Nonprice incentives and energy conservation, Proc. Natl. Acad. Sci. U.S.A. 112 (2015) E510–E515, https://doi.org/10.1073/pnas.1401880112.

[26] M. Harding, A. Hsiaw, Goal setting and energy conservation, J. Econ. Behav. Organ. 107 (2014) 209–227, https://doi.org/10.1016/j.jebo.2014.04.012.

[27] K. Matsui, H. Ochiai, Y. Yamagata, Feedback on electricity usage for home energy management: a social experiment in a local village of cold region, Appl. Energy 120 (2014) 159–168, https://doi.org/10.1016/j.apenergy.2014.01.049.

[28] D. Foster, S. Lawson, M. Blythe, P. Cairns, Wattsup?, in: Proc. 6th Nord. Conf. Human-Computer Interact. Extending Boundaries - Nord. '10 ACM Press, New York, New York, USA, 2010, p. 178, https://doi.org/10.1145/1868914.1868938.

[29] R.K. Jain, R. Gulbinas, J.E. Taylor, P.J. Culligan, Can social influence drive energy savings? Detecting the impact of social influence on the energy consumption behavior of networked users exposed to normative eco-feedback, Energy Build. 66 (2013) 119–127, https://doi.org/10.1016/j.enbuild.2013.06.029.

[30] C. Wilson, H. Dowlatabadi, Models of decision making and residential energy use, Annu. Rev. Environ. Resour. 32 (2007) 169–203, https://doi.org/10.1146/annurev.energy.32.053006.141137.

[31] D.E. Polk, C.M. King, K. Heller, Community-based interventions, Cambridge Handb. Psychol. Heal. Med. 93 (2014) 344–348, https://doi.org/10.1017/CBO9780511543579.075, second ed.

[32] A. Laskey, O. Kavazovic, Opower, XRDS Crossroads, ACM Mag. Students. 17 (2011) 47, https://doi.org/10.1145/1961678.1961687.

[33] D. Wemyss, R. Castri, F. Cellina, V. De Luca, E. Lobsiger-Kägi, V. Carabias, Examining community-level collaborative vs. competitive approaches to enhance household electricity-saving behavior, Energy Effic. 11 (2018) 2057–2075, https://doi.org/10.1007/s12053-018-9691-z.

[34] Schultz, et al., The constructive, destructive, and reconstructive power of social norms, Psycological Sci 18 (2007) 429–434.

[35] H.P. Das, I.C. Konstantakopoulos, A.B. Manasawala, T. Veeravalli, H. Liu, C. J. Spanos, Segmentation analysis in human centric cyber-physical systems using graphical Lasso. http://arxiv.org/abs/1810.10533, 2018.

[36] I.C. Konstantakopoulos, A.R. Barkan, S. He, T. Veeravalli, H. Liu, C. Spanos, A deep learning and gamification approach to improving human-building interaction and energy efficiency in smart infrastructure, Appl. Energy 237 (2019) 810–821, https://doi.org/10.1016/j.apenergy.2018.12.065.

[37] J. Novak, M. Melenhorst, I. Micheel, C. Pasini, P. Fraternali, A.E. Rizzoli, Integrating behavioural change and gamified incentive modelling for stimulating water saving, Environ. Model. Software 102 (2018) 120–137, https://doi.org/10.1016/j.envsoft.2017.11.038.

[38] P. Ponce, A. Meier, J.I. Méndez, T. Peffer, A. Molina, O. Mata, Tailored gamification and serious game framework based on fuzzy logic for saving energy in connected thermostats, J. Clean. Prod. 262 (2020), https://doi.org/10.1016/j.jclepro.2020.121167.

[39] S. Chadoulos, I. Koutsopoulos, G.C. Polyzos, Mobile apps meet the smart energy grid: a survey on consumer engagement and Machine Learning applications, IEEE Access (2020) 8, https://doi.org/10.1109/ACCESS.2020.3042758.

[40] B. Xia, H. Ming, K.Y. Lee, Y. Li, Y. Zhou, S. Bansal, S. Shakkottai, L. Xie, EnergyCoupon: a case study on incentive-based demand response in smart grid, E-Energy 2017 - Proc. 8th Int. Conf. Futur. Energy Syst. (2017) 80–90, https://doi.org/10.1145/3077839.3077862.

[41] C. Grevet, J. Mankoff, S.D. Anderson, Design and evaluation of a social visualization aimed at encouraging sustainable behavior, Proc. Annu. Hawaii Int. Conf. Syst. Sci. (2010) 1–8, https://doi.org/10.1109/HICSS.2010.135.

[42] J.E. Petersen, V. Shunturov, K. Janda, G. Platt, K. Weinberger, Dormitory residents reduce electricity consumption when exposed to real-time visual feedback and incentives, Int. J. Sustain. High Educ. 8 (2007) 16–33, https://doi.org/10.1108/14676370710717562.

[43] P. Fraternali, S. Herrera, J. Novak, M. Melenhorst, D. Tzovaras, S. Krinidis, A. E. Rizzoli, C. Rottondi, F. Cellina, EnCOMPASS - an integrative approach to behavioural change for energy saving, GIoTS 2017 - glob. Internet Things Summit, SAVE Proc. (2017) 1–6, https://doi.org/10.1109/GIOTS.2017.8016256.

[44] J. Novak, M. Melenhorst, I. Micheel, P. Fraternali, C. Pasini, Stimulating energy-saving behaviour through eco-feedback, adaptive gamification and personalized recommendations, 5th Eur. Conf. Behav. Energy Effic. (BEHAVE 2018) (2018) 6–7.

[45] A. Järvinen, Game design for social networks: interaction design for playful dispositions, Proc. 2009 ACM SIGGRAPH Symp, Video Games, Sandbox '09. 1 (2009) 95–102.

[46] V. De Luca, R. Castri, The social power game: a smart application for sharing energy-saving behaviours in the city, proc. Avi 2014 work. Foster, Smart Energy Appl. through Adv. Vis. Interfaces (2014) 1–4, https://doi.org/10.1145/2598153.2602224.

[47] O. Sachs, V. Tiefenbeck, C. Duvier, A. Qin, K. Cheney, C. Akers, K. Roth, Field Evaluation of Programmable Thermostats, 2012.

[48] K. Lusson, SMART THERMOSTATS :Assessing Their Value in Low-Income Weatherization Programs, 2020.

[49] T. Peffer, M. Pritoni, A. Meier, C. Aragon, D. Perry, How people use thermostats in homes: a review, Build. Environ. 46 (2011) 2529–2541, https://doi.org/10.1016/j.buildenv.2011.06.002.

[50] M. Pritoni, J.M. Woolley, M.P. Modera, Do occupancy-responsive learning thermostats save energy? A field study in university residence halls, Energy Build. 127 (2016) 469–478, https://doi.org/10.1016/j.enbuild.2016.05.024.

[51] C. Aarish, M. Perussi, A. Rietz, D. Korn, Evaluation of the 2013–2014 Programmable and Smart Thermostat Program, Inf. Publ. Cadmus Gr., 2015.

[52] M. Pritoni, A.K. Meier, C. Aragon, D. Perry, T. Peffer, Energy efficiency and the misuse of programmable thermostats: the effectiveness of crowdsourcing for understanding household behavior, Energy Res. Social Sci. 8 (2015) 190–197, https://doi.org/10.1016/j.erss.2015.06.002.

[53] K.J. Lomas, S. Oliveira, P. Warren, V.J. Haines, T. Chatterton, A. Beizaee, E. Prestwood, B. Gething, Do domestic heating controls save energy? A review of the evidence, Renew. Sustain. Energy Rev. 93 (2018) 52–75, https://doi.org/10.1016/j.rser.2018.05.002.

[54] National Oceanic and Atmospheric Administration (NOAA), Climate Data Online Data Tools, 2020. https://www.ncdc.noaa.gov/cdo-web/datatools.

[55] L.J. Schoen, P.F. Alspach, E.A. Arens, R.M. Aynsley, R. Bean, J. Eddy, D. Int-hout, E.E. Khalil, P. Simmonds, J.L. Stoops, S.C. Turner, W.F. Walter, R.L. Hall, J. R. Anderson, C.S. Barnaby, S.F. Bruning, J.A. Clark, S.J. Emmerich, J.M. Ferguson, Thermal environmental conditions for human occupancy, ASHRAE Stand. 2013 (2013).

[56] star energy, A Guide to Energy-Efficient Heating and Cooling Change the World , ENERGY, Energy, 2015.

[57] B.J. Andreoni, Cooperation in public-goods experiments : kindness or confusion? Am. Econ. Rev. 85 (1995) 891–904.

[58] K. Dmitrieva, Saving energy with pleasure: designing EnergyRace web application, 2014, Int. Conf. Collab. Technol. Syst. CTS 2014 (2014) 607–610, https://doi.org/10.1109/CTS.2014.6867633.

[59] M.A. Zehir, K.B. Ortac, H. Gul, A. Batman, Z. Aydin, J.C. Portela, F.J. Soares, M. Bagriyanik, U. Kucuk, A. Ozdemir, Development and field demonstration of a gamified residential demand management platform compatible with smart meters and building automation systems, Energies 12 (2019) 1–18, https://doi.org/10.3390/en12050913.

[60] E. Yoeli, M. Hoffman, D.G. Rand, M.A. Nowak, Powering up with indirect reciprocity in a large-scale field experiment, Proc. Natl. Acad. Sci. Unit. States Am. 110 (2013) 10424–10429, https://doi.org/10.1073/pnas.1301210110.

- Despite the heterogeneous characteristics of households, we observed a positive effect of the intervention during the cooling season for all 12 units for which baseline and intervention data were available. These households increased their room air temperature for the sleep and home mode during the intervention period in comparison to their baseline behavioral data. Considering that we have normalized the weather effect in this comparison, we can infer that the changes in the indoor temperatures can be largely attributed to users' behavior changes, i.e. the improvement in users' thermostat settings that increased their usual cooling setpoints.
- From the user interaction analysis, we found that daily popup notifications played a key role in initiating user interactions with the app. Once triggered, users often navigated to the other pages in the app. This contributed to maintaining a certain rate of interaction during the game intervention; however, the rate decreased over time.
- The user interaction with the voice interface was lower than that of the visual interface throughout the experiment. Users tended to use simple and intuitive commands that are related to their everyday life such as querying the temperature information or controlling the thermostat.
- The result from post-experiment interviews with 13 residents revealed that users were overall satisfied with the app and the information provided. The interviews confirmed that the social game popups contributed to increasing user engagement by using social presence avatars to visualize the community's effort. The interview also pointed out directions for additional improvements. To enhance the stability of the app as well as its usability, continuous software updates are required for maintenance. Also, to support the communal goal achievements and user motivations in the game, the shared goals and rewards need to be optimized to account for user characteristics and competencies.

It should be noted that our software platform is developed for the specific context described in the paper. That is a multi-unit residential building with electric heat pumps installed on each unit. Also, our app integrates the functionality of a specific smart thermostat. However, our modularized software infrastructure through API integration can be extended in the future to different contexts. Also, our field study was conducted in an uncontrolled environment to explore the longitudinal effects of feedback intervention on diverse households in an affordable housing community. Considering the various external factors that hindered the experiment, the observed outcome is valuable in that we induced positive user behavioral changes in real-world scenarios in the presence of many limitations. An important follow-up challenge, of course, as with all energy conservation interventions is the issue of "decay" [67], the documented risk that user interactions with the technology, and energy conservation behaviors, may decrease over time in the absence of continued interventions. To further investigate the performance of the software platform and address the aforementioned limitations, we will conduct experiments in shoulder and heating season.

CRediT authorship contribution statement

Huijeong Kim: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sangwoo Ham:** Writing – review & editing, Software, Methodology, Investigation, Data curation, Conceptualization. **Marlen Promann:** Writing – original draft, Visualization, Methodology, Investigation. **Hemanth Devarapalli:** Writing – original draft, Software, Methodology, Investigation. **Geetanjali Bihani:** Writing – original draft, Software, Methodology, Investigation. **Tatiana Ringenberg:** Writing – original draft, Software, Methodology, Investigation. **Vanessa Kwarteng:** Methodology, Investigation. **Ilias Bilionis:** Writing – review & editing, Supervision, Software, Methodology, Funding acquisition, Conceptualization. **James E. Braun:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Julia Taylor Rayz:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Leigh Raymond:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Torsten Reimer:** Writing – review & editing, Funding acquisition, Conceptualization. **Panagiota Karava:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was funded by the National Science Foundation under grant number 1737591 and supported by the Big Ideas Challenge program at Purdue University.

Appendix A. Occupancy logic

Logic was designed to label households' occupancy in three states (i.e., home, away, sleep). The label "home" refers to the state when residents are at home and awake; "sleep" is when users are home and sleeping, and "away" represents the vacancy of the apartment unit. These occupancy states are inferred from the thermostat and power meter data by following the two steps: 1) preprocessing binary occupancy data and 2) labeling home, away, and sleep states.

First, the hourly binary occupancy is derived using two types of binary occupancy data detected from the motion sensor in thermostats and appliance usage. The former is collected from the smart thermostat's built-in infrared motion sensor that detects any movement every 5 min. For each hour, the state is considered as occupied if any motion is detected within the timeframe (i.e., occupied: 1, unoccupied: 0). The latter is inferred from the submeter power data that captures users' presence by detecting users' appliance use activity based on power spikes of the appliances. For this, we calculated hourly electricity energy consumed in the kitchen area (e.g., dishwasher, range, microwave, plug loads), hallway (i.e., dryer), and the plug loads from bathrooms, the living room, and bedrooms. For each timestamp, we calculated power usage differentials for each appliance and detected the interaction moment when it exceeds a threshold. Each appliance's threshold is calculated as an average of the differentials observed during the whole baseline period. For each interaction, the state is considered as occupied.

Then, the three-state occupancy state labels are applied to the binary occupancy data. First, the away state is assigned to the unoccupied hours that continue for more than two consecutive hours. The rest of the hours are labeled as the home state. Once the home and away states are assigned, the sleep state is overridden by the two-state occupancy data based on the users' sleep schedule programmed in their thermostat. The sleep states are labeled over users' sleep timeframe if any home state is observed within this timespan. Each user's sleep timeframe includes one's programmed sleep schedule in the thermostat with 90-min offsets applied before the start and after the end of the scheduled hours, respectively. These offsets are applied to capture user activities before going to bed and after awakening.

References

[1] DOE, Overview of Existing and Future Residential Use Cases for Connected Thermostats, 2016.
[2] EIA, Residential Energy Consumption Survey (RECS), 2015.
[3] J. Rotondo, R. Johnson, N. Gonzalez, A. Waranowski, C. Badger, N. Lange, E. Goldman, R. Foster, Overview of Existing and Future Residential Use Cases for Connected Thermostats, 2016.

to denote the worst possible behavior with 0 and the best possible with 100. The ECB is a function of time-series data from 1) system operation mode O_t ($O_t \in \mathscr{O}$, $\mathscr{O}$ = {"heat", "cool", "off"}), 2) thermostat setpoint used in the mode $T_{O_t,t}$ ($T_{O_t,t} \in \mathscr{T}, \mathscr{T} = [60, 90]$ in °F), and 3) occupancy state S_t ($S_t \in \mathscr{S}$, $\mathscr{S}$ = {"home", "away", "sleep"}) at timestamp t. The three-state occupancy state S_t is derived using logic in Appendix. A with the inputs from the thermostat and power meter data. The building block of the energy score is the ECB that accrues over a 5-min interval in the evaluation timeframe.

Let $O_{1:N}$, $T_{1:N}$, and $S_{1:N}$ be the observed operation modes, setpoints, and occupancy states, respectively, over the N timesteps. The energy score α is a function of these three tuples:

$$\alpha(O_{1:N}, T_{1:N}, S_{1:N}) = \frac{1}{N}\sum_{t=1}^{N} \mathrm{ECB}_{S_t,O_t}\left(T_{O_t,t}\right), \tag{1}$$

where the summand, $\mathrm{ECB}_{s,o}(T_{o,t})$, is the ECB at timestamp t for the state $s \in \mathscr{S}$ and the mode $o \in \mathscr{O}$. The latter has the following generic form:

$$\mathrm{ECB}_{s,o}(T_o) = \begin{cases} 100\ \mathrm{sigmoid}\left(\alpha_{s,o} + \beta_{s,o} T_o\right), \text{if } o = \text{"heat" or "cool"} \\ 100, \text{otherwise} \end{cases}, \tag{2}$$

where $\mathrm{ECB}_{s,o}(T_o)$ is a mode-and-state-specific ECB for the mode o, setpoint temperature $T_o \in \mathscr{T}$ and the state s. When o is "heat" or "cool," a sigmoid function with the shape parameters $\alpha_{s,o}$ and $\beta_{s,o}$ evaluates the observed thermostat setpoint for the given o (i.e., T_o). The shape parameters are obtained by solving equations (3) and (4) analytically for $\alpha_{s,o}$ and $\beta_{s,o}$:

$$\left(1 + e^{\alpha_{s,o} + \beta_{s,o} T_{s,o,H}}\right)^{-1} = 95, \tag{3}$$

$$\left(1 + e^{\alpha_{s,o} + \beta_{s,o} T_{s,o,L}}\right)^{-1} = 5, \tag{4}$$

where $T_{s,o,H}$ and $T_{s,o,L}$ are the upper and lower setpoint thresholds that ensure 95 and 5 ECB scores for specific o and s. The threshold setpoints are selected based on ASHRAE Standard 55 [55] and programmable thermostat user guides from Ref. [56] to guide occupants to save energy without sacrificing their thermal comfort. The threshold temperatures and $\mathrm{ECB}_{s,o}$ curves for the heating and cooling operation modes are presented in Table 1 and Fig. 3. When the system is off, a full score of 1 is assigned.

After calculating $\mathrm{ECB}_{s,o}$ over the N time steps, we can decompose the energy score (Eq. (1)) according to the operating mode and occupancy state. Such decomposition helps to understand the contributions of the user's state-and-mode-specific behaviors to the current energy score. For this, Eq. (1) can be re-written using the following equation:

$$\alpha(O_{1:N}, T_{1:N}, S_{1:N}) = \frac{1}{N}\sum_{s\in\mathscr{S}, o\in\mathscr{O}} \left(\sum_{t=1}^{N} I_s(S_t) I_o(O_t)\right) \mathrm{ECB}_{s,o}(O_{1:N}, T_{1:N}, S_{1:N}), \tag{5}$$

where I_s and I_o are the indicator functions that satisfy the following conditions for the state s and the mode o: $I_s(x) = \begin{cases} 1, \text{if } x = s \\ 0, \text{if } x \neq s \end{cases}$ and $I_o(x) = \begin{cases} 1, \text{if } x = o \\ 0, \text{if } x \neq o \end{cases}$. To get the same result as in Eq. (1), we must define

Table 1
Threshold setpoints of state-specific ECB curves (unit: °C).

Mode	Cooling		Heating	
State	Lower threshold	Upper threshold	Lower threshold	Upper threshold
Home	20 (68°)	23.3 (74°)	27.8 (82°)	23.3 (74°)
Sleep	22.2 (72°)	25.6 (78°)	25.6 (78°)	21.1 (70°)
Away	24.4 (76°)	27.8 (82°)	23.3 (74°)	18.9 (66°)

the following equation:

$$\alpha_{s,o}(O_{1:N}, T_{1:N}, S_{1:N}) = \frac{1}{\sum_{t=1}^{N} I_s(S_t) I_o(O_t)} \sum_{t=1}^{N} I_s(S_t) I_o(O_t) \mathrm{ECB}_{s,o}\left(T_{O,t}\right). \tag{6}$$

Then, using Eq. (5), we update the energy score each day using data collected from the current week (i.e., from the previous Sunday to the end of the previous day). In this way, we can inform users about their daily progress during the calendar week. Eventually, the users get their final weekly score at the end of the week (i.e., Sunday).

4.2. Personal action recommendation

In this section, we present a mathematical framework to generate a personalized message that informs users about the best action that could have resulted in the highest increase in their energy score. The framework consists of the following steps: 1) defining the concept of messages; 2) estimating the effect of messages, and 3) framing action recommendations.

Definition 1. A message is a function from the space of operation mode, setpoints, and occupancy states to the space of setpoints. Mathematically, a generic message is a map:

$$m : \mathscr{O} \times \mathscr{T} \times \mathscr{S} \rightarrow \mathscr{T}. \tag{7}$$

Messages defined as above are too complicated to communicate. Thus, it is useful to define a message concept that only suggests changes in a setpoint related to only one occupancy state and operation mode. We will call such a message a *simple message*.

Definition 2. A simple message m pertaining to occupancy state $s \in \mathscr{S}$ and operation mode $o \in (\mathscr{O} - \{\text{"off"}\})$, is a message that suggests changing the current setpoint $\tau \in \mathscr{T}$ to the recommended setpoint c for occupancy state s and operation mode o, and does not suggest a setpoint change for the states different than s and modes different than o. The setpoint c refers to the upper threshold setpoint used in mode-and-state-specific ECB ($T_{s,o,H}$in Eq. (2)). Mathematically, m is:

$$m(o, s, \tau) = c, \tag{8}$$

while restricted on the set $\mathscr{T} \times (\mathscr{S} - \{s\}) \times (\mathscr{O} - \{o\})$ it becomes the identity map, i.e.,

$$m(s', o', \tau) = \tau, \tag{9}$$

for all $s' \neq s$ and $o' \neq o$. We will refer to such message by $m_{s,o}$.

After listing the type of simple messages, we estimate the effect of each message to find the one that would have had the highest positive effect on the user's energy score if implemented. For this, we define counterfactual scenarios in which the user selects the ideal setpoint c for the specific mode and state in the message. Then we calculate the counterfactual energy scores that users could have achieved by applying the message in the recent past (i.e., the previous day).

Let $O_{1:n}$, $T_{1:n}$, $S_{1:n}$ be the observed operation modes, setpoints, and occupancy states, respectively, over the previous n timesteps of the previous day and m be a message. We can define the application of the message on the time series as follows:

$$m(O_{1:n}, T_{1:n}, S_{1:n}) = \{(O_i, m(O_i, T_i, S_i), S_i)\}_{i=1}^{n}, \tag{10}$$

which represents the time series that results by applying the message on every element of the original time-series data. Then, we can define the relative change that could have resulted from the implementation of the messages by:

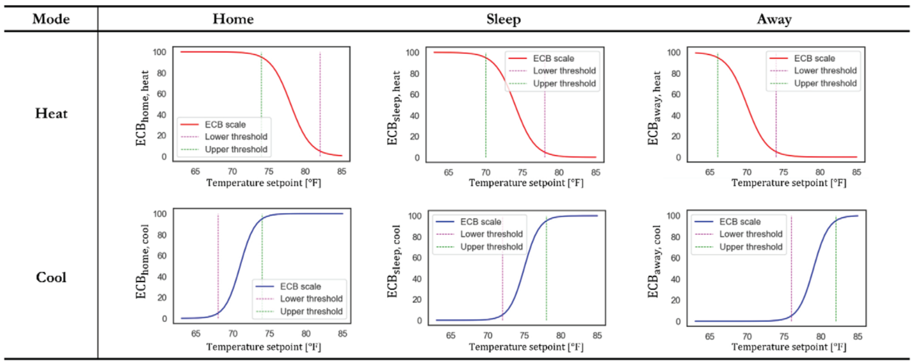


Fig. 3. Mode-and-state-specific ECB curves.

$$\rho(O_{1:n}, T_{1:n}, S_{1:n}; m) = \frac{\mathrm{ECB}(m(O_{1:n}, T_{1:n}, S_{1:n})) - \mathrm{ECB}(O_{1:n}, T_{1:n}, S_{1:n})}{\mathrm{ECB}(O_{1:n}, T_{1:n}, S_{1:n})}, \tag{11}$$

where ρ is the relative change in energy score resulting from applying the message m to the observed time-series data. A positive ρ corresponds to an improvement while negative ρ represents a reduction in an energy score if implemented. The message with the highest ρ is then selected as the best action among the set of candidate simple messages $\mathscr{M}$. That is, the message we select is:

$$m^* = \underset{m \in \mathscr{M}}{\mathrm{argmax}}\, \rho(O_{1:N}, T_{1:N}, S_{1:N}; m). \tag{12}$$

Once the best m^* is selected, we frame it into an action recommendation. For this, we define two types of framing operators i.e., literal operator and relative operator, to mathematically map the simple generic message $m_{s,o}$ to an action recommendation. The literal operator translates the message into a direct recommendation by suggesting users adjust the setpoint to the specific temperature c in Eq. (8). On the other hand, the relative operator phrases an indirect recommendation without mentioning the specific temperature. To help users adjust their setpoint without sacrificing comfort, we use the relative operator for framing messages related to the occupied states (i.e., $s \in${"home", "sleep"}).

Remark 1. In what follows we use the notation "<test>?<option1>:<option2>" as a shortcut for "if <test> is True, then <option1>, otherwise <option2>". Also the binary operator "+" concatenates strings.

The literal operator $\mathscr{L}$ takes a simple generic message $m_{s,o,c}$ for s = "away" and translates it to text as follows:

$$\mathscr{L}(m_{s,o}) = \text{"Why don't you move your"} + (o = \text{"heat"})?\text{"heating"} : \text{"cooling"} + \text{"setpoint closer to"} + c + \text{"°F when you are away?"}. \tag{13}$$

The relative operator $\mathscr{R}$ takes $m_{s,o,c}$ for $s \in${"home", "sleep"}, and it does not report c:

$$\mathscr{R}(m_{s,o}) = \text{"If you feel comfortable, why don't you"} + (o = \text{"heat"})?\text{"lower"} : \text{"increase"} + \text{"your"} + (o = \text{"heat"})?\text{"heating"} : \text{"cooling"} + \text{"setpoint"} + \text{"when you"} + (s = \text{"home"}) : \text{"are home"} : \text{"sleep?"}. \tag{14}$$

Then the intensity of the action recommendation is adjusted based on the level of ρ. For the message with ρ smaller than 5%, we provide an encouraging message such as "Keep going! You have a high score!" to complement the good performance and encourage users to keep up their energy conservation behavior. On the other hand, when ρ is higher than 30% for the away-related message, we frame the message with higher intensity to encourage residents to improve their energy conservation behavior (Table 2).

4.3. Collaborative social game

In this section, we present a collaborative social game that is designed to promote community-level energy savings. For the game, we adopt the Public Good Games (PGG) [57] approach that promotes cooperative behaviors of a group of users with a shared goal and reward. This collaborative structure has been adopted in previous social energy game studies to encourage participants to reach common energy savings goals [33,41,58,59]. These games accompany social norms and social proofs for nudging users toward collective behavioral changes to gain team-based rewards [33,46]. Thus, the game leverages financial motivations with social proof motivations for energy conservation, drawing on the demonstrated power of social proof feedback in energy conservation [60]. Our game incentivizes subsidized housing residents for achieving a shared goal that can contribute to reducing community energy usage. The building manager who is responsible for the community utility bills can benefit from this game strategy by having control

Table 2
Message framing with different intensity.

Simple generic message	ρ [%]	Message framing
All	ρ <5	Keep going! You have a high score!
$m_{\mathrm{home,heat}}$	$\rho \geq 5$	If you feel comfortable, why don't you lower your heating setpoint when you are at home?
$m_{\mathrm{home,cool}}$		If you feel comfortable, why don't you increase your cooling setpoint when you are at home?
$m_{\mathrm{sleep,heat}}$		If you feel comfortable, why don't you lower your heating setpoint when you sleep?
$m_{\mathrm{sleep,cool}}$		If you feel comfortable, why don't you increase your cooling setpoint when you sleep?
$m_{\mathrm{away,heat}}$	$5 \leq \rho < 30$	Why don't you move your heating setpoint closer to 66° when you are away?
	$\rho \geq 30$	Why don't you lower your heating setpoint to 66° when you are away?
$m_{\mathrm{away,cool}}$	$5 \leq \rho < 30$	Why don't you move your cooling setpoint closer to 82° when you are away?
	$\rho \geq 30$	Why don't you increase your cooling setpoint to 82° when you are away?

7.2. Post-experiment interviews

In this section, results from the post-experiment interviews are presented to address RQ3 (i.e., *How was users' experience with the developed platform?*). The interview results are organized into the following main themes: personal eco-feedback, social game, and MySmartE technology usability associated with the visual and audio interfaces. Key takeaways from the interviews are highlighted using participant (P) quotes.

7.2.1. Personal eco-feedback (My energy score)

Although the energy score is an abstracted measure of good energy behaviors from worst (0) to best (100), 10 out of 13 participants said they have actively tried to increase their score with better energy behaviors. Participants said they can increase their score by applying the recommended tips (P3, P4, P5, P9, P11), adjusting the temperature setpoint in their sleep or away schedule (P1, P5, P6, P8), or by turning the thermostat off and opening the windows to cool the apartment in the summer (P3, P4, P9, P10, P12). For example, P5 adjusted the schedule to increase the energy score (e.g., "*Oh, I adjusted my sleep hours and increased my setpoint to try to up my score because I never had low scores like 40s and 50s and it didn't feel good*") while P9 liked to turn off the thermostat to boost the score (e.g., "*I have increased my score by not using air or heat or anything, then it went up.*"). As such, residents learned the tactics to increase scores from their own experience.

From the interview, we learned that personalized tips, automated reminders, and an easy user interface are key to successful user nudging. Importantly, P10 explained the nudging role eco-feedback plays in their energy behaviors: "*Yeah, I actually think that the energy score makes me do better, in terms of where I have my setpoint at […] if I do see it going down, it does kind of give me the initiative to try to set it at a higher setting, […] gives you a push to do better.*" Regular popups that show up on the MySmartE screen are crucial to helping residents remember to check and apply their tips (e.g., "*The popups help remind to rise score and apply tip.*" (P3)).

7.2.2. Social game

Out of the 12 interviewees who were asked about the social games during the interview, four reported to be active participants in the games (e.g., P7, P10, P8, P11), another four were occasional participants (e.g., P13, P9, P3, P5), and the rest were passive participants (e.g., P14, P6, P4, P1). According to the active/occasional participants, the social game elements such as avatars or social proof information excited them during the game. In the game, we allocated each resident a random animal or job type icon/avatar to give an anonymous but identifiable visual image and presented it in the game popups to show the efforts of the peers (e.g., how many households increased their score from the previous week). Overall, such social presence avatars were seen as "fun" (P7), "exciting" (P7), but also "helpful" (P8, P10) and "personal" (P4). The avatars seem to have encouraged participation at both the individual and the social cohesion levels. At the individual level, it triggered social comparative processes (P7, P13, P10): "*It was interesting to see how everyone else was doing and how they were participating. And that made me want to participate.*" (P10). At the collective level, the avatars seem to have increased the community feeling to some users: "*It encouraged me, really encouraged me, oh ok, we are doing this, and we have this to achieve, it was a really fun thing for me,*" (P7). It even encouraged communication with their neighbors. For instance, P5 engaged with their peers to share how they increased the score during the games.

The participants also addressed the obstacles in the game. For instance, the passive participants said they lacked the time or stable network connectivity to engage intently. Others also mentioned that extreme outdoor weather conditions or small rewards relative to the effort prevented participation. For example, P13 explained how they tried to increase the score but couldn't give up their comfort (e.g., "*Sometimes I reached a point where I was like, this is as high as I can take it*") while P7 mentioned difficulty increasing the score (e.g., "*[…] earning the points should be doubled, i.e., a little easier to get to the scores.*"). Lastly, interviews confirmed that clear and timely communication is required to build trust and ensure positive user experiences with eco-feedback that offers tangible rewards such as gift cards. Timely game and reward payment status, procedures, and updates can help residents keep track of the game, their participation, and their benefits.

7.2.3. Visual interface of MySmartE app

When asked about the visual interface usability, 6 of the 13 participants said the app is a lot easier to use than their previous Ecobee thermostat (P4, P3, P5, P7, P10, P13). However, some technical challenges significantly affected user interactions. For instance, occasional network connectivity issues hindered user interactions. Some reported slow synchronization between tablet and Ecobee that led to a touch sensitivity issue. 10 participants admitted to using Ecobee as the backup for MySmartE when it is conveniently on the way (P7, P8, P4, P9) or having connectivity issues (P1, P6, P7, P4, P8, P10, P11, P13, P14). In addition, despite the one-on-one onboarding sessions we had with residents, some still had difficulty setting the system up (e.g., "*I have to go through multiple pages [to change my setpoint]*" (P14), "*I like keeping my cool at the same level and so I don't mess with it [the schedule and uses hold instead]*" (P7)). Considering that people have varying levels of technology literacy, one may need to assess the target users' technology literacy and adoption behaviors to plan a more customized onboarding experience (vs. a single one-time onboarding at the start of the field deployment).

7.2.4. Voice interface of alexa skill

From the interviews, we learned that all the participants were new to the Alexa device except one (P4) who had prior experience with it through family and friends. Even so, 7 out of the 13 participants said they use Alexa either a lot (P4, P5, P10) or at some regular interval (P3, P7, P8, P11). All these active users (7/13) admitted that the user guide we provided during the initial onboarding helped remind them how to talk to Alexa. Most of the active users said that they found learning to use Alexa easy (P4, P5) or slightly difficult, but nothing that cannot be learned with just a little effort (P3, P7, P8, P11, P10). These users mainly used Alexa to adjust the thermostat setpoints, change the operation mode (heat, cool, fan, off), or ask about the current room temperature (e. g., "*I use voice Alexa to change the temperature in the house. I use it a lot. Very convenient*" (P5)). Other use-cases include asking about the current energy score or the community game status. The active Alexa users have come to appreciate, enjoy, and prefer the hands-free voice control interface over other alternatives, such as the original Ecobee thermostat or even the MySmartE tablet app). Many find it "*convenient*" (P5, P8, P11) and "*handy*" (P10) because they can control the thermostat while sitting on the sofa with their kids (P4, P7, P8, P11), lying in the bed (P8, P10), or at other times with their hands are full, like washing the dishes (P10). However, they also noted that they relied on the MySmartE app to confirm whether their command was successfully implemented or not (P5, P11).

On the other hand, 6 non-Alexa users reported difficulty interacting with a pre-programmed voice controllable interface (e.g., "*difficult*" (P1), "*it didn't work*" (P6), "*it just doesn't/cannot understand me*" (P1, P9). Despite their lack of interest, nearly all six of the non-users admitted that their kids use Alexa for alternative purposes, mainly to listen to music, radio, or ask variable questions (P1, P2, P6, P9, P13).

8. Conclusions

In this paper, we presented for the first-time a cloud-based eco-feedback and gaming platform (MySmartE app) aiming to promote energy-conserving thermostat-adjustment behaviors in multi-unit residential buildings. The platform was deployed in a residential community and the main findings from the field study are summarized as follows:

Second, the result of user interaction analysis is presented to answer RQ2 (i.e., *How do users interact with MySmartE app?*). Fig. 15 shows how user interactions with our smart devices changed during the game intervention period. The heatmap shows the percentage of households that interacted with the various features in the MySmartE tablet and Alexa apps. From this, one can notice that the game started with high user interactions for the first week but decreased to a certain extent in the second week. This has been reported in previous studies that addressed the decay of behaviors after an intervention [67–69]. After the 2nd week, the game was interrupted by a lightning strike and resumed from the week of July 5th. Due to the effect of the interruption, the interaction was somewhat low during the week of July 5th, but it increased again from the next week as users started to receive social feedback. Throughout the game, the daily popup notifications played a key role in maintaining a certain rate of interactions by bringing residents' attention to various information about the game. For instance, the users who first interacted with the game popups tended to navigate to the 'My energy score' page or 'Community game' page to check their current energy score and the status of the ongoing game (Fig. 16). Among them, about one-third of users who visited the score page applied the daily action tips to their thermostat schedule using the control panel on the page. For the users who applied tips through the popups, about 50% of them visited the thermostat panel or the schedule panel afterward to check their updated thermostat information. Although user interactions with the 'Apply tip' feature were relatively low during the game, the results in Fig. 14 show significant differences in the room air temperature between the intervention and baseline period. The level of change varies among residents as they followed different approaches to improve their behavior based on their preference and pace, i.e. by applying the tip directly or indirectly by manually adjusting their thermostat panel or schedule to follow tips. The 'apply tip' button in the interface helped users to sustain the improved behavior by maintaining the tip on users' thermostat schedule until they manually override it with other setpoints.

Fig. 17 shows Alexa user interactions during the game experiment.

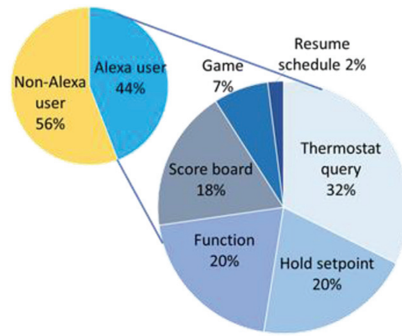


Fig. 17. Breakdown of Alexa user interactions.

During the game, about 44% of users interacted with the voice interface in total. As one can notice from Fig. 15, the rate of weekly Alexa usage was smaller than that of the other features provided in the visual interface. From this, we can infer that many users were somewhat unfamiliar with voice assistant devices throughout the game experiment. The users who interacted with Alexa mainly used it for controlling thermostats by querying their room temperatures, temporarily holding the thermostat setpoint, or turning on and off the system (Fig. 17). A small percentage of users queried eco-feedback information from Alexa along with the game status information. Considering the characteristics of each voice command, we could infer that users tended to use simple and intuitive thermostat-related commands that are more related to their everyday life than the feedback commands.

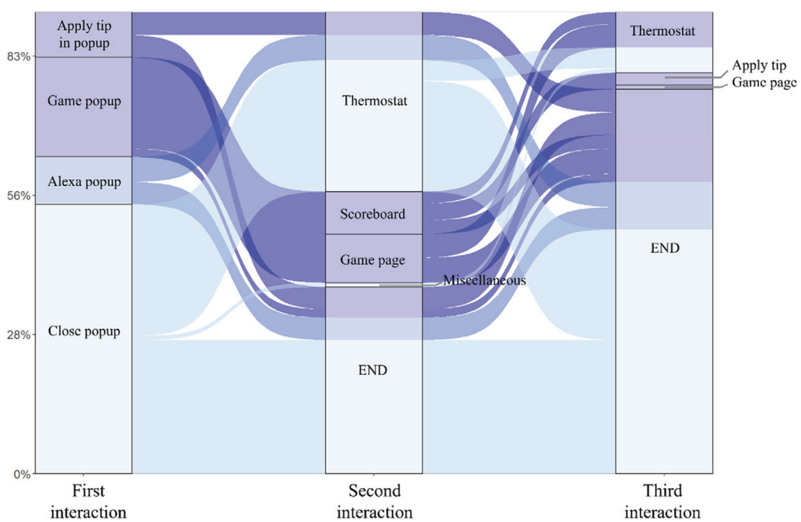


Fig. 16. User interaction flow with popup.

over the goal and the incentives that affect user participation in reducing each households' HC energy usage. The three major game elements such as the community goal (α_{goal}), personal participation threshold ($\alpha_{threshold}$), and incentives (I) are generated every week. Each game constitutes four rounds of weekly games where the new game elements are calculated every Monday with the start of the week. The results of the weekly games are announced every Sunday. The detailed game rules and logic are described below.

4.3.1. Community goals

The first week's community goal ($\alpha_{goal,1}$) is calculated by rounding up the previous week's community average energy score ($\alpha_{com,0}$) to its nearest multiple of 5 (i.e., $\alpha_{goal,1} = 5\left\lceil\frac{\alpha_{com,0}}{5}\right\rceil$, where $\lceil x\rceil = \min\{k\in\mathbb{Z} \mid k \le x\}$). From the second week, the goal increases every week by 5 points only if the community has achieved the goal in the previous round. Otherwise, the goal stays the same as the previous week. This is shown in Eq. (15) as follows:

$$\alpha_{goal,w} = \begin{cases} \alpha_{goal,w-1} + 5, \text{if } \alpha_{com,w-1} \ge \alpha_{goal,w-1} \\ \alpha_{goal,w-1}, \text{otherwise} \end{cases}, \tag{15}$$

where $\alpha_{goal,w}$ is a community goal of the week $w \ge 2$, $\alpha_{goal,w-1}$ and $\alpha_{com,w-1}$ are the community goal and the average score of the previous week, respectively. To achieve the community goal, households are required to increase the community's average energy score above or equal to the goal by the end of the game week (i.e., Sunday).

4.3.2. Participation threshold

To be qualified for the weekly reward, residents are required to keep their scores above or equal to the participation threshold at the end of the game week. This threshold is designed to prevent incentivizing free-riders who may earn rewards without any effort. The thresholds are always 5 points below the weekly community goal (i.e. $\alpha_{threshold} = \alpha_{goal} - 5$).

4.3.3. Personal reward

The personal reward starts from \$5 (i.e., I_0) in the first week and increases by \$5 every week only if the community achieved the goal in the previous week. The new game round's reward will stay the same as the previous week if the community fails to meet the goal. We set the maximum limit of the weekly reward to \$15 to prevent the reward from growing indefinitely. For $w \ge 2$, we can define the reward of the week w as below:

$$I_w = \begin{cases} I_{w-1} + 5, \text{if } \alpha_{com,w-1} \ge \alpha_{goal,w-1} \\ I_{w-1}, \text{otherwise} \end{cases}. \tag{16}$$

The weekly rewards are stacked over 4 weeks and paid out at the end of the game set.

5. MySmartE development

In this section, we present an overview of the software platform and the development process of the user interfaces.

5.1. Software infrastructure

The architecture of MySmartE is designed to support multi-modal communication among multiple software elements (Fig. 4). The three main software elements that structure the system are 1) visual and voice user interfaces, 2) API server, and 3) MySQL database.

First, the two types of user interfaces i.e., visual and voice, were developed to provide users with manual and hands-free options to interact with the eco-feedback and smart thermostat. The visual user

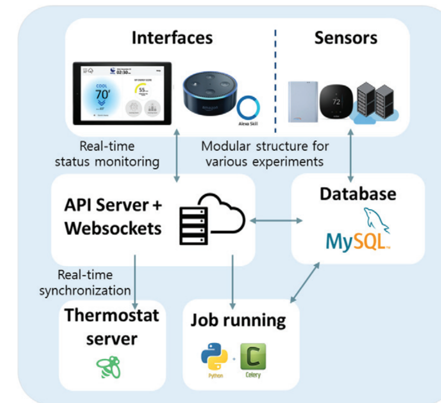


Fig. 4. Software infrastructure.

interface (UI) was designed for the MySmartE web application, which users can access via wall-mounted tablet displays. The front-end of the web application was developed using React[4] in Javascript to selectively add or remove the modularized feedback elements based on the type of experiment. It also incorporates the thermostat control panel that is paired with the Ecobee smart thermostat to control the home HC system. The voice user interface (VUI), accessible from the Amazon Echo Dot, can serve as a functional alternative to the tablet for controlling the thermostat and querying the feedback information.

Second, the API server was built on Google Cloud to enable real-time communications among the web and voice applications, connected devices, and a database with a common API. This unified API contributes to maintaining consistency across the web app and Alexa. The API was built with Falcon[5] in Python,[6] and exposes various REST endpoints for controlling the thermostat, interacting with feedback information, and applying changes. All calls to the API are logged to the unique table in the database, which helps understand the usage patterns across the project site. This also allows distinguishing between touch-based interaction and voice interactions.

The API also enables our web app to replicate Ecobee features in real-time by supporting data synchronization via polling every 5 s and on-demand WebSockets connections. The API and web socket server gets a dedicated host to increase the back-end performance by separating it from a compute-intensive data processing host. A message broker is used to help issue any on-demand tasks on the data processing host from the API host.

Third, the MySQL database was set up for managing multiple data sources such as user interactions, sensor readings, and computation results from the feedback algorithms. The user activities on the UI/VUI are captured in real-time. The data collected from the power meter and Ecobee thermostat are pushed to the database at a set frequency. The raw data is stored as-is for archival purposes, while processed data is incrementally generated every hour. This processed data goes into a feedback algorithm for generating users' daily eco-feedback.

The database structure was designed to enable adding or removing users and apartment units based on the move-in and out status of the

[4] https://reactjs.org/.
[5] https://pypi.org/project/falcon/.
[6] https://www.python.org/.

residents. Similarly, the apartment units can be aggregated into teams and teams can be aggregated into a single group. Each group is assigned to a particular experiment. This hierarchy helps to conduct different experiments and the tablet application adapts automatically without any manual intervention.

5.2. Visual user interface

The UI of the MySmartE app follows a flat information architecture that prioritizes effective thermostat control and feedback visualization using comprehensive visual layouts. Our design process involved the following steps: 1) drawing wireframe and mockups, 2) developing prototype, and 3) iteratively reviewing and modifying the prototype. First, the overall layout of the UI structure was drafted into wireframe and then developed into mockups with refined visual elements using Adobe XD[7] program. Then prototypes were developed and tested to simulate potential user interactions. During the process of 1) and 2), the designs were reviewed and updated by an internal testing group composed of graduate students, professors, and industry partners involved in the project. From this process, the web application was structured into six pages based on their major roles and the corresponding functions required to enhance the user experience.

The 'Welcome page' is designed to onboard first-time users with a virtual walk-through. This page pops up when the new user registers into the web application and allows users to go through a step-by-step thermostat schedule setup as well as introduction slides about the overall features of the application.

The 'Main page' is designed to provide an overview of users' current status (Fig. 5). From the header of the 'Main page', users can view current weather information, a current thermostat schedule icon, time, and online user guides from the help button. By clicking the schedule icon on the header, users can view their weekly thermostat schedule and make any modifications. The page also allows users to navigate the other pages such as 'Thermostat panel', 'My energy score', and 'Community game'. The 'Thermostat panel', which is located on the left side of 'Main page', allows users to manually set up their thermostat by selecting a desired system function, setpoint temperature, and the duration to hold the setpoint. It also provides an 'Eco-Away' button to help users efficiently setback or up the setpoint before they leave the residence.

The 'My energy score' page provides personalized eco-feedback such as energy score and personal action tip (Fig. 6). The score is presented every day along with the past 7 days' data to enable self-comparison. The personal tip is presented with an action button (i.e., 'Apply Tip to Schedule') that enables direct implementation of the tip on the user's thermostat schedule. When users click the button, a small thermostat control panel pops up so that users can adjust the tip-related setpoint temperature programmed in their thermostat schedule. To nudge the use of energy-conserving setpoints, the arrows in the control panel are only activated in energy-saving directions (e.g., downward arrow for heating, upward arrow for cooling). Once the tip is applied, users' new setpoint updates in the thermostat schedule and a short message shows up on the 'Thermostat panel' to confirm the tip activation.

The 'Community game' page provides an overview of the daily game status (Fig. 7). On the page, the community's daily average energy score is visualized with a weekly community goal in a bar chart. The weekly reward is provided right below the community bar. The user's participation status is also visualized on the left side of the page with a custom avatar and a gauge bar comparing one's current energy score and the participation threshold. From the page, users can also see the distribution of their neighbors' energy scores by clicking the 'community participation' button. The navigation button to 'My energy score' is available on the page to nudge users to take actions to improve their performance in the game (i.e., 'See My Daily Tip').

Lastly, the 'Screen saver' is designed to visualize the summary of the current thermostat status and to periodically provide visual cues to nudge users' behavior changes. For this, the animating popups with personal tips and encouraging messages are provided every day at 9 a.m. During the game period, various types of popups that inform game status, social proof information (i.e., how the neighbors are doing), and Alexa commands are also provided.

5.3. Voice user interface

A VUI is designed to allow hands-free Ecobee control through voice interactions. Recent developments in systems tackling voice interactions have enabled dialogues ranging from open-ended to task-oriented commands [61], as well as handling of ambiguities and fuzziness of natural language [62]. To ingest, process, and respond to a user request, tasks such as speech recognition, natural language understanding, dialog management, and natural language generation are performed [63]. In this work, we developed a task-oriented VUI through a custom-developed application or 'skill' using the Alexa Skills Kit.[8] This VUI tends to specific user invocations while addressing explicitly stated as well as ambiguous or fuzzy natural language commands. The skill enables users to interact with an Amazon Echo device, that processes and responds to user requests regarding energy control and monitoring.

5.3.1. Skill development

The skill development process is divided into two parts, i.e., skill design and voice interaction model design. The skill design process requires defining and mapping user intents and utterances. User utterances describe particular goals as enunciated by the speaker. For example, when a speaker provides an utterance, *'What is the temperature in the room?'*, the associated intent class can be formalized as *'Get Temperature'*. The voice interaction model is designed to customize the responses of our VUI to specific user queries. The voice interaction model is hosted on the cloud using an on-demand computing service (Amazon Lambda, AWS) [64].

5.3.2. Skill workflow

Amazon's Alexa is a VUI that allows developers to create a conversational mechanism for interacting with an application. In this project, we use Alexa to give users access to both their application and the Ecobee through the Alexa smart speaker or any mobile device with the Alexa app. Through Alexa, we allow users to control their Ecobee thermostats and review their energy savings.

Through Alexa, users can query the Ecobee for device status, current temperature, temperature settings, the Ecobee's schedule, and outdoor temperature. The user can also control the functionality of the device by turning the device on and off, changing the mode of the device, and adjusting the temperature. Apart from these Ecobee controls, the interface also incorporates game controls. These include querying game rules, user status in the game, game-related tips, etc. The dialogues designed based on all possible voice interaction scenarios were reviewed by the internal testing group during demonstration sessions.

Temperature controls may be adjusted through explicit or implicit commands. The skill recognizes a range of phrases used in natural communication to control the temperature in the unit. For instance, a user can ask Alexa to change the temperature from 76 to 78°F. The user may issue the same adjustment by telling Alexa to increase the temperature by two degrees. Alternatively, the user may request a change using a fuzzy quantifier, such as "a little" without explicitly identifying the number of degrees for the adjustment, which will activate a fuzzy module for the temperature calculation [65]. The user is also able to indicate a level of discomfort on a scale from *freezing* to *burning up*. Synonyms for each level on the scale are included to allow for easier and

[7] https://www.adobe.com/products/xd.html.

[8] https://developer.amazon.com/en-US/alexa.

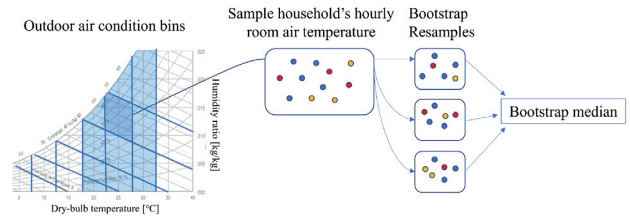


Fig. 13. Data bootstrapping process.

were synthesized into the final results that are presented in Section 7.2.

7. Results

7.1. Field experiment data

First, we present the results of the baseline and intervention data analysis to answer RQ1 (i.e., *How do occupant behaviors change after using the MySmartE app? Does it improve households' thermostat settings and setpoint selections compared to the baseline period?*). Fig. 14 illustrates the impact of the eco-feedback experiment during the cooling season by showing the changes in the households' room air temperatures during the intervention period in comparison with the baseline. The data shows the households' room air temperatures when they were occupied (i.e., home and sleep mode). The baseline data shown in the blue box was collected before the installation of any smart devices while the red-colored intervention data includes all data collected from the feedback experiment. To compare the two datasets, we computed bootstrap medians from the hourly room air temperature data that experienced similar outdoor air conditions. In Fig. 14, each column on the x-axis represents individual households and the y-axis shows the hourly room air temperatures. At a glance, one can notice that most households' median room air temperatures were increased after the intervention. Considering that this is a weather normalized comparison, we can infer that the increases in room air temperatures were caused mainly by users' behavior changes, i.e. the improvement in users' thermostat settings that increased their usual cooling setpoints.

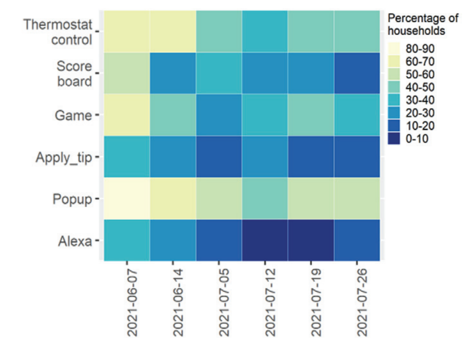


Fig. 15. User interaction heatmap.

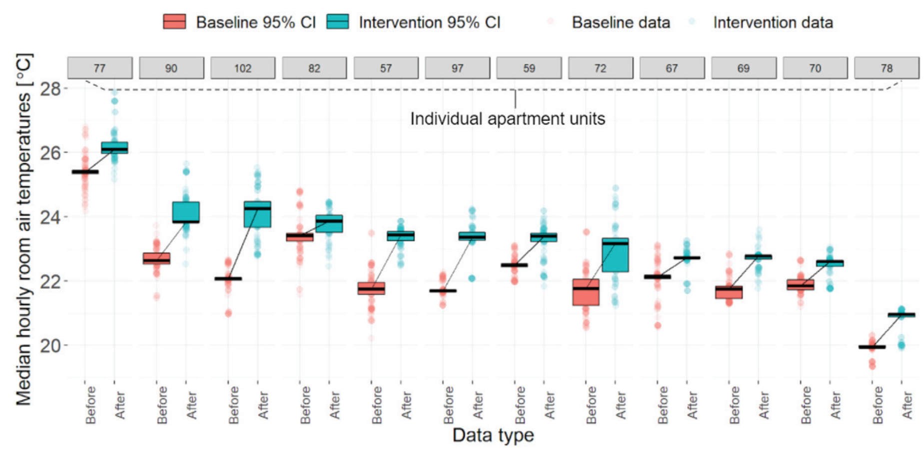


Fig. 14. Cooling season room air temperature during the baseline and intervention period.

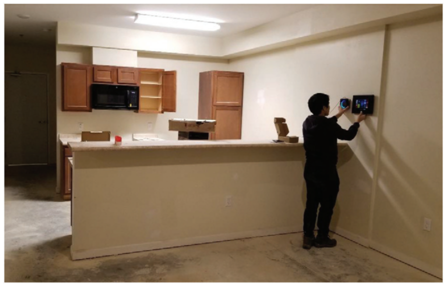


Fig. 11. Installation of the visual and voice eco-feedback devices.

Lastly, we conducted post-experiment interviews to get residents' feedback about the user experience and the app's usability (RQ3). The interviews were conducted by 13 volunteers after the end of the game experiments.

6.2.1. Baseline and intervention data comparison

The effect of eco-feedback was investigated by comparing the households' cooling season room air temperatures during the baseline and intervention period. First, we categorize the data corresponding to its data collection period. The baseline data collected between December 1st, 2019 and December 31st, 2020 are labeled as 'baseline' to indicate data collected before the installation of eco-feedback devices (i.e., a wall-mounted tablet and Alexa). The intervention data measured between January 1st and August 1st, 2021 is labeled 'intervention' to represent user data affected by the presence of eco-feedback devices.

Then, to filter the cooling season data for the analysis, we define the seasons based on the prevailing HC operational mode for different outdoor air conditions. As shown in Fig. 12, we calculated the households' daily averaged HC operation hours with respect to the daily averaged outdoor air temperatures using the whole data collected. In the figure, the blue represents the operation hours of heat pump cooling while the red indicates the operations of the auxiliary heater in the air handling unit or heat pump heating. The threshold temperature for defining the cooling season was set to 18 °C as the households' HC systems were running in cooling mode dominantly when the daily

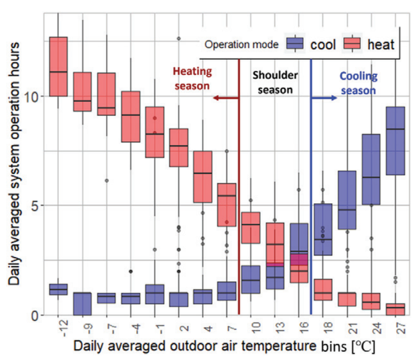


Fig. 12. Daily HC system operation hours for different modes and outdoor air temperatures.

averaged outdoor air temperature was above this threshold. From this, the days with daily average outdoor air temperature higher or equal to 18 °C were filtered as the cooling season for each data set.

After filtering seasonal data, households' room air temperatures during the occupied hours collected from the baseline and intervention data are compared under similar outdoor air conditions. For each data set, the hourly room air temperatures are grouped by the outdoor air condition bin divided based on the range of the dry-bulb temperature and wet-bulb temperature. Then we perform bootstrapping for each bin to compare the properties of the two different datasets under the same sample size (Fig. 13). The bootstrapping allows inference of population properties by estimating the summary statistics of the resampled data. For each bin, we resample the hourly indoor temperature data 50 times for each household and calculate bootstrap medians. For bootstrapping, the sample size is set as the original sample size in the bin.

6.2.2. User interaction analysis

User interactions with the features of the MySmartE app are analyzed using real-time interaction logs. First, we classify the data into 6 categories based on the characteristics of the interaction features: 1) thermostat control; 2) scoreboard; 3) game; 4) apply tip; 5) popup; 6) Alexa. The 'thermostat control' includes all interactions made on the thermostat panel and the schedule setup panel. The 'scoreboard' includes user interactions with the buttons in 'My energy score' and the page log while the 'game' represents all interactions made with the features on the 'Community game' page as well as any actions to enter the page. The 'apply tip' includes all tip-apply actions made through the popups on 'Screen saver' and the 'My energy score' page. One thing to note is that this analysis does not include users' tip-related interactions made directly on the thermostat or the schedule panel. The 'popup' includes the user interactions with the buttons on the daily popups that allowed users to navigate to other pages in the app or apply the tip. The interactions to close the popup pages are excluded from this category. After the classification, the percentage of households that interacted with each feature is calculated using weekly interaction data.

6.2.3. Post-experiment interviews

To solicit user feedback on their experience with our app, we conducted 30-min semi-structured phone interviews with 13 residents living in the community.

To recruit interviewees, we distributed flyers to all participating residents. The flyers listed a phone number where residents had to send a text message to participate in the interview. 15 residents opted in and we texted them a brief survey to collect basic information such as their building (North/South), unit number, and availability from a predefined list of 30 min slots on specific dates. Once users responded to the survey, we sent them a confirmation message. In general, we sent out reminders three times: one week, two days, and in the morning ahead of the scheduled time. Due to scheduling difficulties, two of the 15 volunteers did not complete the interview.

The interviews were conducted over the phone, recorded with iTalk iPhone application version 4.7.5, and transcribed with the Microsoft Word transcription feature into a Word document. We had a general interview script for all interviewees asking general questions about each device (MySmartE app and Alexa), specific feature usage, and confirming how well they noticed changes in the app. Example questions include: *Please describe when and why do you use the MySmartE tablet? How have you tried to increase your Energy score?* and *How many games did you play over the summer?* We also asked specific questions to each resident based on their recent interaction data.

To analyze the interview data, one researcher listened to the audio recordings and transcribed summary notes of each participant's comments to each of the interview questions. We followed a top-down approach for the initial transcription and data organization according to the interview script. Then, two researchers conducted thematic analyses [66] on the transcribed notes and summarized major themes. These

Fig. 5. The user interface of the 'Main page'.

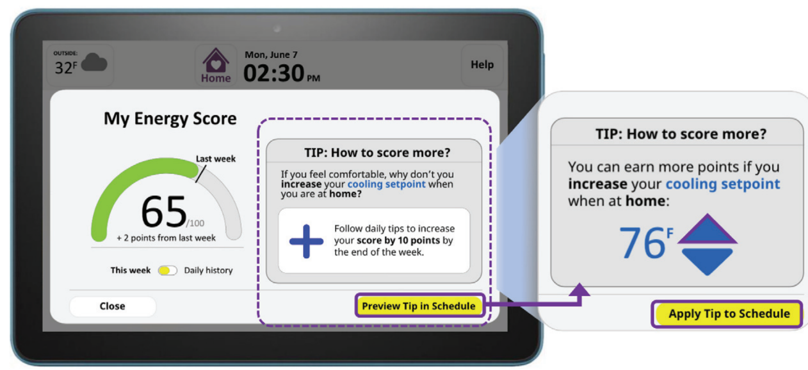


Fig. 6. Applying tip in 'My energy score' page.

more fluid communication. Based on the level of stated discomfort, Alexa will change the set temperature of the Ecobee by increasing or decreasing the temperature by up to three degrees.

In addition to controlling the temperature, the user can use Alexa to interact with the application itself. Alexa can provide users with details about their thermostat usage, recommendations around improving performance, and instructions on how to use the application. An example of interaction is shown in Fig. 8. The Alexa skill has been modified throughout this project to add various functions as well as provide more variability in conversation. It is possible to capture weekly aggregated interaction with the utilities provided in the Alexa Developer Console. Two such interactions have been aggregated using the interaction flow diagrams shown in Fig. 9 and Fig. 10. These flow diagrams show the progression in interaction complexity, pointing towards an increase in the number of dialog turns and the variety of user intents over time. As can be seen from the figures, all interactions start with evoking the skill and then proceed based on the user requests. The red line at the end of a level indicates that a user ended a conversation at that level. A path to the next level shows an intent (a request to be fulfilled) that was called based on the user request. The data in the flow path also shows that the users are not only asking for temperature changes but also communicating when they go to bed which indicates that the temperature can be lowered based on their preferred sleep temperature.

6. Field deployment of MySmartE

6.1. Resident engagement

In collaboration with the housing developer and the facility manager

Fig. 7. Community game page.

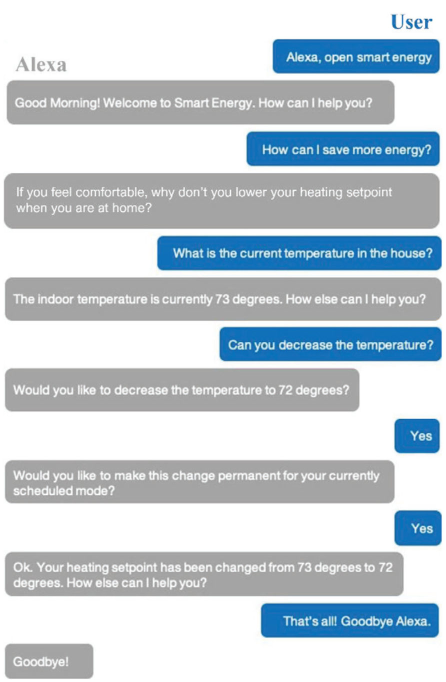


Fig. 8. An example of voice interaction.

of the community, we recruited the residents of 41 households to participate in our study. The residents who agreed to participate received a one-time sign-up payment and a quarterly participation payment. The initial payment was for installing feedback devices and consenting to the intervention while the reoccurring payment was provided in compensation for the technology adoption, any technical inconvenience that may arise during the interventions, and occasional surveys and user interviews that we conducted to collect user feedback. In each unit, we installed the tablet and Alexa on the wall located near the living room and dining area (Fig. 11). After the installation, the residents participated in an onboarding session by going over a virtual walk-through on the 'welcome page'. During this session, residents also received printed user guides about how to use MySmartE app.

Once onboarded, users were given access to all the features in the app except the game page to adapt to the basic functionalities before starting the game. The games were activated during the heating and cooling months (i.e., heating: February, March, cooling: June, July, August). During the shoulder months, users only received personal feedback without any social game information.

For the first four months of the intervention studies, we focused on collecting initial user data and updating the software to enhance the app's usability and increase users' technology adoption. After the software updates, the users were able to participate in the fully-functioning cooling season community game, which was conducted from June 7th to 20th and from July 5th to August 1st, 2021. The game halted from June 20th to July 4th due to a network problem derived from a lightning strike. At the beginning of the game, 36 out of 41 units participated in the studies as two units opted-out and three units had technical issues with heat pumps and smart devices. Due to the increase in the network issue over time, about 26 units were left at the end of the game. During the game season, participating units received weekly game flyers that advertised new updates in the game. The field study was designed and conducted in compliance with Purdue's Institutional Review Board (IRB Protocol #: 1702018811) to ensure that all human subjects research was conducted ethically.

6.2. Data analysis approach

In this section, we present the results from the intervention period of the field study. The analysis is organized around the three main research questions presented in Section 1.3 as follows: 1) intervention and

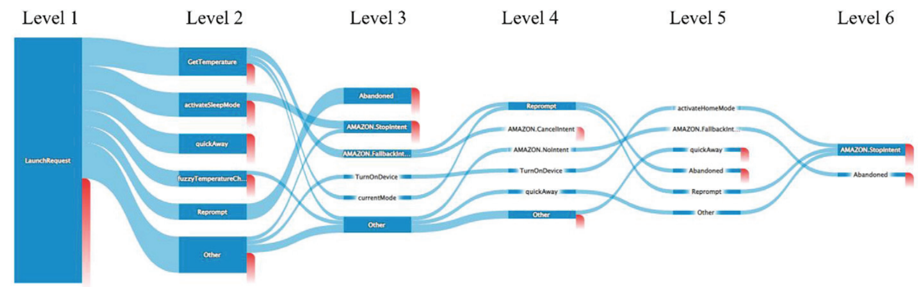


Fig. 9. Early-stage interaction path of the week from 11/09/2020 to 11/15/2020. Thicker paths indicate more interactions and narrower paths mean fewer interactions. During this week, the interactions between users and Alexa lasted at most 6 levels.

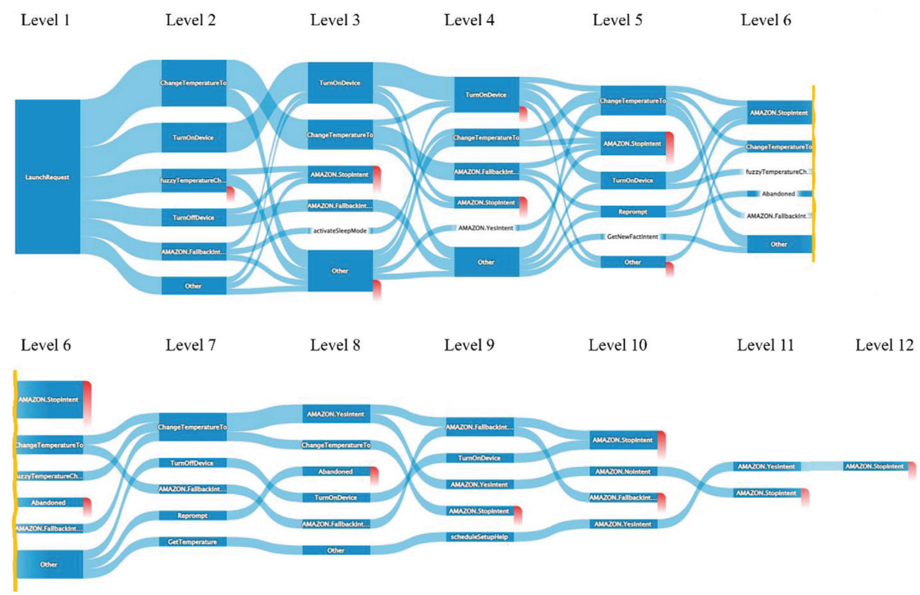


Fig. 10. Late-stage interaction path of the week from 01/31/2022 to 02/06/2022. During this week, the interactions between users and Alexa lasted at most 12 levels.

baseline data comparison (RQ1), 2) user interaction analysis (RQ2), and 3) post-experiment interviews (RQ3).

To address RQ1, we analyzed the effect of feedback intervention on user behaviors by comparing the thermostat data collected during the baseline and intervention periods. For the comparisons, we used field data from the cooling season to investigate the effect of a composite feedback treatment that includes both the personalized feedback and the community game. The heating season data were disregarded in the analysis due to the immatureness of the app during the early stage of the implementation. For the analysis, we filtered the households who moved in before June 1st, 2020 and screened out the units that lost baseline data due to network issues, resulting in a final count of 12 households. One thing to note is that energy usage analysis is not included in this paper due to the loss of power meter data caused by the lightning strike that damaged most of the power sensors on June 20th, 2021.

For RQ2, we examined user interactions with the features in the MySmartE app and Alexa to understand the key features that drive user actions. For this, we used interaction data collected from all 36 households that participated in the cooling season community game in the year 2021.