Designing a Wise Home: Leveraging Lightweight Dialogue, Proactive Coaching, Guided Experimentation and Mutual-learning to Support Mixed-initiative Homes — Comfort-aware Thermostats as a Case

by

Chuan-Che Huang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Information)
in The University of Michigan
2019

Doctoral Committee:

Associate Professor Mark W. Newman, Chair
Assistant Professor Nikola Banovic
Assistant Professor Tawanna Dillahunt
Assistant Professor Walter S. Lasecki

Chuan-Che Huang

chuanche@umich.edu

ORCID iD: 0000-0001-5802-1216

Acknowledgements

Writing this dissertation was one of the most difficult challenges that I have ever taken. Indeed, it takes a village to raise a PhD, and I would not be able to complete this study without the support of many.

If I can only pick a key takeaway that I learned from this exhaustive PhD hike, it will be "[God is] the vine; [we] are the branches. If [we] remain in [him] and [him] in us, [we] will bear much fruit; apart from [him we] can do nothing" (John 15:5). This Bible verses suits my experience particularly well, not just because of its biblical implications that sustain my faith but also because how often the academia circle loves to talk about "low-hanging fruit" in research.

I thank God for leading me to the field of Human-Computer Interaction, and helping me uncover my calling in designing digital technologies that positively improve people's lives. I thank God for opening the door of University of Michigan to both my wife Chih-Chieh Hsu and me, so that we could be together and support each other while being far away from our families. I thank God for placing many good mentors and friends along my PhD hike, helping me pursue a route that brings together my passion in design and in technology.

I would like to give thanks to my beloved wife Chih. It warmed me up every night to come to a home with a long-time friend that would listen to my day and share hers. It was a luxury to have a chef that would cook for me during my long haul. It was a luxury to have a partner that would travel with me when I needed some refreshment. People said that loneliness is the biggest enemy for PhD students, however, with Chih, I never feel lonely.

I would also like to thank my son, Renzo. You certainly served as the biggest motivator for me to finish up my study early so I could spend more time with you and with Chih after graduation. I am not sure if you will read this acknowledgements in the future when you grow up. I found my father's PhD thesis while in college, and was delighted to read his acknowledgements. He thanked my mother for drawing many of the figures used in his thesis, as his work was written at an age without PC and laptops. Perhaps when you read this thesis PC and laptops are also outdated. Being together with you and Chih in the last two years of my PhD study is the most precious time I have. I enjoy playing LEGO with you every morning and evening; I love drawing with you on the giant whiteboard walls in my office; I love taking you to the park right in front of our home and watching you grow.

I thank my long-time advisor Dr. Mark W. Newman for adopting me into his research group. Mark offered me tremendous intellectual freedom, allowing me to achieve many of my goals that are not strongly related to my dissertation: I learned digital fabrication in arts school, founded a design studio/makerspace, worked on several artistic side projects that I failed to publish, interned at a crowd-funded startup, and was heavily involved in the design of a course that took away many of my research hours. Through the freedom Mark offered, I learn how to carve my own career path and *"listen to my inner voice"*, rather than simply following other people's expectations.

In addition to Mark, I was very fortunate to have Dr. Tawanna Dillahunt, Dr. Nikola Banovic, and Dr. Walter Lasecki on my dissertation committee. Dr. Dillahunt served on many of my PhD milestone committees, providing many insights related to Sustainable HCI. Dr. Dillahunt also mentored me in lecturing in my early days of teaching. I also appreciate the fresh perspectives from Dr. Banovic and Dr. Lasecki, in particular on technical HCI and machine learning. I wish I had more chance to collaborate with them, and I am very jealous that new students at the University of Michigan nowadays can access a huge body of knowledge in human behavior modeling and in human-AI collaboration through them.

One of the greatest benefits studying in School of Information (UMSI) is to make friends with students from all over the world and from all kinds of backgrounds. To start, I would like to thank my two Christian colleagues, Rayoung Yang and Jasmine Jones for teaching me how to lead pioneering research while serving people nearby. I also like to thank my long-time collaborator and lab mate Gaurav Paruthi. Gaurav shares many similar interests as me, without Gaurav's accompany, my PhD hike would be very boring. Further, I would like to thank other members of Interaction Ecologies Group, including Tao Dong, Yung-Ju (Stanley) Chang, Shriti Raj, and Xinghui Yan. They shared many feedback on my research projects, helped me rehearse my talks, and discussed with me on many topics related and unrelated to research.

I also like to thank my friends who founded the DoIIIT Studio with me, including Yu-Jen Lin, Hariharan Subramonyam, Shiqing (Licia) He, Cindy Lin and Maulishree Pandey. Although founding and managing DoIIIT Studio took me endless hours (and perhaps reduced my short-term publication throughput), I gained so much energy from *making* things with them. This energy certainly helped sustaining my pursue of PhD.

In addition, I would like to thank many other friends in UMSI, including Priyank Chandra, Xuan Zhao, Youyang Hou, Wei Ai, Shiyan Yan, Yan Chen and Sangseok You. I enjoyed sharing life stories with you by the dinner table after work. Your mental support and encouragement helped me pass through the valleys of PhD study. I will always cherish the times we had hotpot together in the cold winter of Michigan.

Although only one name can be associated with this dissertation, I would not be able to complete this work without my collaborators and mentees, including Shen-Yuan Liang, Bing Wu, Michael Ho, Younghun Lee and Bomi Shim. They provided many support on the design, development and deployment of Project Reef. I am glad to have the chance to mentor and learn from so many talented minds.

I also like to show my appreciation to Pinghui Wu, Stuart Chuang, Paul Huang, Chiao-li Chan, Jingping and CT Lin, Jane and Brandon Girard, the Wicks, the Sanborns, the Reeds, the Cherrys and all our friends from church. It is you who help us turning Ann Arbor our home.

Finally, I would like to thank my father Wen-Kuei Huang, my mother Jui-Ying (Ema) Yang, as well as my father- and mother-in-law Chang-Lang Hsu and Yen-Wen Chen for supporting me to pursue my dream of studying abroad. We made it together!

Table of Contents

List of Figures

List of Tables

List of Appendices

Abstract

Science fiction writers have been dreaming of homes that can understand our preferences, assist our daily chores and teach us to be healthier, more sustainable and more knowledgeable. While we are still far from achieving this dream, the recent development of mobile devices, wearable interfaces, smart home appliances, and machine learning offer unprecedented opportunities for homes to better understand our goals, preferences and contexts, as well as to facilitate our everyday tasks and decision making. These recent advancement open a new possibility to create homes beyond simple automation: they enable creation of homes to coach us to achieve our better selves. That is, homes that are not just smart, but wise as well.

However, to develop a wise home, there are still two key questions: How can a wise home coach its occupants while considering their different goals and needs? How can a home integrates emerging sensors, devices and interfaces to better understand their goals, preferences and contexts in order to support coaching? To answer these two questions, in this dissertation I use residential heating and cooling control as a lens to advance the development of wiser homes.

Based on the three studies conducted, this thesis provides three contributions. First, I show that we can integrate a diverse class of emerging devices, including mobile phones, smartwatches, in-home sensors and home appliances to capture important user contexts, such as individual preferences for thermal comfort. The integration of these emerging devices enables a home to better coach its occupants and potentially better support automation. Secondly, I show that mixed-initiative interaction is an effective approach in the design of a wise home, and further propose four design strategies to support mixed-initiative homes, namely, lightweight dialogue, proactive coaching, guided

experimentation and mutual-learning. Finally, I demonstrate a novel system that integrates the above-mentioned strategies to support the development of wise homes, facilitating home occupants to identify actions to achieve a better balance between their comfort and savings goals.

Chapter 1. Introduction

"*With a bronze mirror, one can see whether he is properly attired; with history as a mirror, one can understand the rise and fall of a nation; with men as a mirror, one can see whether he is right or wrong. Now I've lost my faithful mirror with the death of Wei Zheng" - Emperor Taizong, Tang Dynasty.*[1]

1.1 The Dream of a Wise Home

Hollywood film makers and novel writers have dreamed of a future where our homes can carefully adapt to our needs: they can assist our daily chores, help us to live healthier, be more comfortable and sustainable, and facilitate learning and positive family relationships. The movie *1999 A.D. House of Tomorrow*, released in 1967, embodies such a dream. It depicts vignettes of human desires bond with our home spaces: as a person sits on a coach to read, the home quietly adjusts the lighting, heating, and cooling to fit her needs. As another resident lies on a lounge chair in a home gym, the home scans his body and suggests some exercises to do. The home also serves as a private tutor to a child, among many other capabilities.

This futuristic home, as depicted in these vignettes, is quite intelligent in many aspects. It communicates with occupants to encourage them to engage in healthy behaviors. It accurately senses the activities of occupants (e.g., reading on a couch) and controls

[1] From The Old Book of Tang, https://en.wikipedia.org/wiki/Old_Book_of_Tang

multiple appliances to achieve a goal (e.g., heating and lighting). It has the power to reason appropriate actions to take and suggestions to make; instead of passively responding to user requests, it uses its own morals to offer guidance. Integrating all these capabilities, the home helps its occupants achieve different levels of needs: being safe, healthy, sustainable while comfortable, and knowledgeable.

In some senses, this home is just like a wise minister (Wei Zheng) to an ancient Chinese emperor (Tang Emperor Taizong) — it carefully facilitates positive behaviors in the occupants while knowing that it is not the lord of the house. That is, to succeed, the home must respect the feelings and needs of its occupants. As Wei Zheng is often described as a wise sage, it may be more appropriate to call such an intelligent home a wise home, rather than just a smart home.

What, then, separates a wise home from a smart home? Continuing with Wei Zheng as a metaphor, a wise minister carries two important traits in Chinese history that can be described in two idioms: *cha yan guan se* (察言觀色), and *shan yu jin jian* (善於進諫). The former idiom means “good at reading people”, and the latter means “skillful at offering advice to superiors", often in the form of persuading them to consider the overall situation in making a right choice, rather than merely individual benefits and drawbacks. Drawing parallels between homes and a minister or counsel, a wise home needs to be good at *reading* its occupants’ preferences, goals, emotions and other contexts. Furthermore, based on these *readings*, a wise home should provide appropriate advice to its occupants, helping them engage in positive behaviors that considers the overall situation while respecting their desires. In the context of Chinese imperial rule, a minister may be punished or demoted if an emperor did not like his advice, whether due to the content of the advice, how it was framed, or the emperor’s specific relationship with the minister. While an occupant cannot punish or demote a home, inappropriate advice may lead to abandonment of certain technologies or decreased trust in them. Therefore, different from a smart home, of which Brush et al. (2018) define as “[a home that] consists of a set of connected devices and software that can automate and control these devices”, a wise home is a home that can further integrate these connected devices and software to understand occupants' goals, preferences, contexts and to support coaching.

What is considered right or wrong is a very challenging topic to discuss. In this work, I used the Ethics Guidelines for Trustworthy Artificial Intelligence (AI) [2]— released by the European Union's High-Level Expert Group on AI — as a starting point to design a wise home. One of the guidelines suggests that AI should be designed to support "societal and environmental well-being", including being sustainable and environmentally friendly, as well as taking into account "the environment, including other living beings, and their social and society impact".

1.2 Waves of Smart Homes

Despite the claims of the 1967 film, researchers did not achieve such a dream by the actual year 1999. Three problems hindered their progress towards a wise home. The first problem was a lack of interoperability among home devices. Although more objects and appliances were embedded with computer chips to enable new functionalities, they used varied communication protocols: X10[3], Zigbee[4], Zwave[5], PLC[6], Bluetooth[7], to name a few. It was almost impossible for different devices, made by different vendors, to talk to each other. Another problem was a lack of computing power. Even though many home objects were increasingly embedded with computer chips, these chips were still quite limited. It was not

---

[2] Ethics guidelines for trustworthy AI: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

[3] X10: https://en.wikipedia.org/wiki/X10_(industry_standard)

[4] Zigbee: https://en.wikipedia.org/wiki/Zigbee

[5] ZWave: https://www.z-wave.com/

[6] Power-line communication (PLC): https://en.wikipedia.org/wiki/Power-line_communication

[7] Bluetooth: https://www.bluetooth.com/

possible to run advanced algorithms, such as machine learning models, on these objects to support more advanced interactions. Finally, the third problem was poor user interface medium available then. For a wise home to have an appropriate dialogue with its residents, it must communicate with them at appropriate moments. However, the interface of homes at the time — usually a wall-mounted panel near the doorway — was bulky, not communicative, and unable to attract users' attention when needed.

However, these barriers are lifted now. Lately, we have seen a new wave of devices joining our home spaces that offer superior interoperability. For example, the SmartThings hub[8] now integrates ZWave, Zigbee and Bluetooth in one box, and standardizes all communication with external services to HTTP. This allows easy communication between any devices and services that leverage ZWave, Zigbee, Bluetooth, WiFi and HTTP as protocols. In addition, by allowing a device to access the Internet, this new wave of devices can access enormous computing power available in the cloud, running advanced algorithms with minimal delay. Moreover, as smartphones, smartwatches, and smart speakers become increasingly prevalent, together they form a ubiquitous medium for a home to interact with its occupants at opportune moments. Thus, we live in an era with unprecedented opportunities to advance research on home automation and advance our dream of a wise home.

The Nest learning thermostat is an excellent representation of this new wave of smart home applications. It combines several entities — including smartphones and their GPS sensors, temperature sensors in different rooms (Nest Temperature Sensor[9]), thermostat,

---

[8] SmartThings: https://en.wikipedia.org/wiki/SmartThings

[9] Google Nest Temperature Sensor: https://store.google.com/us/product/nest_temperature_sensor?hl=en-US

machine learning, and the cloud — to facilitate heating and cooling control that better balances occupants' comfort and savings (financial or energy). It is marketed to learn about residents' setpoint preferences for different times of day, and to automatically respond to people's occupancy status (i.e., at home or away). Many other similar systems were studied in the literature (Gupta, Intille, & Larson, 2009; Koehler, Ziebart, Mankoff, & Dey, 2013; Lu et al., 2010; Scott et al., 2011), which shows that they increase savings by around 7-28%, and in some cases, increase comfort as well.

1.3 New Issues Emerged

However, as researchers started to build smarter homes based on these new technology foundations (i.e., better interoperability, interface, and intelligence), new problems emerged. Despite the abilities of these new applications and devices to leverage multiple sensors in capturing human contexts to drive automation, there are still hidden variables that make inference and learning difficult. For example, researchers found that the Nest thermostat's lack of understanding the intention of temperature adjustments can lead to inappropriate learning (Yang & Newman, 2013; Yang, Newman, & Forlizzi, 2014). For example, a person may adjust the temperature due to having parents staying with her for a few days. However, this does not imply that she always prefers this setting. Users' preferences can be quite nuanced (meaning user preferences are very contextual and situated, they can change given many different factors), and human routines can be equally dynamic (meaning there may not be much regularity in daily life, and routines can vary by time and season). These constraints make it challenging to design a wise home.

The above-mentioned issues are further worsened by the invisibility assumption underpinning many of these new technologies. That is, the assumption that these smart home technologies should work silently in the background to complete tasks on our behalf (Bellotti & Edwards, 2001). However, due to the lack of expectations on users to engage with these technologies, occupants are not able to identify errors when they happen, and thus not able to fix those errors in a timely fashion. Researchers have criticized this assumption and proposed an alternative design paradigm: technologies should be designed to better engage with users and better encourage engagement at specific moments (Rogers, 2006). For example, Yang et al. (2014) find that people that are more engaged with their Nest learning thermostats over a long period of time configure their thermostats more appropriately. This, in turn, lead to better savings and comfort. They therefore called for learning thermostats that can better sustain engagement with residents.

Furthermore, current smart home devices are limited to support tasks that can be achieved via simple trigger-action automation. For example, the Nest thermostat will trigger a setback temperature when occupants are away from home. However, to achieve a wise home, a home also needs to persuade its occupants to engage in positive behaviors. Although in the past decade there has been considerable work on smart homes that aims to encourage behavior change by offering behavior feedback (such as feedback on energy usage [Froehlich, Findlater, & Landay, 2010] and exercise time [Bentley et al., 2013]), these feedback devices had limited success in motivating people to take concrete actions. Being able to view previously hidden patterns does not necessarily lead to real actions (Pierce & Paulos, 2012; Pierce, Schiano, & Paulos, 2010).

1.4 Adopt Mixed-initiative Interaction in Smart Home Design

To address the issues mentioned previously, recently researchers have started to adopt mixed-initiative interaction in the design of a smarter home. While there are multiple definitions of mixed-initiative interaction (Allen, Guinn, & Horvitz, 1999; Cohen et al., 1998), in this thesis we focus on Horvitz's notion (Allen et al., 1999; Horvitz, 1999), in which an artificial agent can initiate different actions associated with different degrees of autonomy (Goodrich & Schultz, 2008) based on its understanding of user goals and contexts. For example, an artificial agent may support automation when it has full confidence of user goals and contexts, and retreat to offering suggestions when it only has partial information. It may invite direct control from users in some situations as well. Under this framework, the artificial agent and the users are engaging in a dialogue to achieve tasks together.

Mixed-initiative interaction is essential to the design of a wise home for two reasons. First, although new technologies provide a home with unprecedented opportunities to understand occupants, it would still be challenging for such a home to act appropriately on behalf of the occupants all the time; even humans have trouble doing so (Bellotti & Edwards, 2001). Since mixed-initiative interaction views the interaction between users and systems as a dialogue, it allows a smart system to converse with its users to cope with variables and contexts that are hidden from the system. Further, as mentioned previously, a wise home is also a home that offers advice to help occupants engage in positive behaviors, in addition to supporting automation. Mixed-initiative interaction allows a home to encourage positive behaviors through a conversation.

ThermoCoach (Pisharoty, Yang, Newman, & Whitehouse, 2015) is a representative work that leveraged mixed-initiative interaction in the design of smart home applications. It is a smart thermostat system that can coach occupants about potential temperature schedules to use that can lead to better savings while not hindering their comfort. Based on occupancy sensing, ThermoCoach offers different suggestions of temperature schedules with different comfort and savings tradeoffs. Preliminary studies suggest (Yang, Pisharoty, Montazeri, Whitehouse, & Newman, 2016) that a mixed-initiative system like ThermoCoach can encourage occupants to explore alternative options in decision making, helping them achieve a better balance between their various goals (e.g., comfort and savings).

1.5 Research Questions

However, studies on mixed-initiative homes is still quite limited. We do not yet know what possible actions a mixed-initiative home can take. How can a home employ mixed-initiative interactions to coach its occupants? What are the possible strategies for coaching? How can such a home learns about occupants' preferences and uses this information to offer personalized coaching?

Moreover, compared to pure-software agents such as movie recommenders, these smart home agents impose a different class of risks as they attempt to learn about user goals and preferences and act on behalf of users. Instead of just affecting the digital realm, smart homes nowadays can control physical environments through varying temperature, lighting (M. C. Mozer, 1998), energy distribution (A. J. Brush, Krumm, Gupta, & Patel, 2015), time to start laundry machines (Bourgeois, van der Linden, Kortuem, Price, & Rimmer, 2014), and many other interventions. For example, instead of merely suggesting a bad movie, a smart home agent can cause occupants to feel physically uncomfortable (e.g. too cold or too hot) as it tries to learn about their preferences on indoor temperature. How to design this class of agents with different risk characteristics remains a challenging question.

1.6 Thesis Statement

In this thesis, I argue that lightweight dialogue, proactive coaching, guided experimentation and mutual-learning are four effective strategies to support the design of mixed-initiative homes, leading a home to achieve a more optimal balance between various human and system goals.

- I define **"lightweight dialogue"** as the use of ubiquitous interfaces, such as smartwatches, to offer brief suggestions, quickly remediate errors, and collect user feedback in a low-burden fashion.
- **"Proactive coaching"** refers to a strategy in which a home can actively provide an anchor action that facilitates positive behaviors, such as an action or suggestion that increases sustainability.
- **"Guided experimentation"** is defined as an approach in which a home can facilitate occupants' exploration and evaluation of multiple potential solutions in parallel, particularly in domains where occupants do not yet know the best answer or can be biased by preconceived opinions.
- **"Mutual-learning"** refers to an approach in which a home supports its occupants' learning while it learns about their goals, preferences, and context patterns.

1.7 Thesis Approaches

In this thesis, I used adaptive heating and cooling (a sub-area of the smart home) as a lens to advance research on mixed-initiative homes. This thesis started from the observation that the increasingly available consumer-grade devices, including smartwatches and low-cost indoor sensors, provide an opportunity to advance the dream of wise homes: creating an adaptive heating and cooling system that can understand our individual comfort preferences and coach us to be sustainable while respecting our comfort needs. New smartwatches equipped with physiological sensors allow a home to potentially infer our thermal comfort and activity patterns in unprecedented ways. They also offer a lightweight interface for a home to interact with us at appropriate moments. Given these emerging technologies, we can create a wiser thermostat that I term "comfort-aware eco-coaching thermostats." That is, thermostats that can understand each occupant's comfort preferences and sensations, and use such information in conjunction with other context sensing to coach people to be more sustainable while respecting their comfort needs.

My dissertation work progressed in three phases:

In the first phase, I started from understanding the feasibility of sensing occupants' thermal comfort at home through consumer-grade sensors and via lightweight interfaces. I developed and deployed an experimental smart home system equipped with wearable and indoor sensors and conducted an experience sampling study. The central questions for this phase were: 1) how accurate could a smart home infer its' occupants' comfort level? 2) what were the challenging situations for a home to infer it's occupants' comfort? The result of this study showed that it is promising to leverage consumer-grade devices to infer thermal comfort — my system reached a 0.77 Mean Absolute Error on a 5-level thermal comfort scale in inferring occupants' comfort, superior to previous comfort modeling approaches.

However, we also found that there are situations that were challenging for a smart home system to infer thermal comfort, such as when people were sick or when they were affected by local heat sources (e.g., a laptop on a person's lap). Therefore, although we could achieve reasonable inferences by integrating various off-the-shelf devices, the uncertainty of nuanced user preferences and hidden variables necessitate further design considerations. This led to the question: could a smart home leverage such imperfect but potentially valuable comfort inferences to help people achieve better balance between savings and comfort?

In the second phase, I attempted to answer the above-mentioned question. Could comfort inferences add values to smart thermostats? To address this, I mapped out the design space of comfort-aware eco-coaching thermostats by reviewing trends and insights of heating and cooling systems in Human-Computer Interaction. To combat imperfect inferences, I focused on investigating mixed-initiative interaction as an approach to design comfort-aware eco-coaching thermostats. I drew upon prior literature to distill key design dimensions, created 15 design ideas to illustrate key points in the design space, and conducted a user enactment study to probe the feasibility of these design ideas.

The study showed that there are several possible means to use such comfort inferences. For example, a thermostat may keep track of a person's comfort score based on accumulated results of comfort inferences, and then use such a score to trigger proactive coaching suggestions, such as encouraging occupants to lower the heating temperature by 1°F for an evening to increase savings. However, our findings also showed that there is no one-size-fits-all design — people have different degrees of acceptance to these designs due to their different attitudes toward smart thermostats, and their different expectations on having user control. Personalization and customization are thus needed.

The studies conducted in these two phases also led to two insights that drove the third phase of investigation. First, for a home to better learn about user comfort preferences, it has to more actively alter the indoor temperature to probe an occupant's preference under different contextual conditions. Second, people often do not know about the best setpoints to use at different times of day, especially considering multiple members in the household. As the thermostat learns about people's comfort preferences, it should also support them in learning about their preferences. For users that desire more control, this knowledge can facilitate human-AI collaboration and manual control.

These two insights thus led to the third phase of the investigation. In the third phase, I built and deployed a comfort-aware eco-coaching thermostat that employs guided experimentation and mutual-learning as its mixed-initiative interaction strategies to coach occupants. The study showed that comfort-aware thermostats can use experimentation to help their occupants identify setpoints that can better satisfy their collective comfort needs. It can also help the occupants to better collaborate with the system. Furthermore, this study demonstrated that supporting mutual-learning is beneficial to both user-learning and agent-learning. In summary, through this study, I showed that guided experimentation and mutual-learning are two effective strategies to support mixed-initiative interaction.

1.8 Thesis Contributions

In summary, this thesis makes three contributions: design principles contribution, sensing contribution, and system contribution.

- Design principles contribution: I show that lightweight dialogue, proactive coaching, guided experimentation, and mutual-learning are four effective strategies for mixed-initiative homes to interact with their occupants.
- Sensing contribution: I demonstrate that by integrating emerging consumer-grade wearable and indoor sensors, we can achieve superior inferences on individual thermal comfort compared to prior comfort modeling approaches.
- System contribution: I show that a mixed-initiative home, which employs the four above-mentioned mixed-initiative strategies, can be built and used in typical residences. Such a home fuses data from multiple classes of sensors to obtain a better picture of user preferences, uses ubiquitous interfaces to support lightweight dialogue between occupants and users, and leverages mixed-initiative interaction to coach occupants.

Chapter 2. Background

Smart homes, broadly defined as homes that *"consist a set of connected devices and software that can automate and control these devices"* (A. J. Brush et al., 2018) have been an active area of research over the past few decades. Testbeds of smart homes were constructed to study their infrastructure, intelligence, services and everyday usage. To name a few, *The Aware Home* (Kidd et al., 1999), *House_n* (S.S. Intille, 2002) and *The Neural Network* (M. C. Mozer, 1998) are well-known examples.

Unlike the predominant devices and software created for the workplace, our vision of smart homes are often tinctured with a desire of refuge[10] — the desire of a shelter that protects us from danger or distress. Thus, beyond increasing efficiency and productivity, the devices and software created for homes are often envisioned to also *"increase the comfort of [their] inhabitants in things they already do or enable functionalities that were not possible before through the use of computing technologies"* (Mennicken, Vermeulen, & Huang, 2014). Such functionalities can range from supporting independence (e.g., Mynatt, Rowan, Craighill, & Jacobs, 2001), healthy living (e.g., Adib, Mao, Kabelac, Katabi, & Miller, 2015; Kay et al., 2012), family togetherness (e.g., Dong, Newman, Ackerman, & Schoenebeck, 2015), to sustainability (e.g., Koehler, Ziebart, Mankoff, & Dey, 2013).

Through these various applications deployed in smart home testbeds and in-the-wild, researchers have deepened our understanding of domestic computing, uncovering insights,

[10] Prospect-refuge theory. https://en.wikipedia.org/wiki/Jay_Appleton

design principles, and opportunities of smart homes (Bellotti & Edwards, 2001; A. J. B. Brush et al., 2011; Davidoff, Lee, Yiu, Zimmerman, & Dey, 2006; Edwards & Grinter, 2001; Grinter, Edwards, Newman, & Ducheneaut, 2005; Mennicken et al., 2014). For example, researchers argue that unlike the rendering of full-fledged smart homes in popular press and science fiction, in reality, our homes evolve to be smart gradually and accidentally while parts of the homes remain dumb (Edwards & Grinter, 2001). Therefore, systems and services designed with a mindset of gradual adoption — such as to provide immediate value with the installation of one device rather than a network of devices — are more likely to succeed commercially. Similarly, this insight encourages researchers to evaluate smart home innovations through deployment in-the-wild, i.e., in participants' existing homes, rather than expecting inhabitants to parachute in full-fledged smart homes (A. J. B. Brush et al., 2011).

In addition to the insights generated, the advancement of smart home research has successfully pushed two waves of commercial products, moving smart homes from research labs to the public (A. J. Brush et al., 2018). In the 2000s, there was a burgeoning of everyday devices embedded with digital screens, ranging from digital frames to Internet-connected refrigerators. Around 2015, another wave of products has been enabled by the maturation of deep learning (e.g., speech recognition), cloud computing and mobile computing. Products such as smart thermostats (e.g., Nest Thermostat), security cameras (e.g., NestCam[11]), smart door locks (e.g., August Smart Lock[12]), smart home hubs (e.g., SmartThings), and smart speakers (e.g., Amazon Echo[13]) are touted to revolutionize domestic lives.

[11] Nest Cam Indoor: https://en.wikipedia.org/wiki/Google_Nest#Nest_Cam_Indoor

[12] August: https://en.wikipedia.org/wiki/August_Home

[13] Amazon Echo: https://en.wikipedia.org/wiki/Amazon_Echo

2.1 From Today's Smart Homes to Tomorrow's Wise Homes

Despite the advancement of smart home technologies over the past decades and their commercial success, we have only begun to understand the design of homes that are augmented with AI. The current development of smart homes still falls short of its potential. While the mainstream vision often depicts a future where smart homes can automate all house chores for us, many researchers criticize this vision as inappropriate: it encourages occupants to be dumber as the home becomes smarter. Instead of solely supporting automation, in this thesis, I propose the notion of wise homes: homes that function as wise ministers to encourage positive behaviors while respecting the needs of occupants. Some prior work shares pieces of this vision, and serves as the foundation for developing such wise homes. For example, Intille (2002) argues that the ultimate goal of a smart home is not to fully replace human labor, but to teach people to be "smarter" — to be more sustainable, healthier and knowledgeable. Perhaps one scene that best captures this vision is shown in the movie *1999 A.D. House of Tomorrow* (Madden, 1967). As the protagonist sits on a lounge chair in his home gym, a computer scans his body and suggests exercises to do that would balance his food intake. Although in the movie the framing of the suggestion and the way the computer interacts with the protagonist is quite unintelligent, it does represent a wiser home that not only automates tasks but also tries to coach people to achieve their better selves.

In order to achieve this vision of wise homes, we still need to improve two aspects of smart homes. The first aspect is to design homes that do not merely accept and execute commands, and do not merely perform simplistic trigger-action automation. To create wiser homes, these homes also need to inform, coach and persuade people to help them become better persons. Currently, the commercial products of smart homes tend to be passive in accepting the commands of users or simply allowing remote access to information (e.g., security camera). These smart homes lack the capability to help occupants achieve their high-level goals, such as being sustainable and healthy. While researchers have been exploring persuasive technology in the context of homes, existing research is mostly constrained to feedback devices that display previously hidden information such as detail energy consumption (Froehlich et al., 2010). However, research has shown that offering feedback is insufficient to drive actions and facilitate positive behavior change (Pierce & Paulos, 2012; Pierce, Schiano, et al., 2010). We still need to investigate homes that can actively guide and coach users, researching the appropriate coaching strategies beyond offering feedback.

Further, to facilitate coaching, we need homes to better understand their occupants' context and preferences (M. C. Mozer, 2005). Such an understanding should also be achieved through an approach that is affordable, easily-deployable, user-friendly and respectful of people's privacy to ensure large-scale adoption. In some sense, this is not too different from the vision of home automation, of which the goal is to develop homes that better automate tasks based on a deeper understanding of context. However, a wise home would also need to obtain context and user preferences that can assist coaching, not just to assist automation.

Although research has made significant progress in deep learning that enables smart home applications to better understand speech, natural language and visual scenes such as face detection, many state-of-the-art sensing solutions are still inappropriate for home settings due to invasion of privacy, high cost of ownership, and high complexity of deployment. For example, many state-of-the-art activity recognition solutions rely on deep-learning-based computer vision, while people generally consider camera-based solutions to be privacy-invasive at home (Caine et al., 2011; Choe et al., 2012; Neustaedter, Greenberg, & Boyle, 2006).

Further, current smart home systems still fail to infer contexts and preferences that depend on sparse, subjective and in-situ feedback from end users. For example, individual preference of thermal comfort is a context that can be valuable for developing personalized thermostats (M. Feldmeier & Paradiso, 2010); Similarly, human interruptability can be an essential context for a home to reason when to proactively engage with users (Turner, Allen, & Whitaker, 2015), thus identifying better opportunities for coaching. To infer and learn these contexts and preferences, a system often needs to ask for users' subjective feedback multiple times a day over multiple weeks to collect sufficient ground truths to perform modeling. One common labeling solution is to use video camera or supplementary sensors to collect traces that can then be labeled by the users themselves or by other people (Vondrick, Patterson, & Ramanan, 2013). However, this approach is burdensome for end users and can be privacy-invasive if using other people to label the sensor traces. In some cases such as comfort, it is also impossible for end users to recall their comfort level in a post-hoc fashion — people often forget how they feel at a particular moment. This learning process thus requires a different set of tools than the tools for developing most of today's deep learning algorithms (Stephen S. Intille, Bao, Tapia, & Rondoni, 2004) — of which a third party crowd can be hired to objectively label a large amount of insensitive data. Due to the nature of these contexts, existing crowdsourcing data labeling techniques that power deep learning algorithms become insufficient, and require significant improvement.

2.2 Mixed-Initiative Homes

However, advancing context sensing and coaching is a very challenging task. Human preference is very nuanced and it can vary depending on different contextual factors such as the presence of others (Ackerman, 2000) and locations (Paruthi, Raj, Colabianchi, Klasnja, & Newman, 2018). This makes it challenging for a home to model people's preferences and to determine what to do given certain contexts. Similarly, people's routines may change dynamically given different factors: even if a pattern is found by a smart home this week to support automation and coaching, it may not be generalized to the next week (Yang et al., 2014). In brief, no matter how well a home can understand its occupants' goals, context and preferences, there will always be things unknown to the home, thus preventing it from accurately understanding user context, appropriately automating tasks, and effectively coaching its occupants. Even people have trouble understanding each other's goals, preferences and contexts. Thus to address the difficulty of context sensing, and to support coaching, researchers have started to adopt mixed-initiative interaction in the design of smart home applications.

Mixed-initiative interaction is broadly defined as interaction strategy that "explicitly supports an efficient, natural interleaving of contributions by users and automated services aimed at converging on solutions to problems." (Allen et al., 1999). It strives to leverage the best of human capabilities and the best of machine capabilities to address a problem collaboratively.

Historically the concept of "mixed-initiative interaction" emerged in the domain of dialogue-based systems, and evolved over time as it was used in different types of AI systems. In the context of dialogue-based systems (Walker & Whittaker, 1990), mixed-initiative systems refer to systems that are able to model and control the turn taking of conversation. That is, they try to infer which entity (either a person or a computer) is taking a turn in initiating a conversation at any given moment. They can also interrupt a conversation and take control of it if needed. The concept was later broadened to task-oriented systems, particularly systems designed to support planning, such as vehicle route planning. Mixed-initiative planning views (Ferguson, Allen, & Miller, 1996) a planning activity as a conversation between people to identify solutions to a shared problem. Ferguson et al. argued (1996) that unlike fixed-initiative agents which collect all the planning constraints upfront, mixed-initiative agents function more like people because they start with a "straw plan", and gradually work out a final plan by iterating on the scope, constraints and possible options through conversation. A mixed-initiative planning agent thus acts in a "look and leap" fashion (Ferguson et al., 1996). That is, it proposes a starting plan as soon as possible to the human collaborators and then works with them to revise this plan iteratively. To support revision, such an agent can criticize a plan and support users to explore different solutions. Users still have the final say on the final decision.

Horvitz (1999) further introduced the ideas of uncertainty and expected utility in the framing of mixed-initiative interaction. He proposed that the underlying reasons to create an agent to engage in different actions with different levels of autonomy (e.g., automate tasks, have a conversation with users) are because of the agent's certainty in predicting users' goals, and the expected utility of performing an action. As an agent becomes better at predicting users' goals and context, it can move from supporting manual control, to offer suggestions, to providing automation.

While in the above-mentioned definitions a mixed-initiative agent shares the same goal as its human users, more recently researchers have adopted mixed-initiative interaction in applications where the agent can have a different goal than its users. For example, researchers have adopted mixed-initiative interaction in the design of smart thermostats to encourage energy savings (Pisharoty et al., 2015). A smart thermostat may favor actions that increase energy savings while users may want to prioritize comfort. In this case, a mixed-initiative agent uses dialogue (e.g., offers suggestions) as a way to persuade its users to engage in behaviors that the agent believes to be more beneficial. However, users still have the final say about what they want to do.

Although the definitions of mixed-initiative interaction differ, they all serve as important foundations for creating wise homes. To start, mixed-initiative interaction allows homes to engage in different actions with different levels of autonomy. In addition to providing automation and supporting manual control, homes designed with mixed-initiative interaction can also have a conversation with their occupants. This conversation presents an opportunity for homes to inform and coach their occupants to engage in actions that may lead to better outcomes. For example, a home can use a conversation to offer suggestions, clarify users' goals, preferences and contexts, and provide feedback on their existing actions. Further, depending on different situations, mixed-initiative interaction allows homes to shift their actions from supporting manual control, to engaging in a dialogue and to supporting automation. A wise home may not always need to talk to its users. In some situations it can support automation once its users selected a suggested option. Similarly, in other situations a wise home may retreat back to supporting manual control after offer suggestions to its users. Finally, being able to engage in a dialogue with users also allows homes to re-evaluate their current plans every once in a while, to make sure the homes and the users are still on the same page, to help correct errors if the plans deviate from users' expectations (Yang et al., 2014), and to coach and remind users about the rationales of the chosen plans — this knowledge of rationales may help users generalize an plan to other tasks that are not in control by the system (Brynjarsdottir et al., 2012).

2.3 Heating and Cooling as a Lens

In this dissertation, I focus on residential heating and cooling control as a site of investigation to explore the design of a *wise* home. Heating and cooling control is an appropriate starting point because it faces similar challenges for researchers who are designing smart homes toward the vision of wise homes. Design principles generated from studying this particular domain may be applicable to other application domains of smart homes.

2.3.1 Lack of tools to achieve better balance between goals

Heating and cooling control is an appropriate domain to study because we still lack the tools to help people achieve a better balance between comfort and energy savings. Researchers have found several issues with existing practices of heating and cooling control. They have pointed out that different people living in identical houses can use widely different amount of energy. Even living in identical houses, the occupants with consumption at the top can consume more than two times of the bottom (Sonderegger, 1978). A significant portion of this difference is caused by different usage habits, rather than the physical characteristics of the houses. Further, research found that 40% of people do not currently use a daytime setback temperature (EIA, 2015), 74% do not use a nighttime setback temperature (EIA, 2015), and 17% of people use a default heating setpoint higher than the common 68-72°F recommendations (Parker, 2013).

2.3.2 Eco-feedback displays

To help people better balance their comfort and savings needs, in the past few decades, researchers have designed and investigated numerous eco-feedback displays (Arroyo, Bonanni, & Selker, 2005; Erickson et al., 2013; Gustafsson & Gyllenswärd, 2005;

Houwelingen & Raaij, 1989; Hutton, Mauser, Filiatrault, & Ahtola, 1986; Kappel & Grechenig, 2009; Kuznetsov & Paulos, 2010; Matsukawa, 2004; Seligman & Darley, 1977; Winett & Kagel, 1984). According to Froehlich et al. (2010), eco-feedback displays are devices that "provide feedback on individual or group behaviors with a goal of reducing environmental impact". While these devices are not limited to reducing consumption caused by heating and cooling, surfacing usage information and environmental impact associated with the usage can affect people's heating and cooling usage. Some modern thermostats also display eco-feedback specific to heating and cooling usage. For example, the Nest thermostats display the amount of time heating and cooling systems are in use day by day and provides information on how the use of a household compared to other households.

Research has shown that households with eco-feedback devices use significantly less resources compared to other households. The precise effect of the technology varied study by study, but researchers have demonstrated a scale of saving for 4-10% (Erickson et al., 2013, 2012; Houwelingen & Raaij, 1989; Hutton et al., 1986). However, researchers also found that simply showing consumption feedback is insufficient to drive actions in many scenarios — people may not be aware of the actions to take to reduce consumption, even if they do, they may be unwilling to do so as the effort to manually perform such actions is simply too high (Pierce, Fan, Lomas, Marcu, & Paulos, 2010; Pierce & Paulos, 2012).

2.3.3 Programmable thermostats

In addition to eco-feedback devices, programmable thermostats have also been developed in the past few decades to reduce energy consumption. Unlike eco-feedback devices which simply surface usage information, programmable thermostats attempt to reduce the burden of manual user control by automatically triggering energy-saving setback temperatures based on predefined setback schedules. Users are thus able to set their thermostats to use more energy-saving temperatures when they are away from home and when they are sleeping. Theoretically such thermostats should be able to achieve significant savings if people configure their setback schedules appropriately (Plourde, 2003), however, research has shown that in reality people fail to do so due to the difficulty to configure setback schedules on these thermostats' poorly-designed user interfaces (Meier, 2012; Peffer, Pritoni, Meier, Aragon, & Perry, 2011). In addition, these programmable thermostats often fail to catch up with occupants' changing routines, leading to wasted energy (Peffer et al., 2011).

2.3.4 Reactive thermostats

To improve heating and cooling control, there have been several studies in the past decade to develop *reactive thermostats* — thermostats that can automate control based on various contexts, such as occupancy status (Beltran, Erickson, & Cerpa, 2013; Gupta et al., 2009; Koehler et al., 2013; Lu et al., 2010; Scott et al., 2011; Sookoor & Whitehouse, 2013). In addition to simply reacting to inhabitants' occupancy status, some of the works seek to predict the occupancy trajectory based on past occupancy data, thus to preheat houses before their residents arrive to increase comfort (e.g., Koehler et al., 2013). These systems have been shown to reduce 7-28% of energy compared to an always-on thermostat (Gupta et al., 2009; Koehler et al., 2013; Nest, n.d.).

While reactive thermostats have been shown to be successful in reducing energy consumption, there are still issues with these thermostats that prevent them from achieving an even better balance between inhabitants' needs of comfort, energy savings, and sustainability (Yang et al., 2014).

2.3.5 Lack of coaching

The first issue is that many reactive thermostats passively interact with users without informing or suggesting them alternative means to achieve a better balance between comfort and energy savings, such as using a more appropriate setpoint when occupants are at home and asleep. Occupants often lack an understanding of the impact of their actions on household members' comfort and on the environments (Froehlich et al., 2010). In addition, they are often not aware of the alternative actions to take that can lead to better control (Yang et al., 2016). Even if they are aware of the actions, they often lack the motivation and time to optimize their heating and cooling control because of the limited tools available. When their mental models are unchallenged, and when there is a lack of triggers, their behaviors lead to less optimal usage of heating and cooling at home (Meier, 2012; Yang et al., 2014).

Recently researchers have therefore started to explore eco-coaching thermostats, namely, smart thermostats that coach people to better reduce wasted energy (Pisharoty et al., 2015; Yang et al., 2016). For example, Pisharoty et al. (2015) created a smart thermostat that uses occupancy sensing to generate different temperature schedule options. Some options may result in higher energy savings while others will result in higher comfort level. Pisharoty et al.'s study showed that embedding coaching in thermostat design is a promising approach to encourage occupants to use more energy saving settings (Pisharoty et al., 2015; Yang et al., 2016). However, the design of eco-coaching is still very limited with respect to the strategies used for coaching and the knowledge domains supported by coaching.

Recently researchers have also started to explore coaching systems that encourage adaptive behaviors that people can use to adjust their thermal comfort locally, such as opening windows and adjusting clothing. For example, Clear et al. (2014) built a system to encourage thermal adaptivity by gradually drifting setpoints to more energy-saving ones and limiting the activation of heating to only 90 minutes at a time. This design forces occupants to consider other ways of achieving comfort, instead of simply setting their dormitory rooms to have the same temperature all the time.

This thread of research on adaptive behaviors (A. K. Clear, Morley, Hazas, Friday, & Bates, 2013) is motivated by a design paradigm in building science, which views *comfort as a goal* (i.e., occupants have the agency to achieve individual comfort goals through many means), instead of *comfort as a product* (i.e. comfort is a product delivered to the occupants by the built environments) (J. Fergus Nicol & Humphreys, 2002). In building science, the result of this shift is that thermal comfort regulation for public and commercial buildings start to loosen (Humphreys, Nicol, & Raja, 2007; J. F. Nicol & Humphreys, 2009). Further, green buildings are constructed to rely less on mechanical heating and cooling ventilation but on the use of shading, and on many other means to achieve comfort, such as encouraging occupants to adjust their own clothing or opening windows. Under such a design paradigm, designers create buildings that can better collaborate with their occupants to co-achieve their comfort and sustainability goals (Baker & Standeven, 1996; Brager & de Dear, 1998; J. Fergus Nicol & Humphreys, 2002). While there is progress in the architecture and building science domains, such a concept has not been explored deeply in smart homes and in developing smart heating and cooling systems.

2.3.6 Insufficient context sensing

The second issue with the existing work on thermostats is that these thermostats still have a very limited understanding of user context, particularly contexts that can affect people's heating and cooling preferences. These thermostats assume inhabitants use a fixed temperature setpoint when they are at home, while in reality people's comfort and setpoint preferences are dynamic depending on various factors (Fanger, 1970), such as indoor temperatures in different rooms, humidity, and activity (e.g., cooking and indoor exercise), and the presence of other occupants. Some recent research has started to consider these additional factors that affect people's comfort preferences, such as time of day[14] and dynamic pricing *(Alan, Shann, Costanza, Ramchurn, & Seuken, 2016)*. However, the factors they considered are still quite limited. Creating wiser thermostats faces similar challenges to creating wiser homes. We still need to develop systems that can leverage the emerging sensors and devices to better understand user preferences and context, thus better supporting coaching and temperature control.

While there have been many studies on preference elicitation (Braziunas & Boutilier, 2008; Pommeranz, Broekens, Wiggers, Brinkman, & Jonker, 2012; Pu & Chen, 2008), the

[14] Nest Thermostat: https://en.wikipedia.org/wiki/Nest_Learning_Thermostat

majority of the studies to date are limited to recommender systems such as systems that support flight searching, product discovery and movie finding. However, the approaches to elicit preferences in such application settings are quite different from eliciting comfort preferences to support heating and cooling control of smart homes. Unlike rating a movie, people often cannot recall how comfortable they feel given an environmental condition, they may also be biased by perceived opinions (e.g., believing 68°F to induce discomfort while in reality they may accept such a temperature setting). Further, a smart home also has the ability to vary the physical characteristics of homes such as indoor temperature to facilitate preference learning, while a movie recommender can only control the digital realm. Therefore, the associated cost of incorrectly learning a preference is quite different in the case of comfort than in the case of movie recommenders. A thermostat may cause people to feel unbearably cold, a worse outcome than suggesting a wrong movie. Therefore, to advance preference learning and context sensing of smart homes and smart thermostats, we still need more sophisticated framework to elicit user feedback that leverages and integrates the emerging interfaces, sensors and devices.

As we see, in advancing heating and cooling control to further reduce wasted energy and to increase comfort level, we encounter the challenges of context sensing and coaching. Heating and cooling control is therefore a specific manifestation of the boarder issues of wise homes. The design principles and other knowledge uncovered in designing wiser heating and cooling control can potentially be generalized to other domains of homes, supporting the developing of wise homes.

Chapter 3. The Potential and Challenges of Inferring Thermal Comfort at Home Using Commodity Sensors[15]

3.1 Introduction

Researchers have been exploring ways to model human thermal comfort for more than 40 years. Such exploration is driven by the potential for increasing people's quality of life (e.g., improving building comfort quality); reducing energy consumption (e.g., intelligent thermostats that react to people's comfort level [Feldmeier & Paradiso, 2010]); and advancing knowledge of the connection between physiological and psychological factors regarding comfort. While prior work on modeling thermal comfort such as Predicted Mean Vote (PMV) (Fanger, 1970), and Adaptive Thermal Comfort (J. Fergus Nicol & Humphreys, 2002) provide insights into the major factors affecting people's comfort, there remain several challenges for inferring thermal comfort in real homes and offices. First, devices such as near-body temperature sensors traditionally used to collect measurement data are bulky and cumbersome for people to carry or wear on their body. Second, trained human observers or extensive questionnaires have generally been needed to record data that are difficult to detect with available sensors such as clothing insulation and activity level. Various activities at home such as cooking, dining or cleaning are known to affect one's thermal comfort, but these factors have been difficult to capture without the presence of human observers. Finally, most of the previous models were designed to serve large groups of people (e.g., occupants of an office building) rather than individuals or small groups,

[15] This work was published at the UbiComp 2015 conference.

and thus are not ideal for settings that contain only a few people, such as the home.

These limitations have prevented researchers from developing techniques that could continuously infer one's comfort in naturalistic settings, especially for places where people conduct varied activities and exhibit adaptive behaviors. Furthermore, these limitations make such models unsuitable for UbiComp applications such as intelligent thermostats that intend to improve occupants' comfort by responding to present conditions (A. K. Clear et al., 2013).

Recently, wearable fitness trackers and smart home sensors have become widely available (e.g., Fitbit[16], SmartThings[17]). Although wearable fitness trackers are designed to monitor physical activities or health status, we observe that they could also capture several factors that are influential to human thermal comfort, such as activity level, ambient temperature, and sweat secretion. These emerging sensing devices provide the opportunity to address several challenges with prior approaches and permit the instrumentation of everyday home environments to gather data for inferring thermal comfort. By integrating commodity wearable and in-home sensors, we envision a system that is able to infer thermal comfort in-situ and at home with minimal setup and disruption.

While we are not the first to consider comfort sensing in naturalistic settings, our work offers three novel contributions: First, we present a new technique to infer thermal comfort using off-the-shelf wearable and in-home sensors in a domestic environment. Using wearable sensors allows us to automatically obtain data important to inferring individual's thermal comfort that was previously difficult to obtain, as well as new sources of data that could prove valuable including physical movement (which can be converted into Metabolic

---

[16] Fitbit: https://www.fitbit.com

[17] SmartThings: https://www.smartthings.com/

Equivalent of Task—a factor in PMV models), sweat level (inferred from galvanic skin response (GSR)), and skin temperature.

Second, we demonstrate the feasibility of our technique by conducting a 4-week sensor deployment and ESM study (Hektner, Schmidt, & Csikszentmihalyi, 2007) in seven households. The findings from our feasibility study validate the potential of this new technique in inferring personal thermal comfort at home under naturalistic conditions.

Our third contribution identifies six situations that pose challenges for inferring people's thermal comfort at home. We believe that these situations will be challenging for any system that aims to automatically infer thermal comfort in naturalistic settings.

3.2 Related Work

3.2.1 Models of Thermal Comfort

Numerous studies have been sought to improve our understanding of human thermal comfort. Such studies have covered the influence of physiological factors, acclimatization, and culture on thermal comfort (Baker & Standeven, 1996; Beizaee & Firth, 2011; Brager & de Dear, 1998; Fanger, 1970; Humphreys & Fergus Nicol, 2002; Jones, 2002; Shove, 2003). Researchers (Gagge, Stolwijk, & Hardy, 1967) have investigated the effects of these factors in steady-state and in transition (e.g., changing from cold to warm); applied heat transfer theory to derive formulas for thermal comfort (Fanger, 1970); and used machine-learning to infer comfort models (Megri, Naqa, & Haghighat, 2005). While most of the early research regarded people as passive dwellers with no control of the environment, more recent studies have demonstrated the importance of viewing people as active agents who actively configure the environment to maintain their comfort level (J. Fergus Nicol & Humphreys, 2002). While it is impossible to discuss all the thermal comfort models explored in prior literature, we briefly describe two that are the most widely used and relevant to our work—Predicted Mean Vote (PMV) and Adaptive Thermal Comfort (J. Fergus Nicol & Humphreys, 2002).

Developed by Fanger (1970), PMV operates on the assumption that human thermal comfort is achieved when thermal load, skin temperature and sweat secretion are within a comfortable boundary, given an activity level. Based on these assumptions, the Predicted Mean Vote provides an index that combines six parameters deemed essential to thermal comfort—air and radiant temperatures, humidity, wind velocity, clothing level and metabolic rate. Given these parameters, the model can produce an index ranging from -3 (cold) to +3 (hot), indicating the thermal comfort quality of a building environment. Although this index has been shown to be accurate for buildings with central heating, ventilating, and air conditioning (HVAC) systems and to obtain the mean thermal comfort of large groups of people, it has not worked well for buildings without centrally-controlled systems or for individuals (Jones, 2002; J. Fergus Nicol & Humphreys, 2002).

On the other hand, Adaptive Thermal Comfort (J. Fergus Nicol & Humphreys, 2002) provides an explanation for why such deviation from PMV exists—people are actually more tolerant to warm and cold conditions than PMV predicts. The primary reason is that people adapt by various means, such as opening windows or changing clothes. This suggests that buildings that support adaptive behaviors, like the home, could allow a wider thermal comfort zone.

3.2.2 Thermal Comfort Sensing at Home: Challenges and Opportunities

In this work, we focus on inferring personal thermal comfort at home continuously while allowing occupants to behave naturally. This focus is important for three reasons. First, having a better technique to infer personal comfort at home could make a significant impact, as research has shown that people spend most of their time at home—around 15.6 hours per day in the U. S. (Leech, Nelson, Burnett, Aaron, & Raizenne, 2002). A better way to infer personal comfort at home could potentially improve the comfort of home residents. For example, intelligent thermostats could potentially be developed that react to occupants' comfort level in real-time (M. Feldmeier & Paradiso, 2010). Second, to the best of our knowledge, there is no existing technique that allows inferring of thermal comfort at home while occupants are conducting their natural routines and exhibiting adaptive behaviors. Third, models such as PMV have been shown to be inaccurate for inferring comfort for a small group of people (Jones, 2002). The average U. S. household is only 2-3 people (*U. S. Census Bureau, American Community Survey, 5-Year Estimates*, n.d.), a group too small for models such as PMV to make accurate inference.

Additionally, there are two major barriers that limit previous research from reaching into everyday households. First, many previous studies relied on cumbersome equipment such as bulky near body temperature and humidity sensors (Baker & Standeven, 1996; Beizaee & Firth, 2011). These devices were inconvenient for users to carry continuously. Second, trained observers or extensive questionnaire were often required to collect data that are influential to inferring thermal comfort, but challenging to detect via sensors, such as activity level and clothing insulation (Baker & Standeven, 1996). Due to these barriers, it would be infeasible to apply previous sensing methods in UbiComp applications. For example, Clear et al. (2013) have outlined several possible applications that use thermal comfort as a system input, including, for example, a thermal comfort portal that allows people to reflect on their practice of maintaining thermal comfort.

3.2.3 Comfort Sensing Systems for Naturalistic Settings

To infer human thermal comfort in naturalistic settings, sensors and other tools used for data collection must be minimally disruptive, blending into people's everyday routines. Although a less explored area, some research has investigated various approaches to reach this goal. For example, Feldmeier and Paradiso (2010) developed a system that continuously infers comfort in naturalistic settings. While their studies were conducted in an office setting, it would be feasible to apply their approach in the home, inasmuch at the sensors used are amenable to the home environment. They developed a machine learning-based model that predicts thermal comfort based on input from wearable and embedded indoor temperature as well as humidity sensors. The predictions were then used to control the HVAC system in a large (zoned) office building. However, their models incorporated limited physiological information since they collected only room temperature and humidity with the worn sensors. None of the other essential physiological factors such as metabolic equivalent were included in the models, nor were those factors captured by the sensors used in their study. Additionally, their system was deployed in an office building, thus how to adapt it to perform effectively in home is still a question.

SPOT (Gao & Keshav, 2013b) and the system developed by Nouvel and Alessi (2012) also aim at inferring personal comfort in naturalistic settings. SPOT uses the Microsoft Kinect's skeleton tracking capability ("Kinect," 2015) for inferring metabolism. The system categorizes the skeleton information into four types of posture, namely reclining, seated and relaxed, sedentary activity, and standing. A predefined metabolic rate is then mapped to each of the postures. However, one's metabolic rate might vary significantly in certain postures (e.g., "standing"), meaning that estimates can be wildly inaccurate. The system developed by Nouvel and Alessi expects people to provide explicit comfort feedback whenever their metabolic rate or clothing level is changed. Because people shift between different activities and clothing levels relatively often, asking people to report changes in their metabolic rate or clothing level would seem impractical for a sustained deployment. Finally, these systems have only been deployed in offices, and for only one or two participants. Therefore, it is unclear whether such systems would work in the home.

In the following sections, we will first introduce our experimental system for investigating the potential of thermal comfort sensing in naturalistic settings using commodity wearable devices. Then, we will present our study method and the findings from our analyses.

3.3 Sauna: Experimental Comfort Sensing System

To explore the potential of our approach, we developed an experimental system, Sauna, for collecting the required sensor data and user comfort feedback, as well as for performing the thermal comfort inference. Our experimental comfort sensing system contains five components: (1) HomeHub, (2) wearable sensors, (3) in-home sensors (4) a mobile ESM tool and (5) a web-based diary tool. After the introduction of these five components, the rationale for our sensor selection is then provided.

HomeHub: The HomeHub is the central component for collecting data from the wearable and the in-home sensors. A notebook PC is used for the hardware of the HomeHub (we used an ASUS X200MA), and we built the software of the HomeHub on top of the Lab of Things framework (A. J. B. Brush, Jung, Mahajan, & Scott, 2012).

Wearable Sensors: We employ the Basis B1[18], a wrist-worn fitness tracker, for collecting data useful for comfort prediction. Several of the data sources provided by the Basis B1 (i.e., activity level, skin temperature, and galvanic skin response) have not been employed in previous efforts to predict thermal comfort in real time, and so represent new potential sources of information. Basis B1 measures skin and near-body air temperatures, galvanic skin response (GSR), heart rate, step count, and estimated calorie consumption, calculated once per minute. We use GSR to approximate sweat level. We further use the per minute calorie consumption offered by Basis B1 and the weight of the individual to approximate a person's metabolic equivalent of task (MET). One study has shown that Basis B1 is able to estimate energy expenditure with 76.5% accuracy (Lee, Kim, & Welk, 2014). However, this study was conducted before the introduction of BodyIQ technology, which was available at the time of our study. In a forum post[19], the lead researcher of Basis claimed that with BodyIQ, the accuracy outperforms other major competitors, some of which claim around 90% accuracy.

Basis B1 synchronizes with the user's smartphone, which uploads the data to the Basis cloud storage. We automatically fetch the sensor data from Basis's web service every 15 minutes, which is the maximum rate attainable via Basis' API.

[18] Basis B1: http://www.mybasis.com/

[19] https://support.mybasis.com/hc/communities/public/questions/201390174-Accuracy-of-the-calories-burned

In-home Sensors: We use the AeoTec MultiSensor[20] to track room-level air temperature and humidity. The HomeHub samples each MultiSensor every 3 minutes via Z-Wave.

Mobile ESM Tool: To collect an individual's thermal comfort feedback given different environmental and bodily conditions, we employed an Android-based Experience Sampling Method (ESM) tool called Minuku (Chang, Paruthi, & Newman, 2015). We configured Minuku to send brief questionnaires to participants based on time and location (e.g., "at home"). Each questionnaire asked for the person's thermal sensation, comfort sensation, current activity, indoor location, clothing level, and brief notes that might help them recall the reasons for their sensation and comfort report when completing the end-of-day diary entry. We developed our ESM strategy to minimize interruption, while collecting enough information related to people's thermal comfort to help them to recall more in-depth information later in the day. Our implementation of ESM allows us to collect not just sensor data and user comfort feedback, but also users' explanation of their perception, such as the reasons that might have caused their discomfort. This extra information allows us to investigate the feasibility of our technique in real households and identify potential challenges and opportunities for new approaches.

Web-based Diary Tool: At the end of each day, participants were asked to provide more information to explain the thermal reports they submitted throughout the day. A web-based diary tool (see Figure 1, right) was designed to facilitate these diaries. This tool displays an individual's thermal reports created throughout the day, as well as visualizations of the sensor streams that could help the person recall his indoor location and activity.

Sensor Selection: We chose the types of sensors based on the widely-used PMV model,

---

[20] https://aeotec.com/z-wave-sensor/

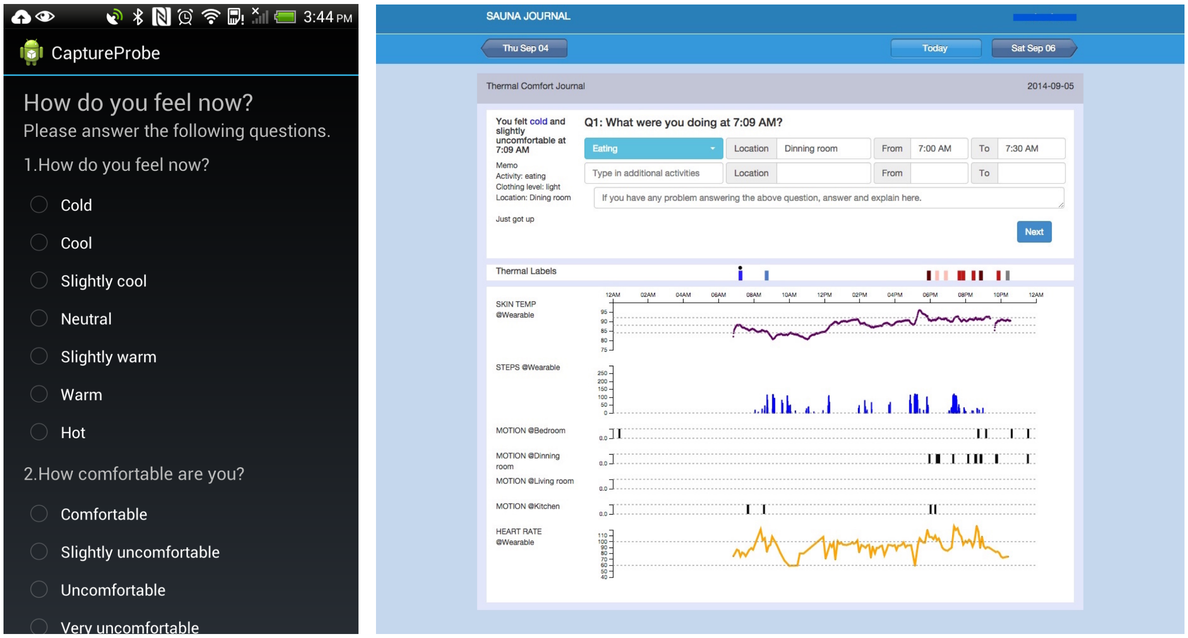


Figure 1: (Left) The ESM interface; (Right) The Web-based Diary Tool.

which states that the primary factors influencing a person's thermal comfort include air temperature, radiant temperature, wind velocity, humidity, metabolic rate and clothing level. In addition to these six parameters, we also track skin temperature and sweat level, which are implicit in the PMV model: when calculating PMV, these two additional factors can be approximated by the six basic parameters using other models (Fanger, 1970) developed in human thermal comfort, rather than directly measured.

While some factors such as radiant temperature and wind velocity are challenging to track precisely, especially in large buildings or outdoors, our setting of a single family home allows us to make a few reasonable simplifications. First, we assume the radiant temperature is the same as the air temperature, since rooms in houses are relatively small. Second, we designate a fixed wind velocity (0.2 m/s) based on the average winds speed of indoor ventilation (between 0.05 to 0.4 m/s) (Thorshauge, 1982). Finally, clothing level is obtained through participants' self-reports, collected with each ESM response. Four options, ranging from nearly naked to heavy clothing level, were provided, along with examples.

While PMV is known to have limited predictive power for individuals (J. Fergus Nicol & Humphreys, 2002), we feel that the factors in the model are comprehensive and well-studied, thus serving as a good basis for our work. We will further discuss how we incorporate Adaptive Thermal Comfort and dynamic transitions by using a machine-learning based approach, along with a person's previous body and environmental states in the data pre-processing section.

3.4 Study Method

The goals of our study were to explore the feasibility of our approach and to investigate the potential challenges of inferring thermal comfort at home in naturalistic settings. We recruited 14 participants from 9 households in Michigan. 11 of the participants from 7 households were able to complete the study. Three people dropped because of important family events that reduced the time they could stay at home. The recruitment was done through Craigslist, mailing lists and the snowball sampling method. We recruited participants who have an Android phone and who stay at home during waking hours for at least 5 hours a day on average. In addition, we recruited half of the households to have two participants in order to explore individual differences with respect to thermal comfort. Participants were compensated depending on the number of reports they provided during the ESM study. The amount of each person varied, with the average compensation being US$44.90. For households with multiple participants, we compensated each of them with US$10 extra.

Our study took place between 30th August and 4th October, 2014, and consisted of a semi-structured initial interview, followed by a four-week sensor deployment and ESM study. We then conducted exit interviews. During the initial interview, we collected information on participants' daily schedules and how satisfied they were with regard to their thermal comfort in the different rooms in their houses. We used participants' daily schedules to help determine the rooms in which to place the multi-sensors. Additionally, we also collected information on how they used their thermostats and other comfort-related appliances such as fans and dehumidifiers, as well as their attitudes towards the trade-offs between saving energy and remaining comfortable.

After the interview, we installed 4 multi-sensors at different locations in each participating house and provided each participant with a Basis B1. Participants were asked to wear the Basis B1 whenever they were awake and at home. In addition, the ESM tool was installed on participants' phones. To ensure every participant understood how to use the tools, we guided each of them to create one thermal comfort report using the ESM tool and to provide a detailed comment using the web-based diary tool.

Immediately after the sensors were deployed, the ESM study was started. During the four-week ESM study period, participants self-reported their thermal sensation, comfort sensation, location within the home, and their activity information. Sensor data detected by the multi-sensors and Basis B1 were stored in our database. Participants were not

Table 1: Participant Information
(*: one member left in the middle of the study; Prog: Programmable Thermostats)

Parti.	Gen-der	Valid Reports	House-hold	House Size (sqft)	Adult (Child)	Type of thermostat
P1	F	187	H1	1400-2000	4	Manual
P2	F	98	H1	1400-2000	4	Manual
P3	M	138	H2	< 800	2	Manual
P4	F	91	H2	< 800	2	Manual
P5	M	143	H3	< 800	2*	Manual
P6	M	131	H4	1400-2000	2*	Nest
P7	F	113	H5	800-1400	2	Prog.
P8	F	10	H6	800-1400	2 (1)	Manual
P9	M	2	H6	800-1400	2 (1)	Manual
P10	M	107	H7	800-1400	2	Prog.
P11	F	112	H7	800-1400	2	Prog.

exposed to any predictions we made, and no prediction results were used to make any changes to their home HVAC system during the study.

For self-reports, we used both the 7-level thermal sensation index introduced in PMV, and also a standard 4-level comfort sensation index used in several thermal comfort studies (e.g., Gagge et al., 1967). Using two indices allowed us to resolve the ambiguity of the labels in the thermal sensation index: for many people, the “cool” label might actually represent a comfortable and preferred feeling.

Our ESM tool prompted each participant approximately every 30 minutes when s/he was at home and during a pre-specified awake time window. Participants could ignore individual prompts, but they were expected to answer at least 6 reports per day. Participants could also actively report whenever they like, although we encouraged them to only initiate a report when they felt uncomfortable. At the end of the study, we found out that some participants had deactivated the GPS tracking of the ESM tool because it drained too much battery from the phone. Some of them therefore only initiated thermal reports, rather than responding to prompts. It’s possible that participants created different number of reports related to uncomfortable situations because of this issue, but we expect no effect on the validity of their answers.

A 30-minute exit interview was conducted following the ESM study. Before the exit interview, we calculated the PMV index of each thermal report a participant created using Fanger’s approach (1970). We used Fanger’s PMV rather than our own model’s prediction because our model was not finalized until after the exit interviews were conducted. While PMV is inaccurate for predicting individual’s thermal comfort, the calculations allowed us to investigate situations in which there were large differences between the PMV and participants’ reports. During the exit interviews, we asked participants to recall what happened at the moment of a report, and the potential reasons for the wrong PMV prediction. To facilitate recall, we asked the participant to review their diary entries and comments.

3.5 Data Processing

From the 11 participants who completed the 4-week study, we collected 1,431 thermal comfort reports (details are provided in Table 2). However, only 1,132 thermal comfort reports were considered to be complete— i.e., containing all the associated information, including thermal sensation, comfort sensation, indoor location and activity, as well as data detected by sensors, including air temperature, humidity, skin temperature, near body air temperature, GSR, and metabolic rate at the moment of report. In addition, only 9 of the 11 participants created more than 90 valid reports; the other two participants were not able to properly maintain the synchronization between their Basis B1 devices and smartphones due to software configuration issues, thus rendering many of their reports useless for training our models.

The "comfortable" sensation dominated the reports: within the 1,132 reports, 50.6% of them are labeled as "neutral" on the thermal sensation index, and 76% of them are labeled as "comfortable" on the comfort sensation index. This may be due to the fact that the weather in the study area was unusually mild during the study period, with average temperatures at 16°C, (max 33°C; min 0°C). However, there was one week (13 Sep. to 20

Table 2: Frequency of reports for each level of our combined 5-level thermal comfort index.

	Comfort Sensation	Thermal Sensation	Thermal Comfort Index	#reports	%reports
1	Uncomfortable/Very uncomfortable	Cooler than neutral	Uncomfortably cold (UC-Cold)	76	6.7%
2	Slightly uncomfortable	Cooler than neutral	Slightly uncomfortably cold (S-Cold)	43	3.8%
3	Comfortable		Comfortable (COM)	862	76.1%
4	Slightly uncomfortable	Warmer than neutral	Slightly uncomfortably warm (S-Warm)	61	5.4%
5	Uncomfortable/Very uncomfortable	Warmer than neutral	Uncomfortably warm (UC-Warm)	90	8%

Sep.) that the temperature dropped below normal. The average temperature in that week was 11°C (max 26°C; min 0°C).

The percentage difference between thermal and comfort sensation also confirms the usefulness of having two indices. For example, some people interpreted “slightly cool” or “cool” as a comfortable and preferred temperature.

3.5.1 Feature extraction

For each of the reports, we further obtained features related to an individual’s previous state, including the metabolic rate, skin temperature, near body air temperature, and sweat level (GSR) 30 and 10 minutes before the report. For each of the features, we smoothed the data by taking an average over a five-minute window. These features were inspired by (Gagge et al., 1967), which demonstrates that, in addition to the factors modeled by steady-state thermal comfort models like PMV, dynamic transitions between warm and cold environments also affect people’s thermal sensation. However, one problem we faced when extracting these features was that some reports had no data from the previous state. For example, when a participant had just woken up in the morning and worn his or her wearable sensors for just a few minutes, there would not be any sensor data available for the previous 15 minutes. To handle this problem, we filled in missing values with the average sensor reading for that participant. Finally, we extracted the temperature and humidity data by referencing the room information participants provided in the thermal report. We smoothed the room temperature and humidity data by taking the average over a 30-minute window. If the room did not contain a multi-sensor, we then averaged the sensor readings from the two adjacent rooms in which we had multi-sensors deployed.

We further categorized the activity provided by the participants using their free text descriptions reported through the diary tool. As we already captured activity-level through the measurement of metabolic equivalent of task, activity type information was collected primarily to investigate the psychological influence of activities. We generated our activity categories by combining the American Time Use Survey Activity Lexicon ("ATUS Activity Coding Lexicons," 2012) with our own heuristics of which activities might effect an individual's thermal comfort sensation, such as "cooking" (see Table 3).

3.5.2 Thermal comfort index

As no standard scale exists in the previous literature for training machine-learning based comfort models, we created a 5-level thermal comfort index for this work. Our 5-level index combines the 4-level comfort sensation index suggested by Gagge et al. (1967) and the 7-level thermal sensation index used in PMV. Table 2 illustrates the mapping from the thermal sensation and comfort sensation indices into our 5-level index. This mapping resulted in a scale ranging from, *Uncomfortably-Cold* (UC-Cold) to *Uncomfortably-Warm* (UC-Warm) at the extremes, with Slightly-Uncomfortably-Cold (S-Cold), Comfortable (COM), Slightly-Uncomfortably-Warm (S-Warm), as intermediate levels.

The intuition behind this transformation is that thermal sensation alone cannot represent an individual's complete thermal comfort assessment. For example, some participants actually interpreted "cold" as a comfortable feeling at times, and most of them enjoyed being "slightly cool". Comfort sensation, on the other hand, only provides the intensity of discomfort, but not the direction of discomfort. However, from our exit interviews we found out that participants tended to have a common interpretation of the comfort sensation index. They interpreted uncomfortable as situations in which they would definitely take action to change the thermal environment, while slightly uncomfortable meant they were uncomfortable but did not experience an urgent need to take action. Therefore, by combining these two indices, we could obtain an index that captures both the intensity of discomfort and the warm-cold direction of the discomfort.

Table 3: Activity Categories

relaxing and leisure, working at home, eating, housekeeping, cooking, waking up, grooming, arriving home, socializing, exercising, sleeping, feeling sick, other

Besides the advantage of integrating the intensity and the warm-cold direction of discomfort, we believe our index is an improvement upon the 3-level scale used by Feldmeier and Paradiso, and the 7-level index used by Nouvel and Alessi (2012). First, compared to the 3-level scale, we allowed participants to report discomfort that was relatively tolerable (e.g., S-Cold, S-Warm), thus taking into account people's tolerance to warm and cold situations, as suggested by Adaptive Thermal Comfort. Second, compared to the 7-level scale, we maintain a smaller set of thermal classes, thus facilitating the training of machine-learning based models.

3.6 Analyses & Findings

Two analyses were conducted in this research. In the first analysis, we developed our comfort model using a machine-learning approach. We compared the accuracy of different feature sets, as well as the accuracy of our approach compared to other baseline models inspired by previous research (Fanger, 1970; M. Feldmeier & Paradiso, 2010). In the second analysis, we explored the challenging situations for inferring thermal comfort at home in naturalistic settings by examining cases where our model failed to make an accurate prediction.

3.6.1 Analysis 1: Accuracy of Comfort-Aware Model

To develop our comfort model, one step is required before the training to handle our imbalanced dataset. In our dataset, 76% of the data are labeled as COM on our 5-level thermal comfort index as shown in Table 3. The remaining classes each cover between 4-8% of the data. To prevent training a classifier that always predicts comfortable, we adjusted the weight of each datum given its class to create equally balanced class labels (following (Kotsiantis, Kanellopoulos, & Pintelas, n.d.)).

3.6.1.1 Baseline models

Random Forest and Gaussian-kernel SVM were selected to train our models, as these two methods are most likely to perform the best on various data sets (Fernández-Delgado,

Table 4: Mean Absolute Error (MAE) and Mean Squared Error (MSE) for Random Forest (RF) and Support Vector Machine (SVM) models with different feature sets. Three baselines are provided for comparison, as described in the text. The best performing model (SVM+BASE) is highlighted with a *.

	Machine Learning Models (RF: Random Forest)								Baselines		
	RF +BASE	RF +NO-CLO	RF +ACT	RF +BASIS	SVM +BASE	SVM +NO-CLO	SVM +ACT	SVM +BASIS	ZeroR	SVM-H/T	DT-PMV
MAE	0.76 (0.15)	0.80 (0.17)	0.75 (0.15)	0.96 (0.20)	0.73 (0.12)*	0.77 (0.12)	0.75 (0.13)	0.91 (0.11)	1.2 (0.0)	1.20 (0.03)	1.10 (0.19)
MSE	1.20 (0.41)	1.31 (0.48)	1.24 (0.45)	1.84 (0.73)	0.98 (0.30)*	1.08 (0.30)	1.07 (0.33)	1.33 (0.28)	2.0 (0.0)	2.00 (0.09)	1.92 (0.78)

Cernadas, Barro, & Amorim, 2014). Three models were chosen for comparison: the first is a ZeroR model that always predicts COM; the second is a decision tree model with the estimated PMV index as the only feature, chosen for testing the efficacy of PMV in predicting individual's thermal comfort as inspired by (Nouvel & Alessi, 2012); the third model is a Gaussian-kernel SVM with humidity and near body air temperature (SVM-H/T) as the only two features. This final model is inspired by Feldmeier and Paradiso's work (2010), which is the only other work that uses wearable sensors for comfort sensing in naturalistic settings. In their work a simpler model, Fisher's discriminant was adopted due to the requirement of monotonicity. This constraint was imposed because the end-goal of their system was to automatically control a HVAC system—monotonicity was required in their model to prevent hysteresis. Removing the monotonicity constraint allows us to employ more sophisticated and presumably more accurate models. We abandoned monotonicity because automatic temperature adjustment is not the only application that could be enabled by comfort sensing. Other research that is more interested in measuring the comfort quality of homes does not require an automatic agent to take actions. Even for applications like intelligent thermostats, mixed-initiative (Horvitz, 1999) approaches might be taken. Such approaches could benefit from a more accurate prediction while remaining tolerant to non-monotonicity.

Since our class labels are ordinal in nature—there is a natural ordering of the classes from uncomfortably warm to uncomfortably cold—we therefore trained our multi-class classifiers on top of a simple ordinal classifier developed by Frank and Hall (Frank & Hall, 2001). This ordinal classifier allows us to transform any multi-class classifier into a classifier for ordinal variables. Note that the class weight adjustment is applied on the 5-level thermal comfort index, rather than the multiple binary classification problems generated by the simple ordinal classifier.

In addition to the prediction models, four different feature sets were investigated: (1) BASE; (2) BASE without clothing level (NO-CLO); (3) BASE with activity information (ACT); and (4) Basis B1 feature set (BASIS). The BASE feature set contains all the major parameters believed to influence a person's thermal comfort. These features are provided by the multi-sensor, Basis B1, and participant self-reports. The NO-CLO contains all the features in BASE except for the self-reported clothing level. This was selected to test if using only sensor data is adequate for comfort prediction. ACT feature set contains all the features in BASE plus activity labels (again, see Table 3). ACT was included as we conjectured that the type of activity a person is conducting might change their expectation about the ideal temperature. Finally, the BASIS feature set was selected to test if using only wearable sensors would be adequate for comfort inference.

Ten-fold cross-validation was performed on the whole dataset ten times, yielding 100 trials of model accuracy. We use Mean Absolute Error (MAE) and Mean Squared Error (MSE) as our evaluation metrics. While MAE is a standard way to evaluate ordinal variable classifiers, we included MSE to further penalize classifiers with large error, as we did not want our classifiers to predict cool while a person is feeling warm. Finally, the Wilcoxon signed-rank test is used to compare model accuracy.

3.6.1.2 Result: SVM with BASE performs the best

Table 4 shows the prediction results for each model. We found that SVM and Random Forest with BASE feature set perform very similarly to each other with respect to MAE, although SVM has significantly smaller MSE. Both of these models outperform the three baselines ($p<0.01$). Compared to the baselines, SVM with BASE reduces the MSE by 51% compared to ZeroR, 49% compared to DT-PMV, and 51% compared to SVM-H/T. SVM-

H/T's prediction result is very close to ZeroR, indicating that having only near body air temperature and humidity is not sufficient for predicting thermal-sensation at home, perhaps because people exhibit adaptive behaviors and conduct various activities. Overall, the relatively small MAE and MSE of our model indicates that it is able to control the prediction error to within one ordinal class distance most of the time—if a person is feeling comfortable, then most of the time the prediction is bounded between slightly uncomfortably warm and slightly uncomfortably cold.

In addition, we also found that adding activity labels (e.g., "cooking," "exercising") directly as features did not improve the model. This could be because there are only a few thermal reports for each type of activity. Furthermore, particular activities may have different effects on different individuals. More advanced graphical modeling might be needed to utilize activity information for improving the models.

Although the SVM model with BASE feature set performs significantly better ($p<0.01$) than the SVM model without clothing level information (NO-CLO), their errors are very close to each other—the MSE for BASE is 0.98 (SD=0.3), whereas for NO-CLO it is 1.08 (SD=0.3). This similar performance may have resulted from relatively low diversity in clothing levels at home, or from participants' inaccurate estimation of their clothing level. We will discuss the latter cause in the second analysis.

Finally, we found that by using only features provided by Basis B1, the MSE is 1.35 times higher than SVM with BASE. It is surprising that even though Basis B1 is able to detect near body air temperature, having no room temperature and humidity information increases the error considerably.

3.6.2 Analysis 2: Challenging Situations

In our second analysis, we further investigated challenging situations for comfort inference. Specifically, we were interested in knowing *the situations and factors behind inaccurate predictions.*

To conduct such an analysis, we selected the best model obtained from our previous study, that is, SVM with the BASE feature set. To generate predictions for all the reports, we conducted two rounds of model training. First, we trained the SVM model by using 50% of the thermal comfort reports (training set), and then performed the prediction on the rest of the reports (testing set). We then trained another model based on the testing set and applied it to generate the predictions for the training set. Table 5 shows the resulting confusion matrix. Note that with only 50% of the data used for training, it is likely that our model would generate more errors than normal (e.g., if we had trained on 90% of the data as in our first Analysis), however, as our goal was to qualitatively analyze the potential situations for inaccurate prediction, we feel that the higher number of error cases is acceptable, albeit slightly less efficient.

To investigate the challenging situations for comfort inference and identify potential solutions, we examined information from participants' comments in their exit interviews, activity information in comfort reports, and raw sensor data.

3.6.2.1 Challenging situations for prediction

We identified 79 error cases where the prediction was more than two classes away from the true label, such as when the model generated UC-Warm while the true label was COM, S-Cold, or UC-Cold. The cells highlighted in Table 5 indicate errors we investigated. Our qualitative analysis of the 79 errors allowed us to identify six situations that lead to inaccurate prediction: (1) *short-term effect or local heat source*; (2) *dynamic transitions*; (3) *extra cover or wind effect*; (4) *lightweight exercising and housekeeping*; (5) *problems with data collection tool*; and *(6) individual differences.*

Short-term effect or local heat source refers to the situation where the indoor temperature is different from the current thermal sensation of the participant. For example, it might be because the participant was drinking a cold or warm beverage, close to a hot stove, or had just taken a shower. Under this situation, a participant's near body air temperature and skin temperature might be within a comfortable zone, while the ambient temperature was low or high. For example, P3 commented *"I felt warmer because I was reading the news and checking email with my laptop on my lap. Even though the room was still cool from earlier, the laptop made me feel warm and kept me comfortable."* He reported COM while the air temperature was 19.8 °C and his skin temperature was only 28 °C, thus making the prediction to be S-Cold.

The second source of the prediction error is a result of *dynamic transitions* between cold and warm situations. We did not have enough thermal reports related to such transitions to train the model effectively for these cases. Furthermore, the sampling rate supported by Basis B1 (1 minute) and our smoothing window (5 minutes) cannot capture quick transitions well. For example, one type of transition occurs when participants enter their warmer homes from colder outdoor environment. Since we conducted the study during late summer/early fall, there were only a few days that the outdoor temperature was low, thus such events are not well-represented in our training dataset. Another type of transition occurs when people move from a warm bed into a colder room. In this situation, participants usually had a high skin temperature from their cozy bed, while the room temperature at the moment captured by the multi-sensor was relatively low. This inconsistency confuses the model. For example, P4 commented that "the room was [at] a

Table 5: Confusion Matrix
(Cells highlighted in gray represent the cases used for our fault analysis.)

True \ Prediction	UC-COLD	S-COLD	COM	S-WARM	UC-WARM
UC-COLD	8	17	0	0	0
S-COLD	7	39	15	8	0
COM	22	186	410	271	10
S-WARM	3	8	17	64	7
UC-WARM	2	1	2	26	9

comfortable temperature" with "waking up" as her activity. The room temperature at the time was only 18.9 °C, while her skin temperature 15 minutes before the report was 31 °C, which is an unusually high skin temperature. Therefore the model predicted she felt UC-Cold, while she actually felt COM.

A third cause of prediction error is when a person might have *extra cover* that was not included in their self-reported clothing level, or might have been affected by un-captured *wind effect*. For example, P11 commented that *"The puppy was in my lap, which warmed me up"* in one case, and noted *"Was still in bed under heavy blankets"* in another report. Additionally, certain locations where participants spent time could have an effect, as an upholstered sofa or carpet can effectively increase a person's clothing level (McCullough, Olesen, & Hong, n.d.). Additionally, our simplified assumption about constant indoor wind velocity occasionally led to prediction problems. For example, P1 commented that she had her fan on while her skin temperature was 33.7 °C and the air temperature was 27.8 °C. The model predicted she felt UC-Warm due to the high temperature while she reported comfortable due to the additional wind effect.

The fourth type of error occurs when an individual is doing *light exercise or housework* during cool days, including activities like moving objects around the house, performing small household repairs, or simply walking around the house. In such cases, an individual might have higher clothing level and slightly higher approximated metabolic equivalent while the other metrics are still low, resulting in a prediction that is cooler than what the individual actually perceives (e.g., predicting COM when the participant reports UC-Warm). Here, it could be that the high clothing level amplifies the effect of the slightly heightened metabolic equivalent. Although the PMV index, a feature included in our machine-learning based model, is supposed to capture such interaction between clothing and metabolism, the misprediction may be generated either because of the inaccurate estimation of the clothing level from the participants, or because of the low skin and air temperatures.

There were *problems with the data collection tool* and data handling. For example, there were several instances that participants did not wear their Basis B1 or had just started to wear it, thus the inaccurate prediction is due to our naive way of handling missing values (i.e., fitting the mean value obtained from all the reports of a participant). In addition, participants may have interpreted the rating scales in different ways at different times. For example, P11 reported her comfort sensation as “slightly uncomfortable” due to the fact that she felt warm. However, she also labeled thermal sensation as “slightly cool” due to the fact that she had just drunk a bottle of cold water. The correct labels to train our model should be “slightly uncomfortable” for comfort sensation and “slightly warm” for thermal sensation, as the “uncomfortable” feeling was caused by the warm temperature. Therefore, the ambiguity in the interpretation of the data collection instructions caused the inaccurate prediction.

Individual differences include the tolerance of heat, cold, sweat and humid environments and the different interpretation of the comfort and thermal sensation scales. For example, P10 reported “comfortable” with comment *“At the desk, my hands were getting cold. I am used to my hands getting cold, though, so it wasn't uncomfortable.”* As his estimated skin temperature was 26.7 °C and the room temperature was 16.5 °C, which is below the typical comfort zone for most people, the model incorrectly predicted UC-Cold. In another case, P1 reported that she felt comfortable after she had exercised at home. The model wrongly predicted her comfort level as UC-Warm as her skin temperature and the air temperature were high—32.6 °C and 28.56 °C respectively—as was her metabolic equivalent. Finally, another source of individual difference is sickness. P11 was sick for more than two weeks during the study, thus her perception and body conditions were different than normal. In one thermal report with mild air temperature at 21 °C, she noted *“[I felt the body] temperature went up because I was feeling sick due to a bad headache.”*

3.7 Discussion

Our work shows that sensing thermal comfort in the wild is promising, but challenges remain. Widely used comfort models and prior techniques that aim at inferring comfort in naturalistic settings are insufficient for inferring individual's thermal comfort at home. In fact, our result shows that these previous techniques—mostly designed for large buildings and offices—did not perform better than a naive ZeroR baseline. The dynamic nature of home activities and people's adaptive behaviors make comfort inference much more difficult in the home than in climate chambers and in offices.

In an effort to infer thermal comfort at home that allows domestic residents to live naturally, we proposed a new technique that uses wearable fitness trackers and in-home sensors to capture various factors that are underexplored in previous in-situ sensing research. By incorporating metabolic equivalent, sweat, and skin temperature sensors, we saw that our technique can reduce prediction error by 50% compared to the baseline models. In addition, our result shows the benefits of having both wearable and in-home sensors. Having only one type of sensor is not sufficient to predict comfort accurately. While wearable sensors are useful to obtain continuous activity-level, sweat and temperature measurements, they are ineffective at capturing the larger thermal comfort picture of the room. Moreover, it would be challenging for the wearable device to understand a transition such as getting out from under a warm comforter in a cold winter while one wakes up. On the other hand, having only in-home sensors is insufficient for capturing discomfort caused by high activity-level and transitions, such as coming back home from exercising.

While we are interested in understanding more specifically which temperature sensors (i.e., skin, near-body and room temperatures) are most critical, our analysis is inconclusive. However, our work is the first step towards understanding the benefit of utilizing these various sensors in reaching better prediction accuracy. We would argue, however, that future work on in-situ sensing should have all three types of sensors, as the combination of the three could potentially help identify various situations. For example, whether users are close to any local heat source could potentially be identified by the combination of near-body and room temperature. Clothing level might potentially be inferred by the difference between skin and near-body temperature, as explored by SPOT (Gao & Keshav, 2013b).

3.7.1 Implications for Improving Thermal Comfort Inference

While there are still some challenging situations for inference, a closer look at the reports reveals that many of these situations could be easily resolved, and some of them could potentially be tolerated depending on the intended application.

First, several of the errors resulted from the way we handle missing data and the way we collect thermal comfort feedback. Note that while we did throw out a report if the user was not wearing the Basis B1 at the time, missing data was filled in for reports that are associated with users just starting to wear Basis B1 (e.g., when they woke up in the morning), as we didn't want to overlook these important moments. When deploying an application based on such a sensing technique, we could easily remove such errors by checking whether the users have been wearing their wearable tracker for a certain period of time. Alternatively, sensor data from previous days that share similar activity patterns could be used to fill in the missing sensor data. For the problem with inconsistent labeling, we could revise the ESM interface to insure the warm-cold direction of the comfort sensation is the same as thermal sensation. For example, we could ask a clarifying question such as, "Do you want the temperature to be cooler or warmer?"

The errors related to rare or quick transitions could potentially be resolved by increasing the sampling rate of wearable sensors, along with better handling of time series data. Currently the Basis B1 has a sampling rate of one time per minute—note that this is the sampling rate accessible by 3rd party developers rather than the true sampling rate of the sensors—which is due to its focus on fitness related applications. On top of this relatively low sampling rate, our five-minute smoothing window makes quick transitions harder to observe. To better identify transitions between cold and warm environments or the presence of additional heat sources, the variance of the sensor reading along with the mean we use in this study could be used. In addition, activity recognition techniques could be applied to identify some transitions such as home/away status (Chen & Khalil, 2011).

Furthermore, if comfort sensing techniques such as ours become useful, it is possible that wearable sensor manufacturers would broaden their services to include comfort sensing, thus tailoring their device to serve such a purpose—e.g., by allowing 3rd party developers to access sensor data in real-time. Finally, a larger deployment with more data to train the models might help to model relatively rare transitions.

Two other challenges we identified include extra "clothing" and wind effects that were not captured by the sensors or by users' self-reports. Through additional insights gained from the exit interviews, we found that whether or not people have extra covering is sometimes related to their locations in the room. For example, P11 usually had her blanket on when she sat on the sofa in her TV room. This suggests it might be possible to infer people's extra clothing level via part-of-the-room indoor positioning, although it would require additional training data from each of the individuals. However, we also found that when such extra clothing exists, the difference between the near body temperature detected by Basis B1 and the air temperature detected by the multi-sensor is larger. We could potentially use this information to help us identify if extra clothing exists.

Finally, individual differences might be resolved by taking a personalization approach. However, the standard ways of training personalized models—such as generating a model for each individual based on one's own data—may not be feasible for our approach since the number of reports required to cover all five thermal comfort labels at a sufficient level may be excessive. While such isolated models may not work for our approach, "Groupization" could be a better solution as it seeks to build personalized models for an individual by using data from other similar people. This approach has been used in personalized information retrieval (Vu, Song, Willis, Tran, & Li, 2014) and activity recognition (Lane et al., 2014) systems, and is particularly suitable for our application scenario where the training data provided by each individual is minimal. Groupization reduces the amount of feedback an individual needs to provide in order to train the model, while improving upon the prediction accuracy based on a population-based model.

3.7.2 Limitations

There are a few limitations of our study. First, study was conducted at the end of the summer/beginning of fall when the temperature was relatively mild. Further deployment is needed to inspect the feasibility of our approach in more extreme weather, as seasonal difference might be an important factor to include in the model. In addition to the relatively short deployment conducted in the summer, our thermal comfort reports are only provided by 9 participants, a small population. More research is needed to consider the role of individual difference and to validate if this approach could work for a larger group of people.

Although extensive collection of thermal comfort feedback from each individual is required in our experimental system, we envision the future system in which online inference could rely on a small group of users who provide feedback for training a population-based model. If derivation from the population-based model is found, the system could then prompt a user for more comfort feedback in order to create a personalized model or reassign them to a more appropriate group.

3.8 Conclusion

In this chapter, we present a new technique for inferring people's thermal comfort at home under naturalistic settings using in-home sensors as well as off-the-shelf wearable devices equipped with sensors that allow estimations of metabolic equivalent, sweat and skin temperature. A sensor deployment and experience sampling study was conducted in 7 households with 11 participants to validate the potential of such approach. Our study results reveal the advantages of this approach, challenging situations for prediction, and potential directions for improve in-situ comfort sensing at home.

Chapter 4. Reef: Exploring the Design Opportunity of Comfort-aware Eco-coaching Thermostats[21]

4.1 Introduction

In the U.S., the domestic sector consumes more than 20% of the total energy produced, and half of this is consumed by heating and cooling usage (*Total Energy*, n.d.). The high level of usage can be reduced. Researchers have pointed out that consumption is highly dependent on occupants' behaviors (Sonderegger, 1978), as well as the inefficient use of existing manual and programmable thermostats (Meier, 2012; Yang et al., 2014).

Over the past decade, numerous intelligent thermostats have thus been proposed to mediate temperature control, helping people to save energy while maintaining comfort (e.g., Gupta et al., 2009; Lu et al., 2010; M. Mozer, Vidmar, & Dodier, 1996; Scott et al., 2011). Such thermostats may have occupancy-responsive control (e.g., Koehler et al., 2013), learning capability (e.g., Alan et al., 2016; *The Nest Thermostat*, n.d.), and/or eco-coaching features (e.g., Pisharoty et al., 2015). However, there are still two problems with these thermostats that limit the energy savings that can be achieved.

First, the majority of intelligent thermostats assume occupants have a fixed temperature preference at home while in reality this can be quite dynamic, changing depending on occupants' activities and other contextual factors (J. Fergus Nicol & Humphreys, 2002).

---

[21] An earlier version of this work was published at the DIS 2017 conference.

While researchers have recently explored comfort-aware thermostats—thermostats that can react to people's changing preferences by predicting their comfort based on sensed conditions (e.g., Feldmeier & Paradiso, 2010)—such thermostats have been principally studied in office settings, and are not suitable for the home due to the sensors used and low accuracy (Huang, Yang, & Newman, 2015).

Second, a fundamental assumption underpinning most smart thermostat development is the view that occupants are primarily receivers of comfort (i.e., comfort is given by the heating and cooling system), rather than active agents that can utilize other means, such as changing clothes, to maintain their comfort. Researchers like Clear et al. (2013) have therefore proposed to investigate approaches based on *adaptive thermal comfort* (Humphreys et al., 2007), an approach that emphasizes occupants' agency in performing adaptive behaviors to achieve comfort, and challenges the notion that comfort and controlled indoor temperatures exist in a fixed relationship.

Our work in this chapter proceeds from the observation that a comfort-aware approach can be combined with *eco-coaching* (Pisharoty et al., 2015; Yang et al., 2016) to provide additional opportunities for reducing heating- and cooling-related energy waste. In the eco-coaching approach, an intelligent system models user behavior and produces recommendations for energy-saving actions that can be taken by users. Recent work has shown that eco-coaching can save 5-12% of energy expended for cooling when recommending thermostat control schedules (Pisharoty et al., 2015), and that users find eco-coaching to ease the selection and execution of more energy-efficient actions (Yang et al., 2016). Previous work, however, only examined recommendations based on occupancy schedules—the opportunity for gaining efficiency by optimizing temperature setpoints was not explored. By including knowledge of users' comfort preferences and recommendations derived from the adaptive thermal comfort model, additional savings can be realized.

Thus in this chapter, we report our initial steps in exploring the design space of *comfort-aware eco-coaching thermostats*, i.e. thermostats that seek to synthesize comfort-awareness, adaptive thermal comfort, and the eco-coaching approach.

Our work makes two contributions: 1) We illustrate the design space of comfort-aware eco-coaching thermostats by systematically mapping the design space and producing fifteen diverse prototypes representing key points in the space; 2) By further carrying out User Enactments (Davidoff, Lee, Dey, & Zimmerman, 2007; Odom et al., 2012, 2014) wherein we engaged 11 thermostat users in scenarios involving the prototypes we had developed, we were able to uncover underlying values and tensions that are likely to drive user reactions to different design directions if and when they are encountered in the wild. Our work provides a critical first step towards the realization of comfort-aware eco-coaching thermostats, and provides valuable insights for future system development and field deployments.

4.2 Related Work

Many researchers have developed smart thermostats to help reduce energy consumption and increase comfort (Beltran et al., 2013; M. Feldmeier & Paradiso, 2010; Gupta et al., 2009; Koehler et al., 2013; Lu et al., 2010; M. Mozer et al., 1996; Pisharoty et al., 2015; Scott et al., 2011). These smart systems have achieved their goals in various ways, including the ability to: react to residents' occupancy status (e.g., Scott et al., 2011), adapt to building characteristics or user preference (e.g., *The Nest Thermostat*, n.d.); and respond to people's thermal comfort (e.g., Feldmeier & Paradiso, 2010).

Occupancy-based thermostats (Gupta et al., 2009; Koehler et al., 2013; Lu et al., 2010; Scott et al., 2011) focus on reducing the heating or cooling time when people are away from home. Most of these systems make a simplistic assumption that people are satisfied with a fixed temperature when they are at home. While occupancy-based thermostats have been shown to be successful in reducing energy usage, this simplistic assumption misses a chance for both further energy savings and increased comfort, as in reality people's temperature preference at home is dynamic according to their activities and other context (Halawa & van Hoof, 2012).

Preference-based thermostats may be able to cope with people's dynamic preferences, for example, by learning the desired setpoints at different times of day (e.g., *The Nest Thermostat*, n.d.) or at different pricing conditions (e.g., Alan et al., 2016). However, users still struggle to understand the smart features of such thermostats, leading to suboptimal use or abandonment (Yang et al., 2014). In addition, current preference-based thermostats only scratch the surface of how smart thermostats can interact with their users. While this work largely focuses on thermostats that learn people's preferences passively, more recent work on *eco-coaching* (Pisharoty et al., 2015; Yang et al., 2016) finds that thermostats that provide actionable recommendations based on their learning capability are more promising in reducing energy consumption.

Comfort-aware thermostats, although a less explored alternative, provide the potential of reacting to people's changing temperature preferences. Such thermostats infer occupants' thermal sensation in real-time using wearable and indoor sensors and predictive models to map environmental conditions and user activity onto inferred thermal sensation (M. Feldmeier & Paradiso, 2010). While this approach seems promising (Huang et al., 2015), the imperfection of comfort inference means that full automation is infeasible. Researchers have proposed employing mixed-initiative design in creating smart thermostats (Pisharoty et al., 2015; Yang et al., 2014), emphasizing the collaboration between machines and humans to reach a shared goal. *Eco-coaching* (Pisharoty et al., 2015; Yang et al., 2016), employs a mixed-initiative approach wherein the system offers suggestions for actions to take and users are held responsible for finalizing the decisions. Although this approach is helpful for dealing with misprediction, it has not yet been applied to design *comfort-aware thermostats*, as previous implementations were constrained to suggesting timing for setbacks (energy-efficient *away* and *sleep* temperatures) rather than alternative temperature settings for times when the home is occupied.

While researchers have demonstrated that occupancy-based and comfort-aware thermostats can reduce energy consumption by 7-57% (e.g., Feldmeier & Paradiso, 2010; Scott et al., 2011), the fundamental design philosophy that underpins such thermostats has been criticized on the basis that it limits the potential to reach more sustainable ways of living (A. Clear et al., 2014; A. K. Clear et al., 2013; Humphreys et al., 2007). Past work on smart thermostats assume *comfort as a product*, something that is delivered to us by the indoor environment. However, recent research on adaptive thermal comfort has proposed to reconsider *comfort as a goal*, a view that emphasizes the role of human agency: occupants can leverage different adaptive behaviors such as adjusting their clothing level to reach comfort in addition to cranking their thermostats up or down. In alignment with this view, Clear et al. (2013) have proposed to create systems that facilitate adaptive behaviors and temperature variations, yet limited work has been done in exploring this design opportunity.

In this work, we therefore explore comfort-aware thermostats' potential in reacting to occupants' dynamic temperature preferences while accommodating mispredictions by leveraging eco-coaching. We also broaden the concept of eco-coaching to include the view of *comfort as a goal* rather than just *a product.* Therefore, rather than only suggesting to users how to control their temperature setpoints, we explore thermostats that might also encourage adaptive behaviors such as changing clothing or drinking warm beverages. We call such thermostats *comfort-aware eco-coaching thermostats*.

4.3 Method & Design

The aim of our study was to advance understanding of how to design comfort-aware eco-coaching thermostats. While others have explored technologies that seek to challenge or overturn people's expectation or values around thermal comfort (e.g., A. Clear et al., 2014), we sought to explore more subtle approaches that work within the existing constraints of people's cultural expectations of mechanically-mediated thermal comfort. We thus pursued designs that encourage more efficient behaviors through being adaptive while respecting people's expectation on thermal comfort.

To illuminate the design space and understand how people would react to possible design approaches, we chose user enactments as our method (Davidoff et al., 2007; Odom et al., 2014). User enactments allow researchers to rapidly explore how potential users' values, expectations, and social identities inform their reactions to different design possibilities. Through simulating real world scenarios involving potential technological futures, researchers gain insights into how users might react to technology designed with different attributes, as well as the values users bring to bear when assessing design alternatives. While prior thermostat work has employed field deployments to validate the effectiveness of various designs (e.g., Alan et al., 2016; Pisharoty et al., 2015), our designs require significantly more complex technology (e.g., personalized comfort prediction and room-level localization) which makes deployment a less reasonable action to take as a first step. Applying user enactments in understanding users' values is essential to ensure that costly system development and field deployment efforts are well-grounded and more likely to succeed.

Table 6: The key design dimensions and the corresponding prototypes.
(NA: Norm-Activation, RT: Rational Thinking.)

Eco-Coaching Style	Persuasive Strategy	Timing of Interaction				
		Real-time Control	Short Engagement	Plan.	Mis-pred.	Refl.
Informative	NA	D1				
	RT			D7		D13
Advisory	NA		D4	D8	D10	
	RT	D2				D15
Proactive	NA					
	RT		D5	D9	D11	
Adaptive	NA					D14
	RT	D3	D6		D12	

Our study involved three main steps. First, through synthesizing prior literature and conducting multiple rounds of brainstorming, affinity diagramming, and expert review, we delineated the design space of comfort-aware eco-coaching thermostats. We distilled three key design dimensions (Table 6) that guided our later designs and summarized the technology constraints of state-of-the-art smart thermostats and comfort prediction approaches. Second, we operationalized our concept through iterative prototyping. We generated numerous ideas and low fidelity prototypes before creating fifteen design probes (D) in the form of high fidelity interactive prototypes (Figure 3). Third, we conducted user enactments to qualitatively probe our target users' opinions: how they perceive the different design concepts and how these designs might encourage or impede users from reaching higher energy savings.

4.3.1 Reef: Thermostat Designs

We first identified the key technology constraints of state-of-the-art smart thermostats and comfort prediction, focusing on constraints that are likely to persist for the foreseeable future. We used such constraints to guide the design of Reef, a hypothetical thermostat that can predict people's comfort, react, and encourage energy savings. First, we assumed Reef employs a sensing approach similar to Huang et al. (2015), namely, it relies on wearable devices that detect activity level and near body temperature (e.g., *Basis B1*, n.d.), as well as in-home sensors that capture humidity and temperature (e.g., "AeoTec MultiSensor," n.d.). Second, we assumed that Reef is able to detect and predict people's house occupancy status, room-level location and sleep status, which have been demonstrated feasible in prior research (Lymberopoulos et al., 2015; Pisharoty et al., 2015). Therefore, Reef can use its users' status to determine whether to trigger an Away, Asleep or Awake mode (i.e., users are at home but not sleeping). Third, Reef can learn personalized comfort preferences by soliciting feedback from users, ultimately generating comfort predictions on a five-level scale ranging from uncomfortably cold to uncomfortably warm (Huang et al., 2015). Fourth, due to limited sensing and inference capabilities, Reef will sometimes make mispredictions (e.g., predicting occupants are comfortable while they are slightly cold) (Huang et al., 2015). Finally, as the smartphone-based control has emerged as a common approach for interacting with smart thermostats (e.g., *The Nest Thermostat*, n.d.), we expected users to interact with Reef through a smartphone application. Our interfaces were therefore all designed for mobile screens.

Through prior literature, we identified three key design dimensions for smart thermostats (see Table 6). These dimensions facilitated the systematic generation of fifteen distinct interface prototypes. The three design dimensions we chose include *eco-coaching style*, *persuasive strategy*, and *timing of interaction*. Here we will describe these three design dimensions in detail as well as their relationship with the various prototypes we created.

Eco-coaching style refers to the approach Reef takes to communicate with users. At one end of the spectrum, Reef seeks to be *informative* by showing useful eco-feedback to help decision-making, letting users remain in control (Froehlich et al., 2010; He, Greenberg, & Huang, 2010; Pousman, Stasko, & Mateas, 2007). This information might include comfort level prediction, estimated financial savings, and environmental impact. While respecting users' agency was one reason that we explored such a hands-off approach, this approach also handles inaccuracy. Due to the inevitability of imperfect prediction, it may not be most favorable for Reef to automatically change the temperature according to its comfort prediction (Fanger, 1970). One prototype that implements such an approach is D1, which shows the predicted comfort of different household members at home in four colors. Similarly, D9 shows users the predicted comfort level in different rooms at different settings while requiring them to manually create their temperature schedules. At the other end of the spectrum, Reef can be *proactive* (Davidoff et al., 2007; Rodden, Fischer, Pantidi, Bachour, & Moran, 2013), making decisions for users and only informing them about the benefits. In some situations users might feel that the benefits of automation outweigh the cost of minor mispredictions. D5 and D7 are two prototypes that implement the proactive approach. In D5, Reef identifies that the user has been comfortable for a few days and therefore automatically lowers the setpoint for an evening. It only informs the user about the decision and the benefits without first asking permission. Similarly, in D7, Reef automatically generates and activates an energy-saving schedule after observing occupants' behaviors and comfort level for a month.

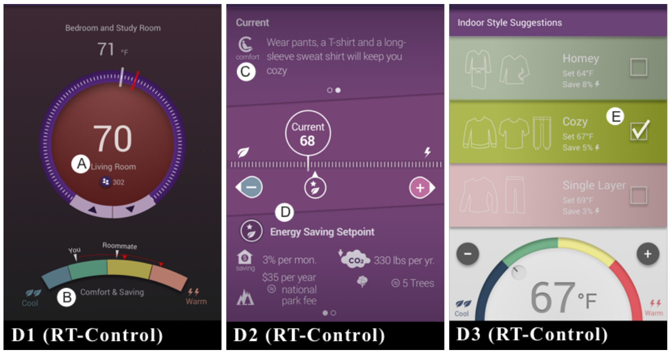
A third approach lies in the middle of these extremes, offering suggestions to users and letting them make the final decision (Pisharoty et al., 2015; Yang et al., 2016). Here we probed two possible design directions: one is to offer suggestions directly related to temperature settings (*Advisory*), the other is to encourage adaptive behaviors *(Adaptive)* (A. K. Clear et al., 2013). For the former type, two different saving suggestions were implemented in Reef's prototypes, namely, personalized saving modes and short-term variation. D2, D4 and D8 are examples that implement such strategies. Inspired by ThermoCoach's (Pisharoty et al., 2015) saving suggestion design, Reef offers four similar modes: high energy saving, energy saving, regular and warm. Each mode has a corresponding temperature when people are at home and awake (i.e., the Awake setpoint). This temperature is determined by Reef's learning of occupants' comfort. In winter, for example, when running in regular mode Reef will pick a temperature that the user mostly feels comfortable with, and in energy saving mode Reef will find an Awake setpoint that is likely to feel slightly cold. The second strategy is short-term variation, meaning that Reef

Figure 2: Participants reacted to our prototypes in five different scenarios. From left to right: real-time control; short engagement; planning; misprediction and reflection. Note that each UE took place in a different part of the home.

might suggest users to lower their setpoints for only a short period of time (e.g., a night or a day). The rationale is that while people care about their own comfort, they might be willing to lower their expectation shortly for saving energy. Whereas in some prototypes Reef suggests users to directly change thermostat settings, in others Reef encourages *adaptive* behaviors. One of the adaptive behavior Reef encourages is to change indoor clothing style. The suggestions are made in different ways. For example, in D2 Reef shows the appropriate clothing to wear at different setpoints; In D3 Reef makes the suggestion more salient—users can pick a clothing option and Reef will adjust the setpoint to a corresponding temperature. Besides clothing level, we also explored other adaptive behaviors suggestions such as drinking a hot beverage (D6), walking at home for a few minutes (D12), lowering the setpoint in the morning when users are leaving soon for work (D13), and delaying the time to heat up one's home after arrival (D15).

Persuasive Strategy refers to the approaches used to promote energy-saving behaviors (Dillahunt & Mankoff, 2014; Fogg, 2002; Froehlich et al., 2010; He et al., 2010; Pousman et al., 2007). Froehlich et al. (2010) highlighted two major strategies: *Rational Thinking (RT)* and *Norm-Activation (NA)*: While some people might respond better to analytical insights, others might be more easily persuaded by emphasizing cultural norms and leveraging social influence.

In our study, we probed different persuasive strategies to encourage saving. Besides information about financial savings, we explored the use of environmental impact framed under a 2015 U.S. government's policy to reduce CO_2 emission by 17% (US EPA, n.d.) (e.g., D2 and D4). In other prototypes, we showed users about friends that use a similar saving mode (e.g., "Mark and 33 others also use this mode" in D8), as well as adaptive behavior tips shared by households similar to the user (D14). These are also two examples demonstrating how we combined the eco-coaching style dimension with persuasive strategy to create the final prototypes.



Ⓐ Show the number of Reef users using this setpoint to provoke reaction about norms.

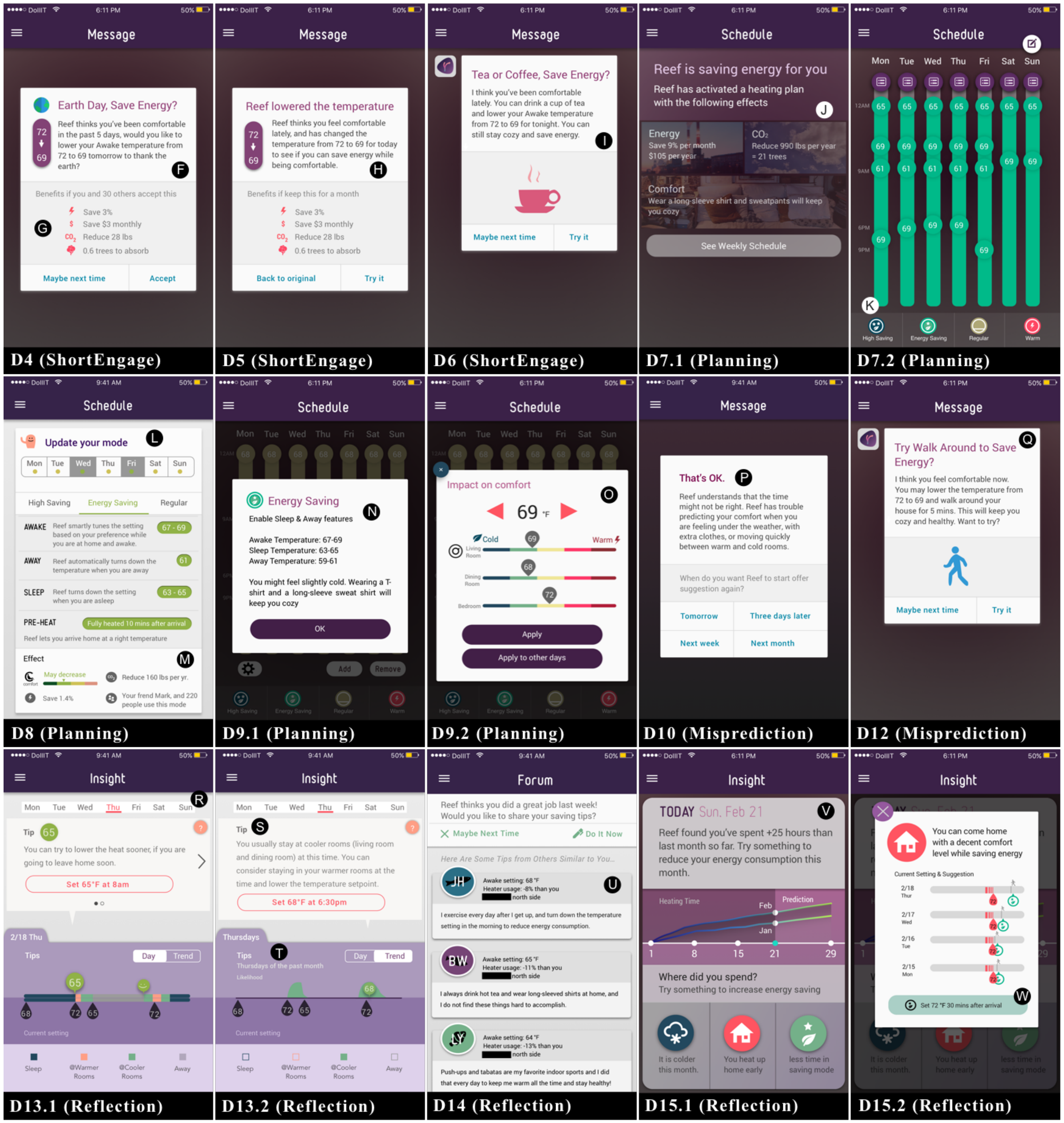
Ⓑ Show the predicted current and future comfort levels for both residents in white and red.

Ⓒ Suggest the clothing needed if the temperature is set at 68°F.

Ⓓ Suggest a personalized energy saving setpoint, 70°F, with benefits as well as recommended clothing in the next card.

Ⓔ Provide three options of clothing with the corresponding setpoints and benefits. Users can click on one option to change the setpoint.

Figure 3: Selected interfaces for our 15 prototypes (see next page for more interfaces). We skip D11 as it looks the same as D5 but was tested in a different scenario.



Ⓕ Track comfort history and only recommend short-term adjustments when users have been feeling comfortable for five days.

Ⓖ Show the benefits if the user and 30 others accept this recommendation.

Ⓗ Automatically adjust the setpoint for its users if they've been comfortable lately.

Ⓘ Suggest users to drink a cup of tea and lower the setpoint briefly.

Ⓙ Automatically activate a personalized energy saving plan without asking users' permission. It shows the benefits and the clothing needed for this plan.

Ⓚ Users can easily switch the thermostat schedule by choosing different plans.

Ⓛ Users can assign different days to different energy saving plans.

Ⓜ Show the benefits as well as friends of the users who use this setting.

Ⓝ Explain how to be rated as "energy saving". Users need to manually adjust the setpoints themselves.

Ⓞ Users need to manually create their own schedule. They can view the effect of different setpoints on comfort and the temperature differences in different rooms.

Ⓟ Reef first sends a notification to recommend users to lower their setpoints. If they don't accept, Reef reveals it's sensing limitations.

Ⓠ Recommend an immediate activity (walk around) to warm themselves up and lower the setpoint.

Ⓡ Show data-driven suggestions. For example, Reef recommends that users could lower their setpoint in the morning since they usually leave home soon after they wake up.

Ⓢ Offer other adaptive thermal comfort tips (e.g., change to stay in a warmer room.)

Ⓣ Show data to support the tips, including people's sleep pattern and when they are usually at a warmer or cooler room.

Ⓤ Show adaptive behavior suggestions shared by people living in the similar area.

Ⓥ Show current usage before the month ends, to give opportunities for changes.

Ⓦ Provide reasons for increased heating usage and actionable items (e.g., encouraging users to only fully heat up their home 30 minutes after arrival).

Timing of Interaction refers to the various situations in which users might interact with thermostats. We chose to explore five situations and created one user enactment scenario for each (as shown in Figure 2), including *real-time control*, *short engagement*, *planning*, *misprediction* and *reflection*. Although timing of interaction is slightly different from other design dimensions—in the sense that the options in this dimension are not mutually exclusive and can be supported by one thermostat—this dimension helps us understand if a particular approach that works well in one situation also works well in another. Here we briefly explain the first three situations while discussing *misprediction* and *reflection* in more depth. Note that all of our scenarios were created for winter settings.

Real-time control (UE-RT) refers to the situation when one feels uncomfortable and wants to change the setpoint. In our scenario, a participant encounters this situation when watching TV in the living room. He feels cold and opens up Reef to see if he can make any adjustment. *Short engagement* (UE-ShortEngage) refers to the situation that a notification is sent from the thermostat to the user to encourage energy saving. In our scenario, a participant faces this situation when reading books in their study room. Reef recalls that she has been comfortable for a few days thus sends a message to encourage saving. *Planning* (UE-Planning) refers to the situation when one is thinking about next week's plan and checking the thermostat schedule. In our scenario, a participant does this on Saturday night before going to bed or another self-selected time that the participant usually plans his next week's schedule.

Misprediction (UE-Mispred) refers to the situation that the thermostat inaccurately predicts an occupant's comfort and makes an inappropriate decision or suggestion. In our scenario, while the participant feels cold and sick, Reef inaccurately predicts she is comfortable, thus suggesting her to lower the setpoint. From previous literature (Huang et al., 2015) we already knew that comfort prediction is not perfect, especially when people are sick, wear extra clothes, or are affected by other local factors that are challenging to track by sensors (Huang et al., 2015). We were curious to explore how to work around technology constraints regarding comfort prediction. For example, a thermostat might apologize

(Srinivasan & Takayama, 2016) and inform its users about its limitations when it makes mistakes (i.e., a form of incidental intelligibility suggested by Yang et al. [2013]). We therefore created a prototype, D10, to present an implementation of incidental intelligibility and compared it with other designs (D11, D12) that do nothing when mispredictions happen.

Reflection (UE-Reflect) refers to the situation when users are thinking about their long-term usage of the thermostats, reflecting on their practices, and reassessing actions they took. We designed a scenario where the participant has just finished checking their energy bill and becomes curious about the heater usage. We explored different ways to support reflection. This includes strategies to support data-driven reassessment proposed by Yang et al. (2016). For example, in D13, Reef allows reassessment by tracking people's in-home behaviors and their schedules. It shows a visualization depicting when the user is likely to go to bed based on behavior tracking. In some prototypes, Reef supports reassessment by considering other real-world factors. For example, in D15 Reef explains how colder temperatures cause additional heater usage, potentially helping users understand why their actions may not lead to the expected outcomes. In addition to supporting reflection by data-driven reassessment, we also explored the inclusion of a discussion board, D14, which helps users share and learn from peers about different ways to maintain their comfort while not turning their thermostat up.

4.3.2 User Enactments Study

Our interactions with study participants involved three steps: an initial interview, a diary study, and the user enactment sessions. The purpose of the initial interview was to understand how participants currently used their thermostats. This interview also shed light on how participants chose their default setpoints, their attitudes toward climate change and sustainability, and how they viewed the relationship between comfort and energy-saving. We then conducted a diary study lasting three to seven days. The purpose of this study was to raise participants' awareness of their own temperature preferences, facilitating better feedback during the user enactments. We installed an experience sampling app, PACO ("Paco," n.d.), onto participants' phones and gave them two thermometers to place in their homes. Participants were asked to report their comfort and the corresponding indoor temperature three times a day. On average each participant created 13 reports (min: 7, max: 24).

After the diary study we then conducted the user enactments in our two-story smart home testbed (a.k.a. the first author's home). Inspired by Rodden et al.'s (2013) approach to ground participants, we presented a series of storyboards to 1) illustrate the current energy problem, 2) envision the future and introduce Reef, and 3) offer the context of enactment.

We first introduced the problem—the high energy usage of heating systems at home and the variation in consumption caused by different usage practices (Sonderegger, 1978). We showed each participant an estimate of their potential financial savings using the average household size in our region of study. By factoring in the energy sources, we also illustrated the potential environmental impact a participant could make in terms of CO_2 reduction. Finally, to offer a frame of reference for a possible environmental impact goal, we introduced the policy announced by the U.S. government in 2015 (US EPA, n.d.), which is to reduce 17% of CO_2 emissions by 2020. We made a simplistic assumption that the domestic sector should contribute equally by reducing 17% of energy use.

Afterward, we asked participants to envision the year 2020, four years in the future from the time of the study, and a new thermostat product, "Reef," has been released. We described the key characteristics of Reef, namely its use of wearable and in-home sensors for predicting people's comfort; the ways it learns from its users; its capability to predict whether residents are sleeping or away from home; and its support for user-defined Away, Awake and Asleep temperatures.

We asked participants to imagine living in a smart home in February 2020, and gave them a simulated calendar showing what their daily schedules might look like at the time. They were also provided with a simulated weather app interface showing the "current" week's weather ranging from 21 °F to 55 °F. We also asked them to imagine living with another person in the house, either a partner or a roommate. After showing them this information, we asked them to walk through the scenarios we designed.

As described earlier, we developed five scenarios and created three prototypes for each scenario. To ensure the session time was bounded in two hours, we let each participant experience a subset of 3-4 scenarios (thus each design was experienced by 7-8 participants and each participant experienced 9-12 designs). We showed one prototype at a time in a scenario, and we repeated the subset of scenarios three times to show all the design variations. In each scenario we first briefly explained the interface of the prototype and asked participants to interact with it. To increase realism, we tried to match the indoor temperature of our smarthome testbed with what was depicted on our prototype thermostat displays (within 1°F difference). At the end of the session, we then displayed all the 9-12 selected prototypes together for participants to compare. All of the user enactment sessions were video recorded for later analysis by the study team.

4.3.3 Participants & Data Analysis

We conducted our study with 11 participants. We recruited participants through emails, online forums, and social networking sites. Participants were compensated with $50 for completing the interview, diary and UE study. Six males and five females participated in the study, representing a range of occupations including teacher, midwife, finance manager, school administrator, software engineer, among others. Most of them were between ages of 26-35 except for U6 (36-45), U7 (>55) and U11 (46-55). Only U8 owned a smart thermostat with the rest using either manual or programmable thermostats. The study was conducted from the end of April to the beginning of May 2016 where the average temperature in the region of the study was 52 °F (max: 79 °F, min: 28 °F). Thus people were still using their heaters at the time, though the outdoor temperature was somewhat warmer than the simulated temperatures depicted in the study.

To analyze the data we conducted a debriefing session for each enactment within 48 hours. During the debriefing, three of the authors reviewed the whole video and discussed emerging patterns and questions to probe during later sessions. After all the UE sessions, we transcribed all the videos and conducted a thematic analysis (Braun & Clarke, 2006) to identify common patterns in participants' reactions and understand the underlying expectations and values.

4.4 Findings

In the following section, we present three major themes that emerged in our thematic analysis: the desire for comfort and its relationship with sustainability; the desire for control and its tension with the desire for convenience (e.g., the desire to have the system make decisions); and the importance of careful allocation of agency while being pertinent.

4.4.1 Comfort & Sustainability

In this section we present findings related to short-term variation, adaptive suggestions, comfort visualization, and personalized saving modes which helped us uncover participants' values regarding comfort and sustainability.

Our design probe (D4) that encouraged short-term variation based on prior comfort history received polarized responses. Before user enactments, we expected participants to be fairly open to this suggestion because the suggestion was only triggered when they had been "comfortable" for five days. However, some participants felt that being comfortable now didn't imply they should compromise their comfort in the future. According to U1, *"That does not make any sense. If I feel comfortable in the past five days, then what I should do is just keep the same temperature, so that I can keep feeling comfortable."* Others were fine with compromising their comfort for a limited time. U2 said, *"That's fine [for me to live less comfortably for a day]. That's reasonable and just for tomorrow. I think it's like, five days I use temperature I like, and one day I lower it. It sounds OK. … Because I am not constantly sacrificing my comfort level, so it's just temporary."* We suspect this discrepancy is due to differences around how our participants valued comfort. Although they all desired to be comfortable, some perhaps viewed comfort more as a necessity while some others viewed it as something more like a luxury. Participants who viewed comfort as more of a luxury might be more willing to compromise their comfort, as long as it's temporary and as long as their comfort needs had been reasonably satisfied.

In addition to short-term variation, some of our adaptive suggestions also shed light on the dynamic nature of comfort. Participants had different expectations of comfort at certain times of the day, such as in the morning when they are leaving home soon (e.g., it would be OK to lower the setpoint in the morning slightly earlier) or when they come back home (e.g., delaying the time to fully heat up home upon arrival.). U3: *"I don't mind waking up a little chilly"* (D13); U9: *"Well as long as it's warmer than outside. [In February] it's going to feel better than what we just came from. I think that would be a great solution that I heat up halfway and heat up rest of the way [after] I get home."*(D15) This suggests that although they valued comfort, their perception of comfort varied within a day. Thermostats designed to fit this dynamic expectation may increase the chance for savings.

Most participants had a strongly positive reaction to the design (D1) that allowed them to view the comfort level of their roommate or partner. Three participants who currently lived with their partners specifically said they were willing to sacrifice their own comfort to keep their partner comfortable. U9: *"I would probably just grab a blanket and leave how it is … I think a lot of changes in the thermostat is the regulation between my wife and myself and the baby honestly. But if I could see that me trying to get more comfortable would [de]crease hers that much … then I wouldn't do it."* These participants valued some household members' comfort more than their own.

In UE-Planning, participants were shown a design (D8) that suggests three different energy-saving modes (e.g., "energy saving" and "regular") that were personalized according to inferred comfort preferences. While we expected that such saving modes would encourage participants to explore a slightly more energy efficient setting, three of our participants raised a similar concern with this design. U6 explained: "*There [is] something interesting to me about describing the level as high saving, energy saving, very comfortable, … Because being very comfortable, is like a qualitative statement about my own personal experience of the temperature, these two levels describing high saving and energy* *saving feels distant from me. It shifts the priority from my comfort to external energy savings and so even though I'm motivated by it ... it seems like to put [me] in a conflict: saving energy or being comfortable.*" Whereas these participants valued both comfort and energy saving, and thought they could achieve both—*"I can be quite comfortable in the colder [mode, i.e.] in the energy saving to high saving [mode] with proper [clothing], like having a blanket, that to me is comfortable, I'm more comfortable in cooler environment anyway"* (U6). It seems that rendering energy saving and comfort in a mutually exclusive relationship created a conflict where none need exist.

4.4.2 Control & Convenience

A major question explored in our study is what types of eco-coaching styles better fit participants' values regarding control and convenience. We were especially interested in the intersection of eco-coaching style and the timing of the interaction.

As we expected based on prior research, participants valued user control and favored a more advisory and informative (as opposed to proactive) approach when presented with short-term saving opportunities in UE-ShortEngage. However, in the UE-Planning scenario, the majority of them preferred a more proactive Reef, even though in this prototype (D7) Reef activates an energy-saving schedule without asking participants' permission. U6: *"I don't mind [Reef acting on my behalf] …I think it's why people engage with smart devices in general. I think that's part of the payoff. [That] is, you have this intelligent device using the data to make data-driven decisions, but does not maximize unless it's making the decision. So I want it to go ahead and use the data that is collected and I would have the real world experience of feeling it. And so if I don't agree … [like] 'oh gosh that was way too far to being cold' then I know I can adjust it."* This indicates that participants' values regarding comfort and convenience may shift based on interaction context. This openness to Reef's proactivity was also contingent on their trust in the system's capability to capture behavioral information, the value placed on convenience, and the ease of control.

Prototypes (e.g., D8) featuring short-term variation also helped us uncover insights relating to the differing weighting of control and convenience among our participants. Our participants' reactions toward short-term variation were polarized. Some favored consistency rather than variation. U7: "*I guess that is [like] my eating philosophy, I have two days when I eat less and the other day I get [what] I want so. I can see that. But it seems like for the temperature I would rather find the lowest temperature during [a] day that are still comfortable. Try maximize that [aspect] as opposed to being uncomfortable on days."* This could be related to valuing convenience—they didn't want to fiddle with the thermostat—, or could be related to the challenges of planning ahead. U6: *"I don't think I would pick two days a week [to be more energy-saving] … because I wasn't sure what will make these two days [more] special than the rest of the days."*

However, some participants thought it was possible to set some days to be more energy-saving than usual. U5: *"I would consider [lowering the setpoint for a day every 6 days]."* Interestingly, this willingness to compromise comfort and convenience is also fluid and negotiable depending on different conditions. U5: *"I think it also depends on, again, like, the temperature outside. If it's going to be significantly colder, and then it asks me to lower the temperature tomorrow, I [would] probably say no. ... If I perceive it's going to be colder, even though the inside temperature theoretically should stay the same… I don't want be colder*".

Whereas the prototypes mentioned above were probed under the situations where Reef made accurate predictions, we also explored participants' attitudes and values when mispredictions happened. We expected our "incidental intelligibility" (Yang & Newman, 2013) design (D10, UE-Mispred) to receive positive feedback from participants. In this enactment, we told participants to imagine that Reef had misjudged their comfort, and suggested lowering their setpoints while they were sick and cold. If they chose not to, Reef then showed a sample interface attempting to explain why it may have made that mistake. However, almost all the participants expressed that they didn't even want to look at Reef's explanation when they were sick and when Reef made a misprediction. U1, for example completely dismissed the explanation, saying that, *"I don't want to debug it."* Participants' first reactions were to just increase the temperature to a comfortable point—they didn't necessarily want to fix Reef. This finding suggests that for some people, understanding the system models and capabilities is not as highly valued as we expected based on prior work (e.g., Yang & Newman, 2013). To them, convenience and error recovery are more important.

However, some participants found that Reef's transparency made it more considerate. U5: "I like this one better [D10] because it's more honest…, it makes me feel like it does understand like you are a human, you are doing different things, different things might affect you, and we are an app, and we might not be able to get that all the time... It feels kinder, doesn't feel quite as harsh" (U5). This transparency could also help users trust the thermostat more by mitigating the negative perception that might result from mispredictions. U4: "If I never knew that Reef knew that [i.e., its limitations], and I came to my own [conclusion] and said like 'oh, Reef is terrible at figuring out when I am sick' then that makes me lose faith in the device."

4.4.3 Impertinence, Irrelevance & Allocation of Agency

By engaging participants through different scenarios and prototypes, we were able to gain insight into participants' attitudes about the appropriate allocation of responsibility between a thermostat and its users. Finding the boundary between what is and isn't appropriate for a smart thermostat to ask of its users is critical for mixed-initiative design, and more generally for designing technology that will be perceived as appropriate.

Participants' responses toward designs that suggest alternative indoor clothing were polarized (e.g., D3 and D7). Some expressed negative feedback on these designs because they thought they knew how to dress already. For some, such suggestions were seen as impertinent as they felt indoor clothing was a personal domain that the thermostat should not be involved with. U11: *"It's OK [for Reef to change] the temperature but I'm not very excited about it picking what clothes I want to wear. I find that's more personal… part of yourself. Temperature, you know, in your house, that's not part of you. What you wear, it's* *part of you. If I want to feel comfortable today, I'll wear this, if I don't, if I want to wear shorts around, I will wear what I want to" (D7)."* For others, clothing suggestions were acceptable but they were not enthusiastic about it as they had been wearing more layers of clothes at home already and viewed it mainly as reminders. Clothing selection, including what to wear at home, emerged as an area of personal choice incorporating lifestyle choice and personal expression, in addition to comfort preferences. The difference in reaction may be partly attributed to participants' views of the role of the thermostat, e.g., as a passive instrument versus a cooperative agent.

Besides clothing, the majority of participants reacted negatively to adaptive suggestions that required them to change their routines and living styles (e.g., D13, D14). Suggesting a move to a different room or exercising at certain time of the day to reduce heater usage are examples of such suggestions. Most of the user-generated tips we prototyped fell into this category. Participants hoped to see tips and suggestions that fit their current lifestyle. For example, U10 mentioned a tip he would like to share on the Reef forum (D14). After describing how warm he had to keep the house in order to keep his baby comfortable since safety guidelines prevent the use of blankets on infants, he added: "*It's like impossible. And we are on Amazon, and we found a wearable blanket that they can like clip on and zip into it, it's like a sleeping bag, that's attached to them, which then let us…keep us house colder,… If I was on this [forum], that would be a tip I would have."* This tip reflects a useful adaptive behavior tip that similar users would be more likely to accept. However, participants felt that tips that were irrelevant to them failed to resonate with their identity (e.g., as "parents") and damaged their trust in the system.

4.5 Discussion

Our findings allowed us to gain a deeper understanding of how various designs of comfort-aware eco-coaching thermostats might align with or oppose people's values with respect to comfort, sustainability, convenience, control and agency. Here we discuss how our study extended or challenged insights produced from prior research, specifically in terms of coordinating comfort and sustainability, balancing control and convenience, and allocating of agency while avoiding impertinence.

4.5.1 Coordinating Comfort & Sustainability

While prior systems have incorporated comfort sensing for temperature automation in offices (M. Feldmeier & Paradiso, 2010), there has been limited work investigating how people value comfort-aware systems in the home. Although we knew that people desire to be comfortable and hope to be environmentally responsible (Yang et al., 2014), it was unclear if people are open to smart thermostats that actively encourage sustainability while considering their comfort. Our study explored this question by surfacing comfort in different ways, including showing sensed comfort in visualizations (D1), suggesting short-term variation only when people have been comfortable for a few days (D4), suggesting adaptive behaviors when people may have a lower expectation on comfort (e.g., D15), and offering personalized saving modes (D8).

First, our findings point out that for some people, the comfort of certain others is more important than individual comfort. Interestingly, while prior research (M. Feldmeier & Paradiso, 2010) resolves multi-user conflict around comfort through automatically identifying a middle ground, our findings point to the possibility that in small households, visualizing comfort or using an important person as a reference might also be a solution, even a better outcome in some cases. In addition, our findings suggest that some people might value their comfort more as a necessity while others view it more as luxury. People who view comfort more as a luxury are more likely to compromise their comfort, and respond more favorably to recommendations of short-term variation. Furthermore, we found that while people desire comfort, their expectation and perception of comfort change throughout the day. This finding challenges previous measurements of comfort for smart thermostats: For example, PreHeat (Scott et al., 2011) aimed to optimize comfort by minimizing “MissTime”, which the authors define as the amount of time where the users are at home but the temperature does not reach the target temperature. Our findings point out that this measurement of thermostat efficiency is not in line with people’s changing expectations of their comfort. Finally, our design of saving modes raised some concerns. Inspired by the success of prior research (Pisharoty et al., 2015; Yang et al., 2016), we chose to use a similar framing of saving mode suggestions (i.e., "energy saving", "comfortable"). However, our findings point out that this framing renders energy saving and comfort in a mutually exclusive relationship. This implicitly highlighted conflict may steer people to weight their comfort as more important their adaptive potentials.

4.5.2 Balancing Comfort & Convenience

We knew from prior work that leaving users in control is essential, yet in a situation like thermostat scheduling, people appreciate some degree of automation (Alan et al., 2016; Yang & Newman, 2013). However, these prior insights were produced from reactions to smart thermostats that take a passive stance on energy savings (e.g., Nest). We were unclear whether these insights would still hold for eco-coaching thermostats that actively probe comfort boundaries by lowering the setpoint. It was also unclear whether implementing incidental intelligibility would help strike a balance between convenience and control, making them more tolerant to a proactive design even in the face of mistakes.

Our findings related to eco-coaching styles suggest that due to people's value of convenience, many people are accepting of a proactive design when doing temperature scheduling. This is of course contingent on people's trust in the system's capability to make data-driven decisions, and the perceived ease of control if adjustments need to be made. Although people might still adjust the temperature back if they feel uncomfortable, this brings up an opportunity to view eco-coaching as a negotiation process. Prior work has only explored eco-coaching in a more passive way (Alan et al., 2016; Pisharoty et al., 2015; Yang et al., 2016), but our findings point to the potential for smart thermostats to negotiate a more energy-saving configuration by setting a first reference point.

Our findings also point out that incidental intelligibility (Yang & Newman, 2013) was helpful for some participants, although error recovery was even more important when mispredictions happened. For these people incidental intelligibility helped them sustain their trust in the system. However, some other participants didn't value intelligibility as highly as we expected. We suspect there are multiple factors that contribute to the observed reactions. These factors include context (i.e., participants were asked to imagine being sick), the nature of interaction (i.e., system-initiated push notification), and the presentation (i.e., verbal explanation). More research is still needed to see whether implementing incidental intelligibility in user-initiated actions, and with other forms of presentation, might yield better results.

4.5.3 Being Pertinent & Respecting Allocation of Agency

Researchers (A. Clear et al., 2014; A. K. Clear et al., 2013) have already proposed to incorporate adaptive thermal comfort into the design of energy-saving systems. However, it was still unclear to us how to operationalize this concept in an appropriate way that can result in a useful system. Our findings suggest the importance of making adaptive suggestions that are pertinent and relevant to people's identity and situations.

Our findings point out that supporting appropriate user agency in making personal decisions at home is critical. Indoor clothing suggestions, even for passive designs like D3 and D7, might step into some people's private domain and be perceived as impertinent. Our findings also point out that many people already engage in adaptive behaviors. Suggesting these already-familiar adaptive behaviors without bringing in new insights comes across as ignorant and does not represent an effective approach to impacting behavior. Further, it is important to offer adaptive suggestions that fit people's lifestyles and routines. People are usually quite resistant to changing their lifestyle or routines. Many of the common adaptive behaviors mentioned in prior adaptive thermal comfort work (e.g., drinking hot beverages) elicited negative responses from our participants due to their highly user-dependent nature.

4.6 Limitations

The insights from our study were obtained through User Enactments, which involve short-duration engagements with possible future designs. We expect that users' reactions to the designs proposed in our study would evolve during a longer engagement. Our prototypes were also all probed in a single-participant setting. While we did ask participants to envision co-habitation, and we designed our interfaces with multiple users in mind, our study did not capture the complex social dynamics that may emerge in a longer field deployment. Given the early stage of design for both comfort-aware and eco-coaching thermostats, however, we felt that an important first step would be to survey a broader set of potential design directions using an approach that gives us insight into the underlying expectations and values that would drive responses to different approaches, as such values would influence both immediate and longer-term reactions. We look forward to future research that addresses the evolution of usage and complex social dynamics at home in a long-term deployment.

4.7 Conclusion

This work represents the first attempt to explore the design opportunity of *comfort-aware eco-coaching thermostats*: smart thermostats that can understand occupants' thermal comfort while persuading them to save energy. To research this opportunity, we created fifteen prototypes covering a diverse set of design attributes and conducted a user enactment-based study with 11 participants. Our study provides insights into how different designs of comfort-aware eco-coaching thermostats might align with or against people's values related to comfort, sustainability, control, convenience and allocation of agency.

Chapter 5. Leverage Eco-experimentation and Mutual-learning to Facilitate Residential Heating Control with ReefSetpoint

5.1 Introduction

Space heating alone consumes more than 43% of energy used in homes (EIA, 2015), resulting in significant CO_2 emission due to the heavy reliance on fossil fuel as an energy source in the U.S. ("U.S. Energy Information Administration (EIA)—Total Energy Monthly Data," n.d.). Much of the heating consumption can in fact be reduced by creating better tools to facilitate heating control, as researchers have shown that individual habits significantly impact a person's energy consumption (Sonderegger, 1978). For example, a large survey conducted in the U.S. shows that while 46% of households use thermostat-controlled heating temperatures (i.e., "setpoints") lower than 70°F (21°C), more than 31% of households use setpoints higher than 71°F (22°C) (Parker, 2013). Based on the rule of thumb that "1% of energy can be saved for each 1°F setback (i.e., a decrease of setpoint) over an 8-hour period" (Plourde, 2003), a significant amount of energy consumption can be reduced if people use lower heating setpoints.

To reduce energy consumption, numerous learning thermostats have thus been developed (e.g., Koehler et al., 2013; *The Nest Thermostat*, n.d.). Such thermostats aim to facilitate heating, cooling and ventilation control (HVAC) by taking into account various environmental and human signals (Yang & Newman, 2013). For example, the Nest thermostat detects occupancy status of a house and triggers energy-saving setpoints when occupants are away from home (*The Nest Thermostat*, n.d.). As wearable devices become increasingly available, there is also a growing number of studies that aim to take into account individual's comfort preferences by sensing physiological signals such as skin temperature and learning their correlation with comfort (Gao & Keshav, 2013b; Huang et al., 2015). While such learning thermostats were shown to reduce energy consumption, long-term evaluation studies reveal that imperfect learning capabilities and lack of user control lead to wasted energy and reduced comfort (Yang & Newman, 2013; Yang et al., 2014).

To address these issues, researchers recently proposed investigating eco-coaching approaches (Pisharoty et al., 2015; Yang et al., 2016). Eco-coaching approaches aim to remediate artificial agents' limited intelligence by broadening possible actions to take and considering the human in the loop. They consider users as part of the system, helping to achieve goals together with the agent. Along with automation, smart thermostats designed with eco-coaching approaches can also suggest actions or inform users about potential alternatives. However, studies of such eco-coaching systems are limited in the strategies considered. Questions remain on how to design a learning thermostat to best support eco-coaching.

In this work, we propose two new eco-coaching strategies, *eco-experimentation* and *mutual-learning*, for learning thermostats to further facilitate heating control. This work is inspired by prior studies (Huang, Liang, Wu, & Newman, 2017; Huang et al., 2015; Snow, Auffenberg, & Schraefel, 2017) and our experience designing thermostats with coaching and comfort inference capabilities. First, we believe that for domains where users do not yet know a good solution or are biased by preconceived opinions (e.g., regarding what setpoints to use at home), a smart system may facilitate problem solving by actively trying out different possible solutions in parallel. We refer to the eco-coaching approach that uses this strategy as *eco-experimentation*. Second, similarly to what Snow et al. (2017) proposed, we believe that to design a learning thermostat, we should not assume that learning happens only on the system side; it should happen on the user side as well. We define a learning process that supports both users and systems to learn as *mutual-learning*.

Based on these ideas, we created a comfort-aware thermostat, ReefSetpoint (abbr. Reef), which can leverage both *eco-experimentation* and *mutual-learning* to support eco-coaching.

Reef facilitates both occupants and itself to learn about better settings by actively trying different temperature setpoints. Further, it facilitates user reflection by displaying data collected for model building back to user in an easily understandable fashion.

In studying the lived experience of Reef in five households with ten participants for four to five weeks, we make two contributions in this chapter:

1) **Eco-experimentation:** We show that a smart thermostat can help its occupants identify less resource-intensive setpoints to use by actively experimenting with different options during a learning period.
2) **Mutual-learning:** We show that a learning thermostat that also supports user learning can help users identify actions that can lead to a better balance of comfort and savings considering the needs of multiple household members.

5.2 Background & Related Work

In this section, we briefly summarize the research on intelligent thermostats from three perspectives: 1) What contexts do they consider? 2) What actions can they perform? 3) How do they learn?

5.2.1 Contextual Inputs to Smart Thermostats

In the past two decades, there has been considerable work on smart thermostats that aims to automatically detect relevant contextual states in order to manage heating and cooling control (M. C. Feldmeier, 2009; Gao & Keshav, 2013b; Gupta et al., 2009; Koehler et al., 2013; M. Mozer et al., 1996; Ranjan & Scott, 2016). These context-aware thermostats may detect occupancy patterns (Koehler et al., 2013), setpoint preferences at particular times of day (*The Nest Thermostat*, n.d.), and dynamic pricing changes (Alan et al., 2016), and use this information to control the temperature. Lately, with the advancement and increasing availability of wearable devices, researchers have begun to take physiological signals such as skin temperature, activity level, and correlated personal comfort levels into consideration (M. C. Feldmeier, 2009; Gao & Keshav, 2013b; Huang et al., 2015). Such comfort-aware thermostats are designed with the goal of sensing and inferring more fine-grained individual levels of comfort. Early studies of these thermostats (Chaudhuri, Zhai, Soh, Li, & Xie, 2018; Gao & Keshav, 2013b; Huang et al., 2015) show the feasibility of inferring personal comfort using wearable devices and the potential to dynamically adjust heating and cooling based on multiple people's comfort preferences. However, it is still unclear how systems that take into account such information can support a household in making HVAC-related decisions.

5.2.2 Actions of Smart Thermostats

Smart thermostats that automate HVAC control seek to reduce user burden by completing tasks with little to no input from users (Yang et al., 2016). For example, occupancy-based thermostats eliminate the need to manually configure a setback schedule (i.e., a schedule that indicates when to use particular setpoints) by using sensors to detect occupancy status and to trigger predefined setpoints that correspond to the home's occupancy status (i.e., occupied or not).

While these automation approaches have resulted in savings of roughly 10% (Koehler et al., 2013; Nest, n.d.), new issues arise as the thermostats start to consider more facets of context to assist control. First, studies show that existing systems have a hard time catching up with family members' changing daily routines and the corresponding preferences (Yang et al., 2014). As these automated systems expect little to no input from users, small adjustments made by the users have been observed to lead to unintended system configurations and wasted energy. In addition to the issue of coping with people's changing routines, some contextual information such as individual comfort (while valuable) cannot be inferred accurately all the time (Huang et al., 2015). Furthermore, there may not exist a simple policy that maps a detected context to an expected action due to variables hidden from the system, such as presence of visitors. How to use such a valuable but uncertain context-based predictions with potentially complex decision policies remains a challenge.

To address these issues, mixed-initiative approaches have been proposed, and some preliminary systems have been built and deployed (Pisharoty et al., 2015; Yang et al., 2016). Mixed-initiative interaction (Allen et al., 1999; Horvitz, 1999) views artificial agents as capable of acting with different levels of autonomy (Goodrich & Schultz, 2008; Sheridan & Verplank, 1978) (e.g., supporting reflection or making suggestions rather than solely automating tasks) based on their knowledge of users and their confidence of prediction. Moreover, it views users and systems as partners, achieving goals together; agents can ask for help and users can provide it if needed.

One type of thermostat that leverages mixed-initiative interaction to facilitate energy-saving has been called an "eco-coaching" thermostat. For example, Pisharoty et al. (2015) created a thermostat that can provide different setback schedules with different comfort and energy-saving tradeoffs based on learned occupancy patterns. Users are then presented with recommended schedules and can decide whether to adopt one of the system's recommendations. In generating these schedules, the system mainly varies the timing of energy-saving setbacks based on occupancy patterns. It does not question the default setpoints used by the participants. The authors showed that by making personalized schedule suggestions, their thermostat helped occupants identify alternative schedules leading to more energy savings. However, the work thus far on eco-coaching has been limited to setback scheduling, while there are many other areas a thermostat may coach, such as ideal setpoints to use when people are at home.

5.2.3 Learning Process of Smart Thermostats

In the past, learning thermostats were often assumed to be artificial agents that passively learned about users' context and preferences (Huang et al., 2015; M. Mozer et al., 1996). They would go through a learning period to collect data, and then act according to the models built from the data afterwards. Under this paradigm, researchers assume agents are the ones that require learning, not the users. However, in many situations, it is beneficial for users to learn about their own preferences as well, such as for households with multiple occupants.

Recently, researchers have proposed the notion of co-performing agents (Kim & Lim, 2019; Kuijer & Giaccardi, 2018), an idea that emphasizes an artificial agent's capability to perform tasks together with users. As both of them learn about the tasks and each other, the roles and functions of the agent evolve over time. In this concept, users are considered as learners and performers during the learning process. To the best of our knowledge, we have not seen work on learning thermostats that view learning in this fashion. The closest work we can identify is (A. K. Clear, Mitchell Finnigan, Olivier, & Comber, 2018), where the authors created a system to facilitate multiple occupants in an office to reflect on an acceptable range of indoor temperature by collecting and visualizing individual comfort feedback on a public display. However, the focus of their work was in office settings rather than homes. It is well-known that homes have very different social dynamics than offices and often require different design considerations (Crabtree, Rodden, Hemmings, & Benford, 2003).

Our work differs from prior research in three aspects. First, our thermostat system supports coaching on temperature setpoints, rather than setback scheduling. Second, we explore *eco-experimentation* as a strategy for coaching rather than offering suggestions. This means that our system actively facilitates users to try out and evaluate different potential solutions. Third, inspired by the notion of co-performing agents, our system is designed to support both users and the system to learn about comfort preferences during the learning process, rather than simply constructing models based on comfort feedback. As users learn about their own preferences and the agent's capability, they may configure the agent to act differently and expect different degrees of control. We call this learning process *mutual-learning*.

5.3 System

To investigate the feasibility of leveraging *eco-experimentation* as an eco-coaching approach, and to explore how to design thermostats that support *mutual-learning*, we developed ReefSetpoint (abbr. Reef). Reef is a comfort-aware thermostat that leverages

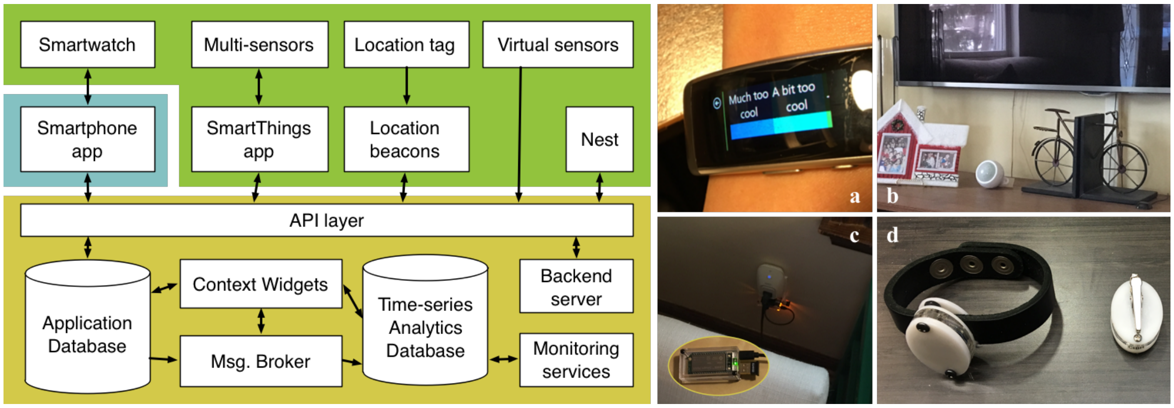


Figure 4. Left: A simplified system architecture of Reef; Right: a) Reef's Microsoft Band 2 interface; b) A multi-sensor deployed in a participant's home; c) A deployed location beacon; d) Location tags with different form factors.

the two above-mentioned strategies to encourage less resource-intensive heating choices.

In this study, we focus on investigating the initial learning process of Reef, a period where it collects comfort feedback from users in preparation for creating personalized comfort models. Reef did not perform any online learning nor did it attempt to recommend or enact setpoints or setback schedules based on inferred comfort models. Demonstrating the accuracy and acceptability of Reef's modeling, prediction and control is ongoing work and beyond the scope of this chapter.

5.3.1 Features of Reef

Reef supports four major functions. First, it can actively explore different possible setpoints to use when people are away, at home and sleeping, namely, the AWAY, AWAKE, and ASLEEP setpoints. Second, as a thermostat designed to support learning of comfort preferences, it aggregates a variety of data relevant to personal comfort from indoor, wearable and Web-based virtual sensors. Third, Reef leverages experience sampling (Hektner et al., 2007; Stephen S. Intille, Rondoni, Kukla, Ancona, & Bao, 2003) to request comfort feedback from users via micro-questionnaires (S. Intille, Haynes, Maniar, Ponnada, & Manjourides, 2016) implemented on smartwatches. Reef uses its context sensing capability to trigger different micro-questionnaires. Finally, while Reef collects users' comfort feedback for performing comfort modeling, it also displays the collected data back to users to facilitate user reflection on preferred setpoints (see Figure 5).

Note that there are many possible models of comfort preferences. We mainly considered two types of models in our work. The first is a simplistic model where the system simply learns to associate a satisfaction score with a setpoint, given a setpoint mode (AWAKE, ASLEEP, AWAY). Once it learns, it can then pick a setpoint for a mode that maximizes satisfaction. The other approach is a machine-learning-based model (Huang et al., 2015), where the system can leverage multiple sensor data streams to infer comfort sensation at the moment.

In the following sections, we describe the technology components of Reef that allows it to support the above-mentioned functions.

5.3.2 Perception & Control Component

Reef uses a Nest thermostat (*The Nest Thermostat*, n.d.) to control heating and cooling and integrates four types of sensors to collect information relevant to individual comfort. First, a set of *multi-sensors* ("AeoTec MultiSensor," n.d.) are used to detect room temperature,

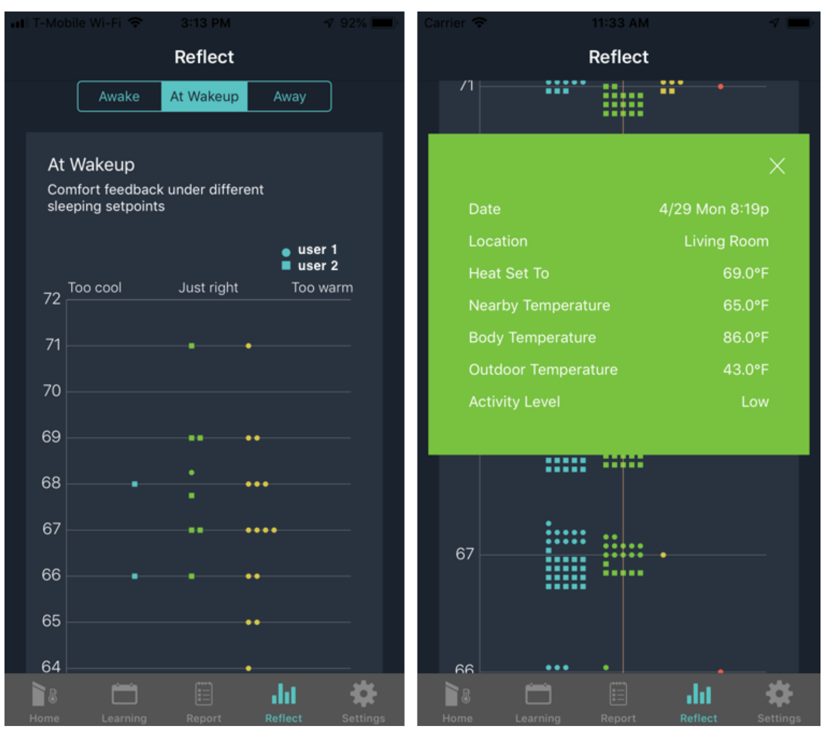


Figure 5: Phone interfaces to support reflection. Left: Feedback for **ASLEEP** mode; Right: Feedback for **AWAKE** mode. Clicking on a dot will reveal detail context.

humidity, and motion in different areas of a house. Second, a *Microsoft Band 2* ("Microsoft Band 2," 2019) is used to detect physiological signals including skin temperature, galvanic skin response, and activity level of each user. Third, a *set of cloud-based virtual sensors* is used to detect outdoor temperature and humidity in the area of a house, and the heating and setpoint status of the Nest thermostat. And finally, to track the location of house occupants we built a *Bluetooth-based room-level indoor positioning system* (similar to [Faragher & Harle, 2015]) to infer indoor location in real-time. Using this system, a location model was trained for each house during initial home visits. Our system reaches higher than 90% accuracy for the test datasets we collected in each house.

All these sensor signals are pushed to the inference and decision-making (I&D) component in near real-time with a streaming-based approach, thus allowing the I&D component to infer secondary context and to take actions such as sending context-aware micro-questionnaires.

5.3.3 User Interaction Component

A watch app and a mobile app were built on Microsoft Band 2 and iOS. The watch provides an interface for users to quickly report their comfort feedback. In this study, we use five-level comfort labels similar to (Huang et al., 2015). These labels include: "much too cool," "a bit too cool," "just right," "a bit too warm," and "much too warm".

We built a mobile application to let users manually adjust their thermostat setpoints. The app also allows occupants to configure the parameters relevant to *eco-experimentation*, such as the temperature range Reef uses for eco-experimentation. Further, to support household members in reflecting on their collective preferences, the app provides visualization of occupants' comfort feedback and the contexts associated with each feedback (e.g., where were they when they felt cold). We designed multiple dot plots to highlight the relationship between setpoint, indoor temperature, and comfort among multiple household members (see Figure 5). Users can also tap on a data point to view detailed context, including the time, location, activity level and other sensor data associated with comfort feedback.

5.3.4 Inference & Decision-making Component

The inference component infers several secondary contexts, including if occupants are asleep, their indoor locations, and the home/away occupancy status of the house as a whole. For simplicity, Reef determines asleep status based on users' predefined wake up and sleep times, though future versions could infer sleep status more dynamically (Pisharoty et al., 2015). For occupancy status, Reef integrates data from all the motion sensors and the indoor localization system to make an inference. Some households may have occupants that do not wear a watch or location tag, such as children. In this situation, Reef would still use motion sensors to detect their presence and avoid triggering its AWAY mode.

In addition to inference, this component also selects different setpoints given different activity contexts (AWAKE, ASLEEP, AWAY), and sends micro-questionnaires to obtain comfort feedback via remote notifications.

Random Sampling of Setpoints: For Reef's learning phase, we employed a random probing policy in choosing the setpoint to use for a given time of day. Reef first randomly selects a setpoint for each possible activity context of a day (AWAKE, ASLEEP, AWAY). The range of allowable setpoints for each context is configured by the users before the learning begins. Once each context's setpoint is chosen, Reef monitors the users' context and determines which setpoint to use for a particular time. During learning, Reef also detects if the user overrides the setpoint and, if so, suspends control of the thermostat and maintains the user selected setpoint until the next morning.

We chose random sampling over other approaches (e.g., reinforcement learning [Sutton & Barto, 1998]) as it was unclear to us how to balance exploration (i.e., trying out alternative setpoints) and exploitation (i.e., use users' prior setpoints) in this application setting. Random sampling provided us the ability to repeatedly sample a single setpoint on different days, providing data about that setpoint under different conditions. Even though people may be uncomfortable with a setpoint some days, they may be comfortable with it some other days due to factors other than temperature (A. K. Clear et al., 2013; de Dear & Brager, 1998).

5.4 Methods

5.4.1 Participants

We recruited participants through local mailing lists, online forums and Facebook ads. Participants were required to be iPhone users and live within 40 miles of our university campus. In addition, while each participating household had to contain exactly two adults, we imposed no constraints on the number of children and pets.

In total we recruited 5 households and 10 participants (2 per household). Our participants ranged in age from 29 to 65+ and had different numbers of children and pets (see Table

Table 7: Participants information

PID	Pseudonyms	Children & Pets	Thermo. Before Study	Prior Setpoint	AWAKE Range
H2a	Steve	1 Dog	Nest	70	68-71
H2b	Sharon				
H3a	Andrew	2 Children & 1 Dog	Programmable	68	66-69
H3b	Mia				
H4a	Jane	2 Dogs	Programmable	68	66-69
H4b	Tyler				
H5a	David	None	Programmable	70	67-70
H5b	Diana				
H6a	Chloe	1 Child	Manual	72	67-70
H6b	Mason				

7). The living space of these houses ranged from 800 to 1400 sq ft (74 to 130 m^2). The setpoints used by these households prior to the study also differed — ranging from 68°F (20°C) to 72°F (22°C). The parameters they used for experimentation also differed. Participants were compensated based on their completion of home visits, interviews, as well as optional comfort feedback and a weekly diary. Each participant received between \$16.2 and \$49.1 (mean=\$29.7). Each household also received a Nest thermostat as a gift.

5.4.2 Field Deployment Study

Prior to our deployment study, a pilot study was first conducted in two of the authors' homes (H0, H1) with two adult participants in each home for 1-2 months to ensure stability of the system and to sharpen the user experience of the thermostat. Afterwards, the actual deployment happened between late March and early May, 2019. At the time of the study, the study area had a mean temperature of 49°F (9°C) and a minimum of 22°F (-5°C). All participants actively used their heating systems throughout the study.

Our study included three phases: pre-deployment, deployment, and post-deployment. During the pre-deployment phase, we visited each home two times for two hours each. During these visits, we installed all the study equipment and software. More than 20 sensors and devices were deployed in each household, with a total of more than 100 sensors and devices running during the study. Between the two visits, an interview and a questionnaire were also conducted with each participant remotely. This helped us understand each participant's attitudes toward comfort, heating, energy savings and sustainability.

During this period, participants were also informed that this was a research study to investigate thermostats that could learn about their comfort preferences, and the focus of the study was on its initial one-month learning period. As Reef was still learning, it would not control their setpoints based on the comfort feedback collected in this study.

Afterwards, we started the deployment phase which lasted 4-5 weeks for each household. Participants were encouraged to respond to micro-questionnaires sent from the system and to initiate comfort reports themselves six times a day. We received a median of 175 comfort reports per participant (6 per day) during the deployment. Participants were also encouraged to respond to weekly diaries sent via email. These diaries helped us better understand their experience living with Reef. If participants did not like the setpoint experiment by the system, they could also override it or reset the learning parameters. Two households (H5, H6) updated their learning parameters in the first week of the study, as they found the minimum setpoint used by the system to be too uncomfortable. While H5 successfully updated the range to match their preferences, H6 accidentally lowered the range to be 67-70°F, a range much lower than their previous setpoint at 72°F. None of the research team and the participants found out this issue until the exit interview.

After the deployment phase, we then conducted a 1-hour exit interview and questionnaire remotely with each participant to gain a deeper insight of their experiences.

5.4.3 Interview Study

In this chapter, we focus the analysis on our participants' experiences living with Reef; this is mainly based off of the analysis of the qualitative data obtained through interviews, questionnaires and weekly diaries. We first transcribed all the exit interview recordings, then used thematic analysis to create an affinity diagram mixing all the data from the interviews, questionnaires, and diaries to identify themes.

5.5 Findings

In this section, we will first report findings on the usefulness of having an initial learning process that leverages eco-experimentation and supports mutual-learning. We will then reflect on Reef's design attributes that may contribute or hinder its capabilities to support agent learning and user reflection. The names used in Table 7 and the following sections are all pseudonyms.

5.5.1 Eco-experimentation and mutual-learning led to better understanding of preferences

5.5.1.1 Eco-experimentation helped users identify better AWAKE *setpoints*

Even though all of our participants had been living with a thermostat for a long time and had been trying to reduce their heating costs, five of the participants found out they could live with a lower setpoint after going through Reef's learning process. For example, Mia found that while she had already used an eco-friendly setpoint prior to the study, she did not realize that she could live with an even lower setpoint: *"Just before we always thought that 67°F and 68°F were the most common eco-friendly temperatures and we didn't know that we would be fine at an even lower temperature [66°F]. So we never really thought to experiment with a lower temperature. Now that we have, especially [living] with the Reef app and the Nest Thermostat [as part of the Reef system], it's nice to know that we can have an even lower temperature in the house."*

While the other five participants did not identify a new setpoint to use while they were at home, two participants found that they could be comfortable under a wide range of setpoints — even though sometimes they felt cold under cooler settings, often times they were not. Steve explained that going through this learning process made him aware of his wide comfort zone: *"I didn't realize that I felt comfortable at such varying temperatures because [the temperature detected by Nest while I reported being comfortable] ranged from 65°F all the way up to 75°F."* Note that Steve was reflecting based on the visualization that showcased the correlation between indoor temperature detected by his Nest and occupants' comfort. Although Reef only experiment 68-71°F in Steve's household for AWAKE setpoint, the indoor temperature could sometimes reach lower or higher due to transition between setpoint modes (e.g., from AWAY to AWAKE) and warmer outdoor temperatures. This caused some comfort feedback to be collected with an associated indoor temperature outside the range of experimentation.

For these two participants, while they did not prefer to stay with a lower setpoint all the time, this understanding of their "lower but mostly OK" setpoint may help them accept a cooler setting for a short period of time. For example, when asked about if he was willing to use a lower setpoint for one day a week, Andrew responded: *"I think that I would, based on the data that we have. ... I think before doing this [setpoint experimentation], if I had a thermostat that said that [use 66°F for a day], I would be, 'Absolutely not.' But now I think, just looking at how often I was comfortable at 66°F, yeah, I think I would be comfortable with saying we're going to try that out, one day a week [could] be at 66°F. Yeah."*

5.5.1.2 Eco-experimentation helped people identify better ASLEEP setpoints

It was challenging for Reef to collect people's comfort feedback with regard to ASLEEP mode. This was due to our limited instruments for experience sampling. Reef only sent one micro-questionnaire per day with regard to ASLEEP mode at a predefined time in the morning. Participants often still had their watches charging and their phones on nightstands. This caused the micro-questionnaires to be easily overlooked. Due to this reason, only three participants (Sharon, Mia, Jane) provided more than 10 days of sleep time feedback (across the entire 4-5 week study period).

Despite the limited feedback, two participants among the three found out that they could live with a colder ASLEEP temperature. For example, Mia (H3b) had previously changed the setpoint in her house from 68°F to 71°F at night when she slept as her bedroom tends to be colder. After living through Reef's experimentation process, she found out that she could sleep comfortably with a lower setpoint. *"I really like that it dropped the temperature a little bit at night. There were times I was like, 'Oh, it's really, really cold,' to start off, but throughout the night I would get really warm because the dog's in there with me and then the two kids are on both of my sides. So it gets hot. That was nice that I learned that eventually it's nice that it starts out cold because it's going to get hot."*

5.5.1.3 Mutual-learning helped occupants understand each other's preferences

While all of our participants had been living with their partners for multiple years and they felt that they generally understood their partners' temperature preferences prior to the study, almost all of the participants responded that they learned more about their partners' temperature preferences after the study. For example, Sharon stated: "*I actually thought* *that we would be more different. We ended up being pretty close together. That's good to know."*

In three households, this understanding helped them identify a more energy-saving but acceptable setpoint to use. When asked about the preferred setpoints for their households in the exit survey, six participants (from H2, H4 and H6) responded with a setpoint that was lower than what they had previously used. According to the visualization shown by Reef, they believed this lower setpoint would work for the household. For example, before the study, Steve and Sharon typically used 70-71°F. After living with Reef, Steve suggested that 68°F would work for them and Sharon suggested 69°F. If these participants decided to control the setpoints themselves after the learning process, they would likely reach an agreement between 68°F and 69°F, resulting in a setpoint around 1-2°F lower than what they used previously. This could reduce heating usage in their house by one to four percent based on the heuristic that reducing 1°F for an 8-hour period each day can lead to 1% saving (Plourde, 2003).

5.5.2 Human-in-the-loop led to better eco-experimentation

5.5.2.1 Anchoring experimentation on prior settings led to higher acceptance

The majority of the households found the learning range used by Reef to be acceptable, even though they were uncomfortable sometimes. For these households, they used a default AWAKE learning range recommended by Reef, which ranged from 2°F lower than their previous setpoint to 1°F higher. For example, Taylor and Jane (H4) preferred to use 68°F before the study and they decided to use 66°F to 69°F as the range of temperature for Reef to try out. Afterwards, they found this range to be acceptable. Tyler explained that although there were times that he felt uncomfortable, the learning process was painless for him: *"I mean, there was one time in the middle of the night I woke up pretty cold. Other than that, it was a painless situation with [Reef] trying to learn my likes and dislikes."*

However, for two other households that configured Reef to use an AWAKE or ASLEEP learning range that reached more than 2°F lower than their previous setpoints, they felt that the temperature range was too extreme and had to adjust the range in the first week of the learning period. For example, Chloe and Mason (H6) configured their Reef to use an ASLEEP learning range that ranged from 65°F to 69°F while they previously used 71°F while sleeping. Noted that for ASLEEP temperature, Reef suggested *3-6*°F lower than their prior setpoint as the learning range to encourage sleep-time energy savings, although participants could still edit it to whatever they wanted. After a few days, Chloe found the temperature to be unbearable and she decided to adjust the range. *"So I thought we had set [Reef] to 68°F for when sleeping but last night I froze all night and when I checked it around 5 am it said it was at 65°F with most rooms in the house having dropped to 61°F [detected by Multisensors], so I adjusted its learning mode for night time hours."* Similar things happened in H5 where the participants had to adjust the AWAKE learning range in the first week. The adjusted learning range still allowed Reef to explore 2°F lower than their prior default setpoint, but not lower.

5.5.2.2 Supporting easy override and "pause of experimentation" led to better user control

In our design, despite Reef actively trying out different temperatures, it would also pause its experimentation at any point until the following morning if participants overrode the temperature. Under this design, all of the participants felt that they were still in control of their thermostats even if Reef experimented with different setpoints. *"I felt I was [always] in control. ... I felt the system was providing suggestions to me. Then I could act accordingly if I didn't like the suggestion. I felt I didn't feel it had control of the house or whatever of the temperature. Because at any time I could control it and adjust it." (Tyler)*

5.5.2.3 Micro-questionnaires reduced burden of system training

Reef prompted users for comfort feedback every 40 minutes while they were home, and the majority of participants found this frequency to be acceptable. We believe Reef's usage of micro-questionnaires on a smartwatch reduced the burden of reporting, making the high frequency of prompting acceptable. For example, Chloe, a participant that reported her comfort feedback 97% of the time from her watch, stated: *"The notification didn't bother me. ... because the [watch] interface, how you reply, is so simple and automatic. I could be in the middle of watching a movie, and hear it go off and just press on it and it didn't even* *interrupt me."*

However, depending on the versions of their phones, participants with older iPhones tended to have connectivity issues with their watches. They therefore reported their comfort on their phones more often, which potentially caused annoyance given the high frequency of experience sampling. For example, Steve — a participant who reported his comfort 90% of the time from his phone, as he had trouble getting his watch connected –– stated: *"It was a little annoying to get the notifications when you're at home every couple of hours, every hour. ... I mean, the first week wasn't bad, but then after a while it would really start to be, yeah, I'd say annoying. Irritating."*

5.5.3 Agent learning and user reflection were reciprocal

5.5.3.1 Agent learning supported user reflection

We found that providing views of previous comfort reports during the learning period enabled user reflection, as participants were motivated to provide comfort feedback to train their thermostats. This therefore offered enough data for reflection. All of the participants found the benefits of having a system learn their preferences outweighed the risk of discomfort. For example, Jane stated: *"Overall, seeing the end result and knowing what the goal is, I think [this learning process] is fine. You just have to stick with it to have it learn where your comfort zones are."*

If agent learning (i.e., collecting comfort feedback to train comfort models that can be used to facilitate thermostat control) were to have been removed from user reflection, we may have had difficulty obtaining comfort feedback and convincing people to reflect on their preferences under a lower temperature. For example, Mason explained that if the thermostat did not show any sign of learning after training it for a month, it would be frustrating for him to go through this experimentation process. Even though reflection could provide benefits at the end, it appears that Mason was primarily motivated to train a thermostat that could understand his preferences, rather than reporting for self-reflection: *"If we would have done this study over the course of an entire season, and if it didn't seem to be learning, I think that would be a little frustrating."*

5.5.3.2 User reflection supported agent training

Almost all of the participants found the visualization feature to be useful for making sense of their comfort preferences. For example, Sharon explained: *"I think it's mostly from looking at the graph [to make sense of my preferences]. I don't think I consciously was looking at what is the temperature now [on the physical thermostat] and how comfortable I feel. I think I needed the summary, the reflection report to see that."*

While we designed the visualization to support reflection, we found that displaying data back to users also made them felt that the system was making progress, potentially leading participants to feel that the system was more trustworthy and intelligent. For example, Andrew explained: *"I felt more comfortable probably having the opportunity to review the data and anything like that. ... It's transparent enough that I got to experience it. ... And I found it [Reef displaying data] more interesting than just being a part of [the learning process], I actually got to go in [to the mobile app's visualization] and see some of how it [Reef] worked."*

5.5.4 Shortcomings of Reef

5.5.4.1 Failing to remediate when people reported being uncomfortable

In our study, Reef did not adjust the setpoint right away when people reported being "much too cool" or "much too warm". We expected people to manually adjust the setpoint via the phone interface after offering comfort feedback. It was a challenge during design to balance the two needs: the need to make *repair* easier if errors happened, and the need to facilitate mutual-learning and adaptive behaviors (that is, discouraging immediate setpoint adjustment and encouraging users to engage with alternative practices to keep themselves comfortable). We thus designed the phone interface to allow adjustments once users provided a comfort feedback; there was no way to adjust temperature on the watch. Through this design, we increased the steps required to achieve temperature adjustment in the hope to discourage frequent manual changes.

However, two participants found this to be annoying, as they expect if a thermostat was intelligent enough to ask people "how do you feel now?" it should know to take action if the answers were "much too cool." For example, David explained: "*The thing I'm not sure what … Reef is supposed to do. It didn't seem like if I kept saying it was cold, it never really adjusted. That's the part that was a little bit, I would say confusing, or sometimes irritating a little bit.*" Since David primarily interacted with Reef through his watch, the lack of a temperature adjustment on the watch made him feel like this "intelligent" system was ignoring his request.

5.5.4.2 Inability to differentiate override intention

Although Reef's ability to pause its learning in response to user override helped participants to feel in control, the duration of the pause could cause confusion. In our study, Reef would disable its experimentation until the next morning's first AWAKE mode occurrence. However, this meant that if participants overrode its setpoint in the morning, the setpoint would stay the same throughout the whole day and night. For some participants (e.g., H5) that used a night-time setback prior to the study, they experienced some warm nights and were not able to sleep well. They had to remember if they overrode once in a day, they had to provide manual control the whole day — this could be annoying and confusing as people did not always remember these system behaviors.

5.5.4.3 Lack of iterative and online learning

Participants were informed about Reef's one-month learning period prior to the deployment. While the majority of the participants accepted this duration, two participants expected the benefits of learning to start earlier. They expected Reef to learn more actively — collecting feedback for two weeks, act based on the feedback, and then continue to learn with a reduced frequency of experience sampling. For example, David explained: *"I could live with getting cool for one to two weeks, and [having Reef] getting that [comfort] information and I could see the [result of] learning. … If I see learning, and I see where there's intelligence going on there, I can live with it. If I can add one more thing here. Even after a month or two… If it tries a little bit lower every now and then, just to see if data is [still] correct, I don't see that as a problem."*

5.6 Discussion

Reef facilitates mutual-learning and leverages eco-experimentation as two strategies to support eco-coaching. It actively explores different possible setpoints to use that may reach a better balance between comfort and savings, and it shares training data back to users to support reflection among multiple household members. These approaches allowed Reef to offer four benefits: it helped participants identify better setpoints to use when they were active at home (AWAKE mode), and when they were sleeping (ASLEEP mode); it helped some participants to accept short-term saving suggestions (e.g., lower the heating setpoint for 1-2°F one day a week (Huang et al., 2017)); and it helped occupants better understand each other's comfort preferences, which can lead to better setpoint settings that increase comfort or savings. In the following sections, we highlight implications for design to better support eco-experimentation and mutual-learning.

5.6.1 Providing mildly-challenging while respectful eco-experimentation

5.6.1.1 Personalize experimentation settings and allow pause of experimentation

During design, our team extensively debated the choice of setpoint range that should be used to support experimentation. If selected inappropriately, we either hindered the ability of the thermostat to learn, or put users at risk of discomfort (which could, in turn, threaten user acceptance). In the end, we opted for a mildly-challenging approach, where the system would offer a default suggestion to start with that was generated based on people's prior setpoints with a small deviation (e.g., in the case of AWAKE setpoint, from two degrees lower than the prior setpoint to one degree higher). This design was inspired by Yang et al.'s (2016) finding that many people are willing to explore settings that are slightly outside of their comfort zone. While users could adjust the learning range beyond the defaults the system would enforce a four-degree difference between the minimum and the maximum setpoints, ensuring sufficient space for exploration. Similar to (Pisharoty et al., 2015), Reef also displayed different benefits and drawbacks given different ranges.

While the design of Reef prioritizes experimentation, we were aware that life is full of unexpected events that are challenging for a smart system to anticipate (Bellotti & Edwards, 2001). To cope with these events, Reef allowed users to override its setpoint, either through the physical interface of Nest, or through Reef's mobile application. User overrides paused the experimentation until the next morning. With this approach, Reef respected people's comfort needs, treating them with higher priority than its learning objectives.

We found that our design worked well for the majority of our participants. Although there were times where they felt uncomfortable, they felt that it was easy enough to quickly adjust the setpoint themselves. They also felt that the benefits of having a system to learn about their comfort preferences outweighed the potential for temporary discomfort.

However, for a few participants that (for unknown reasons) started with a range that reached lower than 2°F compared to their prior setpoints, they quickly found out that these setpoints caused them to be "much too cool" consistently and updated the experimentation range in the first week. These participants were also frustrated about Reef's lack of reaction after they reported being "much too cool." The thermostat was perceived to be inconsiderate as it did nothing when people felt uncomfortably cold.

Our override design also caused issues as participants did not always remember that once they overrode the temperature, they had to manually adjust it until the next morning, including manually setting a night-time temperature. A better design should ask people about the duration of "pause," allowing participants to disable it until the next morning, before sleep, or for the next 3 hours.

Our findings therefore indicate that a 2°F deviation based on prior setpoints may be a sweet spot for an initial round of experimentation. We also show that if users consistently report "much too cool" for a certain setpoint, the system should significantly reduce the chance to sample this setpoint, and it should adjust the temperature right away. Based on our findings, we suggest that future eco-experimentation systems support personalized learning parameters and offer more flexible override mechanisms.

5.6.1.2 Provide benefits of learning sooner

To support eco-experimentation, one question we have is how long the initial learning period should last. If it is longer, the system can obtain more training data to build a more accurate comfort model, and thus deploy control policies that better respect users' comfort needs. If it is shorter, users may get some benefits earlier that motivate them to continue training and using the system. In the end, we decided to explore a one-month learning period as we estimated that this would provide sufficient comfort feedback for each experimented AWAKE, AWAY and ASLEEP setpoint.

We found that one month of initial learning worked well for the majority of the participants. They were willing to continue offering feedback even though the system did not leverage their comfort feedback to perform actions (besides supporting user reflection). However, for a few participants, they felt that this learning period was too long, and they expected to see benefits earlier — within 1-2 weeks. That is, they expected the system to learn more iteratively, taking in feedback for 1-2 weeks, then begin using this information to perform actions while continuing to request comfort feedback — ideally with a reduced frequency. Thus, to improve the experimentation process, we recommend future systems offer benefits sooner, such as considering a more iterative learning process (Horvitz, 1999; Sutton & Barto, 1998). However, how to design an iterative experimentation process is still a challenging question, as it involves creating smart home systems that can dynamically adjust their coaching strategies based on certainty of their knowledge of users' preferences. More research is still needed to explore such mixed-initiative systems (e.g., Horvitz, 1999) in the context of homes.

5.6.1.3 Leverage context-aware micro-questionnaires

Another design challenge we faced was determining the frequency of requesting comfort feedback. We wanted as much feedback as we could get while not burdening our users. To address this, we used micro-questionnaires (S. Intille et al., 2016) where users easily tapped a button on their watch to respond to comfort sensation questions. This can be done within 1-3 seconds, and for many participants, they expressed that this was easy enough that they could, for example, answer while watching movies without feeling interrupted.

However, while this micro-questionnaire approach worked well for collecting comfort feedback for the AWAKE setpoint, it did not work well for the ASLEEP and AWAY feedback. To collect feedback with regard to the ASLEEP and AWAY setpoints, the system used sensors and user-provided information to infer context and to trigger questionnaires at specific moments. For example, the system used a pre-defined wakeup time to trigger micro-questionnaires for the ASLEEP time feedback. It turned out we collected very few reports with regard to AWAY setpoints (feedback for which was requested shortly after users returned home), and only some reports for ASLEEP setpoints due to the limited instruments for experience sampling. For example, participants who still had their watches charging when receiving an ASLEEP questionnaire easily overlooked this question. To improve the evaluation of ASLEEP and AWAY setpoints, future research may consider instruments more situated in the context of reporting (Kummerfeld, Tang, Kay, & Yekeh, 2015; Paruthi, Raj, Baek, et al., 2018). This may involve using a situated self-report device (Paruthi, Raj, Baek, et al., 2018) near a doorway or on a nightstand that serves as an ambient reminder as well as an interface to collect comfort feedback when people arrive home or wake up.

5.6.2 Embedding reflection in agent learning

5.6.2.1 A simple probing strategy is adequate for balancing learning and user satisfaction

A challenge we faced during design was the choice of experimentation policy to support mutual-learning. In the end, we chose a simplistic approach; that is, we randomly selected a setpoint from a range configured by the occupants. We believed that this simplistic approach could provide sufficient opportunities to explore more energy-saving setpoints (i.e., support learning), and offer enough time to be at setpoints where occupants are highly likely to be comfortable (i.e., maintain user satisfaction). The parameters we chose allows the system to experiment a setpoint more than five days during its learning period. If users offered multiple comfort feedback a day, then this resulted in more than ten comfort feedback per setpoint, a number that we believe is sufficient to support reflection and agent learning.

We found this simplistic learning policy worked well to support user reflection. It allowed people to identify alternative setpoints that would also work for their households — they did not explore these solutions before due to the lack of tools to encourage experimentation. Our probing approach allowed the system to repeatedly sample a setpoint even if people reported being "a bit too cool" sometimes; at other times, they felt "just right" with these setpoints. We found that this approach was helpful in identifying setpoints which make people feel cold, but bearably cold, an insight that is helpful for the system to suggest short-term setpoint adjustments to save energy (Huang et al., 2017).

Our learning policy also made it likely that participants were comfortable at least half of the time, which may have contributed to the users' general acceptance of the mutual-learning process. Other reinforcement learning policies may further improve the effectiveness of this mutual learning process, for example, an epsilon-greedy policy that reduces the chance of using a setpoint if uncomfortable feedback is detected (Sutton & Barto, 1998). However, more research is still needed to evaluate these policies in terms of their effectiveness on supporting both users and agents to learn optimal settings.

5.6.2.2 Agent training and user learning are reciprocal

We found that agent training facilitates user reflection. That is, knowing that they were training a "smart" thermostat, participants were willing to provide a lot of comfort feedback to the system; this in turn helped with user reflection. If our system was only designed to support reflection as (A. K. Clear et al., 2018), it would be unlikely that it would collect enough feedback from each individual to be able to allow them to determine if an alternative setpoint would work.

On the other hand, we also found that data displayed by the system for reflection also helped the system gain users' trust and motivate their participation in training. This finding provides evidence to support the development of more scrutable (J. Kay & Kummerfeld, 2013), intelligible (Lim, Dey, & Avrahami, 2009) and transparent (Cook & Kay, 1994; Cramer et al., 2008) agents. For example, some participants commented that looking at the visualization gave them a sense of progression in terms of agent's learning. This visualization also helped them confirm that the agent was doing what they expected it to do — to collect, organize, and extract insights from the comfort feedback they provided. Taken together, our findings suggest a reciprocal relationship between displaying learning data for user reflection and gathering data for agent training.

5.6.3 Implications for other smart home applications

While we focused on HVAC as an area to investigate eco-experimentation, there are other application domains that may benefit from eco-coaching that incorporates eco-experimentation and mutual-learning. For example, researchers have investigated the distribution of energy between appliances, in-home batteries, and electric vehicles (EVs) (A. J. Brush et al., 2015). This problem shares a similar structure with our system: the system can experiment with different distribution strategies and it requires users' feedback with regard to impact on their daily living. This approach may also be used in other applications such as smart laundry machines that consider dynamic pricing (Bourgeois et al., 2014). Indeed, there are many other opportunities where mutual-learning can be applied in smart homes. For example, home robots that aim to learn about household routines to support chores can share data related to activity patterns back to users to support reflection on family time use (Dong et al., 2015). We believe more research is needed to explore the effectiveness of these strategies in different problem domains.

5.7 Limitation

Our studies took place primarily in the northern United States in April, while the weather was relatively mild. Despite the weather still being cold enough in the area of the study that participants were still using heaters, reactions to the system may have been different in mid-winter, when outdoor temperatures in our area can reach below 0°F (-18°C). Furthermore, as with many prior deployment studies (e.g., Koehler et al., 2013), the high complexity of system deployment meant that we were only able to conduct our studies in five households with ten participants. As technologies evolve, a smart home system like Reef will be much more inexpensive and less complicated, allowing larger scale deployment to study how people with different attitudes respond to such a system. Although limited by these factors, we believe that our findings show that *eco-experimentation* and *mutual-learning* represent two promising strategies to support eco-coaching that deserve further study.

5.8 Conclusion

In this chapter, we have introduced two strategies for eco-coaching, namely, *eco-experimentation* and *mutual-learning*. We developed ReefSetpoint, a comfort-aware thermostat that can actively experiment different setpoints for different activity contexts (AWAKE, AWAY, ASLEEP) to facilitate both the occupants and the thermostat in learning about user comfort preferences. By deploying the system to five households for 4-5 weeks, we demonstrated that a system leveraging these two eco-coaching strategies can facilitate occupants in identifying setpoints that can reach a better balance of comfort and savings.

Chapter 6. Discussion & Future Work

In Chapter 1, I introduced the vision of wise homes, homes that are augmented with artificial intelligence capabilities to carefully coach people to help them achieve goals such as to live more sustainably and healthier. In such a vision, homes can also leverage the burgeoning of ubiquitous devices to better understand their occupants' goals, preferences and contexts, thus to facilitate their everyday decision making and to automate tasks. To investigate how to design such wise homes, I show that residential heating and cooling control is an appropriate starting point as this application domain shares the same challenges as we advance smart homes toward wise homes: that is, they both encounter challenges in supporting coaching and in sensing contexts. To further advance homes toward this vision, although limited, researchers have started to adopt mixed-initiative interaction in the design of smart home applications. Based on the three studies conducted, this thesis contributes to the design of mixed-initiative interaction in the home setting, providing insights on how to approach our vision of wise homes through better collaboration between humans and homes. However, there are still many unanswered questions that are important to the development of wise homes. In the following paragraphs, I will therefore briefly summarize the findings from each of the studies, discuss the limitations of this work, following by the steps that may be taken to further advance the research on smart homes.

6.1 Improving Comfort Inferences

The study presented in Chapter 3 demonstrates the promising opportunity to leverage the increasingly available wearable and in-home devices to infer individual thermal comfort, thus enabling the development of comfort-aware thermostats. That is, smart thermostats that can take into account people's thermal comfort preferences to facilitate heating and cooling control. However, in this chapter, I also show that there are limitations of this approach due to nuanced human preference and context that are challenging to be detected by a computing system. Smart thermostats that leverage such a comfort sensing technique to support heating and cooling control should therefore use mixed-initiative interaction in its design.

Although there are limitations identified in this chapter that may be challenging to address by a computing system (e.g., detecting if people are sick), there are many places where more advanced sensing and modeling can help. As discussed in Chapter 3, future research can explore better ways to model human thermal comfort, such as using thermal camera or other sensors to infer clothing level (Gao & Keshav, 2013b, 2013a), temperature gradient in a room, and wind velocity. Further, the machine-learning models explored in this Chapter 3 is still quite limited with respect to the consideration of the transition between warmer and cooler conditions (Ugursal, 2010). Future research may explore other sequence models, such as Hidden Markov Models (Juang & Rabiner, 1991) or, if there is enough data, Recurrent Neural Networks to better consider transition (Sak, Senior, & Beaufays, 2014).

6.2 Probing Comfort-aware Eco-coaching Thermostats in the Field

Chapter 4 articulated the design space of comfort-aware thermostats, exploring designs alone three design dimensions: timing of interaction, coaching approaches, and ways to support different decision-making models (norm activation and rational thinking). we created 15 design ideas with high-fidelity interfaces, and probed these ideas through an user enactment study.

Overall, we identify several proactive coaching designs that can further facilitate energy savings or increase comfort. However, we see no one-size-fits-all solution, our participants often gave polarized feedback toward each design idea: while some can imagine living with a comfort-aware thermostat that supports a coaching feature, some participants dislike these features. Among these ideas, the promising features include designing thermostats to more proactively make a decision before informing users, thermostats that offer suggestions of short-term adjustments, and thermostats that can display the impact of different setpoints on multiple household members' comfort level.

While there is no one-size-fits-all solution, through these 15 designs, we obtain a clearer picture on how to balance people's various desires, including comfort, financial savings, energy savings, and desire of control, which inform the design of ReefSetpoint in Chapter 5. However, Reef was studied in a simulated setting, rather than in the wild with a real functioning system. More research is still needed for investigate how these ideas play out in real households, and how they affect comfort, savings and user experience.

6.3 Investigating the Effects of Eco-experimentation and Mutual-learning to Machine-learning-based Comfort Modeling

In Chapter 5, I report findings on the lived experience of ReefSetpoint — a system that implements a subset of features that can be supported by comfort-aware thermostats. In particular, in the design of ReefSetpoint, I explore two interaction strategies to support mixed-initiative interaction and coaching, namely, eco-experimentation and mutual-learning. When the *"guided experimentation"* strategy is used in settings to facilitate energy savings, we call it *eco-experimentation* instead to highlight the energy-saving goal of the experimentation. Chapter 5 shows that eco-experimentation and mutual-learning can help household members identify better temperature setpoints to use to achieve a better balance between their comfort and saving needs.

While I collected a rich amount of qualitative and quantitative data, there are still many questions that I was not able to cover. These questions include: 1) In addition to supporting user learning, do eco-experimentation facilitate artificial agents to better learn about user preferences? 2) How accurate can ReefSetpoint model occupants' comfort preferences? 3) What are the ways to improve learning? In the following paragraphs, I will therefore provide some insights to these questions, based on a preliminary analysis of the

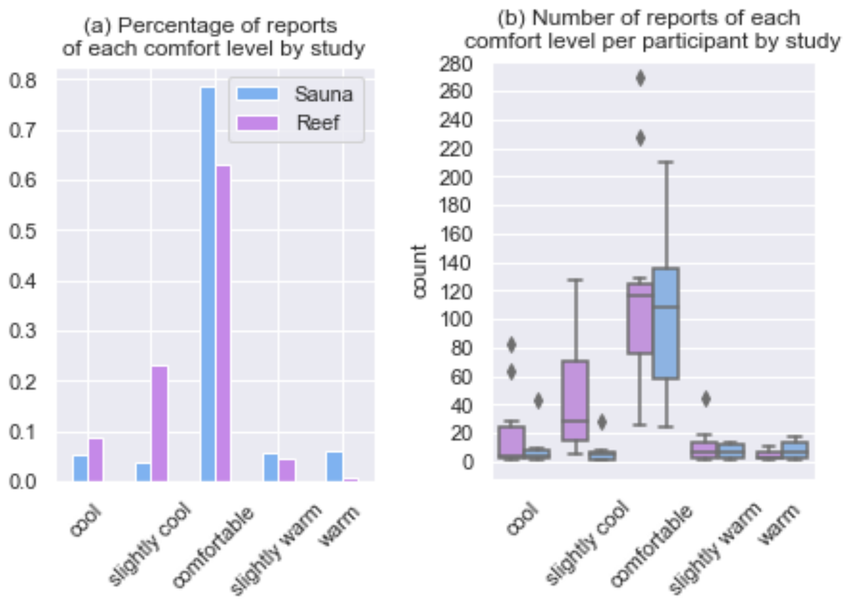


Figure 6: (a) Percentage of reports of each comfort level by study. In ReefSetpoint study we received much less "comfortable" labels and much more "slightly cool" labels. (b) The number of reports of each comfort level by

quantitative data collected in deploying ReefSetpoint.

6.3.1 Experimentation helps create more balanced dataset

To start, based on my preliminary analysis, I do find evidence showing that experimentation benefits the learning of agent as well. For example, I find that there is a significant increase in the number and the percentage of comfort feedback labeled as "a little too cool" and "much too cool" compared to the data collected in Chapter 3 (Figure 6). This effectively addresses the problem I encountered in conducting the study presented in Chapter 3. To construct personalized models, we need sufficient comfort reports covering different classes of labels. However, we collected very few comfort feedback on labels other than being "comfortable" in Chapter 3, as participants often used a fixed setpoint that would make them feel comfortable most of the time. Although collectively we received sufficient data to train population-based models, we were not able to train potentially more accurate personalized comfort models. In addition, this deficiency of data under different temperature conditions may also reduce the generalizability of the comfort

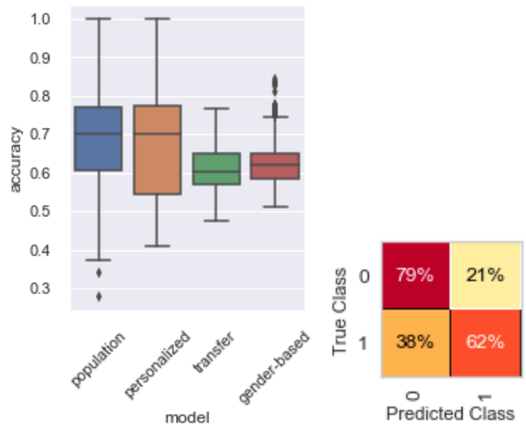


Figure 7: (left) Comparison of different models. (right) Confusion matrix of the population-based model. 1: cool; 0: not cool.

models. By using eco-experimentation, we did receive more diverse labels this time, potentially helping the agent to construct more accurate models.

6.3.2 Accuracy of comfort modeling: population-based vs. personalized models

Based on this more balanced dataset, I also see that ReefSetpoint can achieve 0.98 Mean Absolute Error (MAE) and 1.29 Mean Square Error (MSE) at inferring thermal comfort on a 5-level scale. Unfortunately, this performance is lower than the 0.77 MAE shown in Chapter 3, possibly due to the increased diversity of temperature conditions and labels in this dataset. If we simplify the inference to predicting whether people feel “cool” or “not cool”, we can achieve an accuracy of 70% based on the dataset collected in Chapter 5. For some participants, we can achieve an accuracy of 80% in predicting their comfort (see Figure 7).

By leveraging experimentation and mutual-learning, I am also able to train a personalized model for each participant, thus comparing the accuracy of population-based models and personalized models. As mentioned previously, we were not able to do so through the data collected in Chapter 3, as many participants did not report enough feedback in addition to being “comfortable”. My preliminary analysis shows that population-based models and personalized-models have a similar performance in this dataset. Unsurprisingly, as we obtain more comfort feedback from an individual, the personalized model starts to

Figure 8: The relationship between the numbers of comfort labels and the accuracy of personalized models. We see that for a few participants that reported more than 200 comfort feedback, their personalized models outperform the population-based model by around 2%.

outperform the population-based models (see Figure 8). Nevertheless, the improvement is very marginal in this dataset.

However, adding comfort feedback may not necessarily make a personalized model performs better. Based on my analysis using multi-dimensional scaling, certain people's data are more easily classifiable while others are not (see Figure 9). Perhaps some people just have a more fixed preference, while others are not. Perhaps there are also variables hidden to the system. For example, the accuracy of indoor positioning system can affect how accurate ReefSetpoint can infer people's comfort. While our indoor positioning system shows higher than 90% accuracy when tested on our testing dataset collected during our initial home visits, it needs further evaluation to see if the accuracy holds true once people start to use ReefSetpoint. More research is needed to investigate the effectiveness of personalized models, as well as how to improve such modeling.

6.3.3 Feature importance

Finally, based on the datasets collected in Chapter 3 (Sauna dataset) and Chapter 5 (Reef dataset), I am able to obtain preliminary insights on the importance of different sensors across different seasons. By using tree-based feature selection and recursive feature elimination, my preliminary analysis shows that nearby room temperature and outdoor

Figure 9: Multi-dimensional scaling charts of each participant's comfort reports. "Much too cool" and "a little too cool" reports are marked in blue, "just right" is marked in gray and "a little too warm" and "much too warm" are marked in red. Some participants' data (e.g., H5a and H5b) are much more separable than others (e.g., H2b, H4a).

temperature are always selected as the top features in both Reef and Sauna dataset. I also find that temperature detected by the Nest thermostat (an approximation of living room temperature), skin temperature and outdoor humidity are quite useful in the Reef dataset, while metabolic equivalent of task and galvanic skin response are quite useful in the Sauna dataset. The difference may be a result of seasonal difference, as the Sauna dataset was collected in late fall while the Reef dataset was collected in winter and spring.

6.4 Limitations

While I have discussed the limitations of this thesis in each chapter and in the previous sections, I will briefly summarize them again in this section. In this thesis, I demonstrate four effective strategies to support the development of wise homes, namely, light-weight dialogue, proactive coaching, guided experimentation and mutual-learning. These strategies were investigated through different research methods with different levels of validity. To start, in Chapter 4, I demonstrate that proactive coaching is a promising design strategy to help people further reduce energy consumption: many participants were open to such a strategy, such as a thermostat that asks them to lower their heating setpoint one day per week to save energy. However, this study was conducted in a simulated environment with user enactments. It is still unclear how people would actually live with a thermostat that coaches them proactively. For example, while many participants accepted suggestions to lower their heating setpoints one time per week in the simulated scenario, it is unclear whether they would actually follow the suggestion in an actual deployment. Despite this drawback, user enactments allow us to probe a large number of concepts, helping us identify ideas that are more likely to succeed. Through comparing multiple concepts, we also gain a deeper understanding on the tradeoffs between comfort, savings, convenience, control and agency. This understanding allows us to more carefully design the ReefSetpoint system introduced in Chapter 5.

To address the shortcomings of user enactments, we therefore conducted a field deployment in Chapter 5. Field deployment allows us to investigate *light-weight dialogue*, *guided experimentation* and *mutual-learning* in a fashion that allows more authentic user feedback. However, due to the high cost of system deployment — it took me almost three years to complete the ReefSetpoint system —– I was only able to study a small set of ideas. Many of the promising concepts proposed in Chapter 4 were left out in Chapter 5 due to insufficient time for development. In addition, even though I was able to deploy a complex smart home system to five households for four to five weeks, this study only represents a small population sample and a very limited time frame. It is still unclear how people with different attitudes from different areas would live with ReefSetpoint in different seasons.

Finally, the system created in this thesis still consists of many devices and sensors that are costly (roughly $2000 per household), and they are still not easily deployable by end users without the help of researchers. In the current age, the proposed system will only be applicable for people that are well-resourced, with more technical expertise. However, as technology evolves, I believe that many of the technology components (e.g., indoor positioning) would be dramatically simplified and the cost would be dramatically reduced, therefore lowering the barriers of adoption.

Chapter 7. Conclusion

"*With a bronze mirror, one can see whether he is properly attired; with history as a mirror, one can understand the rise and fall of a nation; with men as a mirror, one can see whether he is right or wrong. Now I've lost my faithful mirror with the death of Wei Zheng" - Emperor Taizong, Tang Dynasty.*

In this concluding chapter, I will revisit the core research question described in chapter 1, namely, how to design a wise home? Through the lens of residential heating and cooling, this question becomes "how to design homes to better coach their occupants to live more sustainably?" In this thesis, I argue that we can leverage mixed-initiative interaction to design such a wise home, and such a wise thermostat. I further show that lightweight dialogue, guided experimentation, mutual-learning and proactive coaching are four effective strategies to support mixed-initiative interaction in homes, helping us to move closer to the vision of a wise home. In the following paragraphs, I will summarize these four design strategies and elaborate how each of the three studies mentioned in this dissertation contributes to the development of these strategies.

7.1 Lightweight dialogue

First, I show that smart homes can leverage lightweight dialogue to better learn about user contexts, preferences and goals. For lightweight dialogue, I refer to the use of ubiquitous interfaces to engage in brief conversation with users, such as to collect feedback on preferences and contexts. In Chapter 3 and Chapter 5, I demonstrate two systems that leverage low-burden mobile and wearable interfaces to collect comfort feedback from occupants, thus helping the systems learn about users' comfort preferences. In addition, I show that micro-interaction enabled by smartwatches offer an even better experience than phone-based dialogue. In sum, lightweight dialogue allows a mixed-initiative system to collect feedback on user preferences and contexts in a high frequency while introducing minimum interaction burden, facilitating the training of system that can better understand people's preferences and contexts.

While in these two studies the use of lightweight dialogue is limited to comfort feedback collection, in Chapter 5 I also highlight the importance of supporting repair if smart homes make an inappropriate decision. These ubiquitous interfaces could thus be used to support other types of lightweight dialogue beyond collecting feedback for personalization, such as remediating errors, and displaying behavior change messages.

7.2 Guided Experimentation

Secondly, I show that a mixed-initiative home can leverage guided experimentation to help occupants identify better solutions to their problems. A home can help participants evaluate different possible solutions by actively exploring different options. In Chapter 5, I demonstrate a system that can actively help household members explore different possible temperature setpoints to use. This study shows that guided experimentation can help household members identify temperature setpoints that would better satisfy their collective needs.

7.3 Mutual-learning

Thirdly, I show that a mixed-initiative home should facilitate occupants to learn about their own preferences and system constraints while the home is learning about their goals, preferences and contexts. I argue that a system which supports mutual-learning can better facilitate occupants to collaborate with the system, and identify better solutions to their problems. In Chapter 5, I show a thermostat system that leverages mutual-learning in its design, displaying data back to users through friendly visualizations while the system is collecting user feedback for training supervised machine-learning models.

7.4 Proactive Coaching

Fourth, I show that smart homes can be more proactive in supporting decision making, such as by actively setting an energy-saving schedule to use before informing users, or lower temperature setpoints for one evening per week to reduce energy consumption. While not everyone accept such proactive system behaviors, many people find them to be acceptable. In Chapter 4, I evaluated many of these proactive coaching ideas through user enactments. The study shows that it is feasible to leverage proactive coaching, however, we will need to allow customization and personalization as the acceptance of these behaviors are influenced by people's attitudes toward sustainability and artificial agents.

7.5 Sensing and System Contributions

In addition to the contribution on design strategies, this dissertation also contributes to the development of thermal comfort sensing. In Chapter 3 and Chapter 6, I demonstrate that we can leverage emerging wearable and in-home sensors with machine-learning techniques to achieve thermal comfort inference with an accuracy of around 70% on a 2-level comfort scale and 0.77-0.98 Mean Absolute Error on a 5-level comfort scale.

Finally, this dissertation also demonstrates a complex system that serves as a foundation for the development of a wise home. In Chapter 5, I demonstrate the feasibility of integrating a variety of sensors, home appliances (e.g., Nest Thermostat) and ubiquitous interfaces to support coaching and context sensing.

Although it took me more than six years to develop the systems and the design strategies highlighted in this dissertation, this thesis only makes a small progress in moving smart homes towards our vision of a wise home. We are still far from this dream. I was hoping that my son Renzo will be able to live with such a wise home when he grows up, however, it seems I still have to wait for a while. It will likely take a few more decades of research before we can live in a house as wise as the minister Wei Zheng.

Appendices

Appendix A:
Initial Interview Protocol in the ReefSetpoint Study

A semi-structured interview will be conducted remotely with each participant after the initial home visit.

Script

Hello, my name is _______, and I am a researcher at the University of Michigan School of Information. Today, I would like to understand how you manage your heating and cooling at home, how you think about energy savings. This interview will take about 45 minutes, and it will be audio-recorded.

Interview Questions (semi-structured)

- How do you usually use your thermostat in the winter?
    - What's the temperature you set when you are at home in the winter now?
    - Are you satisfied with this setpoint?
    - Have you change your thermostat setting this winter? Why?
    - Do you think this temperature setting is higher or lower than other households? Why?
- According to (initial survey), you stated that you (have/have not) tried out many different temperature settings on your previous thermostat, why did you say so?
    - (If tried out many) What have you tried?
    - (If not) Why didn't you explore other possible settings?
- According to your response to the initial survey, you are (satisfied/not satisfied) with the comfort level provided by your previous thermostat, why is that?
- According to your response to the initial survey, you state that you (would/do not) like to reduce energy consumption at home. Why is that?
- (Clothing-level): What do you usually like to wear when staying at home?

  - (If wearing lightly at home in the winter) Have you consider wearing more?
- Do you find certain areas of your house to be warmer or colder than other areas?
  - Do you use that areas in the house?
  - How do you cope with the (warm/cold) when you use that area in the house?
- In addition to the main heater/AC you have, do you use any other portable heaters or coolers? How do you use that?
- **(Multi-user)** Do you feel you and your housemate have similar feeling of warm and cold or do you find it to be different?
  - Do you remember any experience where you feel differently than your housemate at home?
  - If different, how do you cope with the difference? E.g., did you adjust your thermostat settings to satisfy both of your needs?
- **(Attitudes toward heating and cooling as a mean)** Do you think changing your thermostat setting is a useful way to reduce CO2 emission and resource consumption? Why?
- According to your response to the initial survey, you stated that you (would/would not) like to live with a thermostat that automatically controls temperature settings based on your household member's comfort preferences, why is that?
- According to (initial survey), you stated that you (agree/disagree) to live with a thermostat that encourages you to put on sweaters sometimes for the purpose of reducing energy consumption in the winter, why is that?
- According to (initial survey), you stated that you (agree/disagree) to live with a thermostat that encourages you to turn down heating temperature for a few degrees in the winter once a week to increase energy or financial savings, why is that?

Appendix B:
Initial Questionnaire Protocol in the ReefSetpoint Study

Attitudes toward sustainability

Please rate your opinion on the following statements (in terms of strongly disagree, disagree, neither agree nor disagree, agree, strongly agree:

- We are approaching the limit of the number of people the Earth can support.
- Humans have the right to modify the natural environment to suit their needs.
- When humans interfere with nature it often produces disastrous consequences
- Human ingenuity will insure that we do not make the Earth unlivable
- Humans are seriously abusing the environment.
- The Earth has plenty of natural resources if we just learn how to develop them.
- Plants and animals have as much right as humans to exist.
- The balance of nature is strong enough to cope with the impacts of modern industrial nations.
- Despite our special abilities, humans are still subject to the laws of nature.
- The so-called "ecological crisis" facing humankind has been greatly exaggerated.
- The Earth is like a spaceship with very limited room and resources.
- Humans were meant to rule over the rest of nature.
- The balance of nature is very delicate and easily upset.
- Humans will eventually learn enough about how nature works to be able to control it.
- If things continue on their present course, we will soon experience a major ecological catastrophe.

Usage and attitudes toward heating and cooling system

- I spent a lot of effort adjusting the settings of my previous thermostat (refers to the thermostat you had before we install the Nest device.)
- I have tried out many different temperature settings on my previous thermostat.

- I am satisfied with the comfort level provided by my previous thermostat.
- I would like to reduce energy consumption at home to save money.
- I would like to reduce energy consumption at home to reduce consumption of environmental resources and CO_2 emission.

Attitudes toward smart thermostats

- I would like to live with a thermostat that automatically controls temperature settings based on my household members' comfort preferences.
- I would like to live with a thermostat that encourages me to put on sweater sometimes to reduce energy consumption in the winter.
- I would like to live with a thermostat that encourages me to turn down heating temperature a few degrees in the winter once a week to increase energy or financial savings.

Appendix C:
Diary Study Protocol in the ReefSetpoint Study

- Have you ever changed the thermostat settings used by Reef this week?
    - Yes/No/I don't remember
- What happened so that you have to change the setting? Please describe the incidence(s).
- Have you ever feel uncomfortable because of the settings experimented by Reef?
    - Yes/No/I don't remember
- If so, how did you react to the incidence(s)? (E.g., did you override the settings used by Reef? Did you perform other actions to cope with the discomfort?)
- Have you learned something new because of your usage of Reef this week? (E.g., does Reef help you find a different setpoint to use? Does it help you understand your household members' comfort preferences?)
    - Yes/No
- What have you learned from Reef this week? (E.g., does Reef help you find a different setpoint to use? Does it help you understand your household members' comfort preferences?)
- Have you encounter any other issues with Reef this week?
    - Yes/No
- What are the issues? Please provide more detail.
- Is there anything else that you wish Reef to improve in this week?

Appendix D:
Exit Interview Protocol in the ReefSetpoint Study

Overview

After four weeks of Reef deployment, an semi-structured interview will be conducted remotely. During this exit interview, we will gain deeper insights on participants' attitudes toward Reef. This interview will take 60 minutes.

Interview Questions

- How do you feel about when living with a smart home that can actively experiment different temperature settings under your permission?
  - Do you feel comfortable during the learning period?
  - Do you feel that you have control of your heating and cooling system?
- According to our data, in ___% of days during the study period, you have performed some manual override that override the temperature experimented by the system. Could you recall in what situations did you overrode the settings?Do you find it acceptable to live with a system like this?
  - Do you think it is OK for your smart home to go through a learning process like this, in order to learn your comfort preferences?
  - (I want to know, even if they override the settings, do they feel fine with it?)

[Open Reef app on the screen, using participant's account]

- **[Reflect]** Can you describe what you learned from the data provided by Reef in the reflect tab?
- **[Reflect - Housemate preference]** After reading the Reflect tab, do you think you and your household have a similar temperature preference?
  - Is there anything unexpected about your housemate's temperature preference?

- Why it is unexpected?
- After you know about your housemate's preference, how does it impact your daily behavior?

- **[Reflect]** After reading Reef's Reflect tab, what setpoint do you think may work the best for you?
    - I remember you say that you prefer ___, is it the same?
    - (If change) Why you think this one may work better?

[Open the exit interview doc, show the relationship between temperature and comfort]

- **[Reflect]** What about this graph, do you think this works better?
- **[Reflect]** Also, I notice that you previously say you use __ as the setpoint, and even with __ as the setpoint, you feel a little too cool some of the time. Do you think a thermostat should choose a temperature that will make you comfortable all the time, or a temperature that may let you feel slightly cool some of the time. (What's the threshold?)
- **[Reflect]** How do you want Reef to use this knowledge to help you control the temperature at home?
- **[Reflect - Different factors]** Aside from the temperature setpoint of the thermostat, were you able to identify any factors that affect your comforts via the Reflect Tab?
    - How did you learn about it?
    - After you know about the other factors, how does it impact your daily behavior?
- **(Perception of AI)** Are you willing to use the thermostat in the future?
    - Why do you want to use it?
    - What benefit do you expect?
- **Conflict resolution:**
    - Do you prefer Reef to use:
        - A person's preference to control the temperature

        - Find something in the middle
        - Why?
- Review temperature difference in rooms.
    - This is another graph we are thinking about in the future. Do you find anything surprising?
- Review weekly survey
    - I notice you stated that ... , ..

Appendix E:
Exit Questionnaire in the ReefSetpoint Study

- Please look at the "Reflect" tab on the Reef mobile app, what range of temperature do you think will work for your house when occupants are active at home (that is, not sleeping) in the winter? (e.g., 69)
- If Reef uses the data shown in the "Reflect" tab to automatically set a temperature in your house, you do you want it to consider individual differences when both persons are at home?
- Please rate your experience living with Reef on these scales

Incompetent	◯◯◯◯◯	Competent
Ignorant	◯◯◯◯◯	Knowledgeable
Irresponsible	◯◯◯◯◯	Responsible
Unintelligent	◯◯◯◯◯	Intelligent
Foolish	◯◯◯◯◯	Sensible

- Please rate your emotional state while living with Reef on these scales:

Anxious	◯◯◯◯◯	Relaxed
Calm	◯◯◯◯◯	Agitated
Unmoved	◯◯◯◯◯	Surprised

- Please rate your experience living with Reef on these scales:
    - I liked the values of the system
    - The system helped me find out better temperature settings to use

  - The system helped me understand my housemate's comfort preferences
  - Using a system like this would help me live more comfortably
  - Using a system like this would help me reduce energy consumption
- Please rate your experience with Reef on these scales (5-level scale, Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree):
  - I could trust the system
  - The system was assisting me
  - I was assisting the system
  - I agreed with the way tasks were divided between me and the system
  - I would like to live with a system like this (a smart home system that will first undergoes a learning process for a month before activating other features) in the future
- Please rate your experience living with Reef on these scales (5-level scale, Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree):
  - I found the amount of interaction required by the system to be appropriate
  - I was annoyed by the number of notifications sent from Reef

References

Ackerman, M. S. (2000). The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility. *Hum.-Comput. Interact.*, *15*(2), 179–203. https://doi.org/10.1207/S15327051HCI1523_5

Adib, F., Mao, H., Kabelac, Z., Katabi, D., & Miller, R. C. (2015). Smart Homes That Monitor Breathing and Heart Rate. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 837–846. https://doi.org/10.1145/2702123.2702200

AeoTec MultiSensor. (n.d.). Retrieved February 10, 2015, from Aeotec website: http://aeotec.com/z-wave-sensor

Alan, A. T., Shann, M., Costanza, E., Ramchurn, S. D., & Seuken, S. (2016). It is Too Hot: An In-Situ Study of Three Designs for Heating. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5262–5273. https://doi.org/10.1145/2858036.2858222

Allen, J. F., Guinn, C. I., & Horvitz, E. (1999). Mixed-initiative interaction. *IEEE Intelligent Systems and Their Applications*, *14*(5), 14–23. https://doi.org/10.1109/5254.796083

Arroyo, E., Bonanni, L., & Selker, T. (2005). Waterbot: Exploring Feedback and Persuasive Techniques at the Sink. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 631–639. https://doi.org/10.1145/1054972.1055059

ATUS Activity Coding Lexicons. (2012). Retrieved February 10, 2015, from http://www.bls.gov/tus/lexicons.htm

Baker, N., & Standeven, M. (1996). Thermal comfort for free-running buildings. *Energy and Buildings*, *23*(3), 175–182. https://doi.org/10.1016/0378-7788(95)00942-6

Basis B1. (n.d.). Retrieved from http://www.mybasis.com/

Beizaee, A., & Firth, S. K. (2011). *A comparison of calculated and subjective thermal comfort sensation in home and office environment*. Retrieved from https://dspace.lboro.ac.uk/dspace-jspui/handle/2134/14021

Bellotti, V., & Edwards, K. (2001). Intelligibility and Accountability: Human Considerations in Context-aware Systems. *Hum.-Comput. Interact.*, *16*(2), 193–212. https://doi.org/10.1207/S15327051HCI16234_05

Beltran, A., Erickson, V. L., & Cerpa, A. E. (2013). ThermoSense: Occupancy Thermal Based Sensing for HVAC Control. *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*, 11:1–11:8. https://doi.org/10.1145/2528282.2528301

Bentley, F., Tollmar, K., Stephenson, P., Levy, L., Jones, B., Robertson, S., … Wilson, J. (2013). Health Mashups: Presenting statistical patterns between wellbeing data and context in natural language to promote behavior change. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *20*(5), 30.

Bourgeois, J., van der Linden, J., Kortuem, G., Price, B. A., & Rimmer, C. (2014). Conversations with My Washing Machine: An In-the-wild Study of Demand Shifting with Self-generated Energy. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 459–470. https://doi.org/10.1145/2632048.2632106

Brager, G. S., & de Dear, R. J. (1998). Thermal adaptation in the built environment: A literature review. *Energy and Buildings*, *27*(1), 83–96. https://doi.org/10.1016/S0378-7788(97)00053-4

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Braziunas, D., & Boutilier, C. (2008). Elicitation of Factored Utilities. *AI Magazine*, *29*(4), 79–79. https://doi.org/10.1609/aimag.v29i4.2203

Brush, A. J. B., Jung, J., Mahajan, R., & Scott, J. (2012). HomeLab: Shared infrastructure for home technology field studies. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, 1108–1113. https://doi.org/10.1145/2370216.2370450

Brush, A. J. B., Lee, B., Mahajan, R., Agarwal, S., Saroiu, S., & Dixon, C. (2011). Home Automation in the Wild: Challenges and Opportunities. *Computer Human Interaction*, 2115–2124. https://doi.org/10.1145/1978942.1979249

Brush, A. J., Hazas, M., & Albrecht, J. (2018). Smart Homes: Undeniable Reality or Always Just around the Corner? *IEEE Pervasive Computing*, *17*(1), 82–86. https://doi.org/10.1109/MPRV.2018.011591065

Brush, A. J., Krumm, J., Gupta, S., & Patel, S. (2015). EVHomeShifter: Evaluating intelligent techniques for using electrical vehicle batteries to shift when homes draw energy from the grid. *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 1077–1088. Retrieved from http://dl.acm.org/citation.cfm?id=2804274

Brynjarsdottir, H., H\a akansson, M., Pierce, J., Baumer, E., DiSalvo, C., & Sengers, P. (2012). Sustainably unpersuaded: How persuasion narrows our vision of sustainability. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 947–956. Retrieved from http://dl.acm.org/citation.cfm?id=2208539

Chang, Y.-J., Paruthi, G., & Newman, M. (2015). *A Field Study of Comparing Three Crowd-Based Annotation Approaches to Collecting Transportation Data*.

Chaudhuri, T., Zhai, D., Soh, Y. C., Li, H., & Xie, L. (2018). Random forest based thermal comfort prediction from gender-specific physiological parameters using wearable sensing technology. *Energy and Buildings*, *166*, 391–406. https://doi.org/10.1016/j.enbuild.2018.02.035

Chen, L., & Khalil, I. (2011). Activity Recognition: Approaches, Practices and Trends. In L. Chen, C. D. Nugent, J. Biswas, & J. Hoey (Eds.), *Activity Recognition in Pervasive Intelligent Environments* (pp. 1–31). Retrieved from http://link.springer.com/chapter/10.2991/978-94-91216-05-3_1

Clear, A., Friday, A., Hazas, M., & Lord, C. (2014). *Catch my drift?: Achieving comfort more sustainably in conventionally heated buildings*. 1015–1024. https://doi.org/10.1145/2598510.2598529

Clear, A. K., Mitchell Finnigan, S., Olivier, P., & Comber, R. (2018). *ThermoKiosk: Investigating Roles for Digital Surveys of Thermal Experience in Workplace Comfort Management*. 1–12. https://doi.org/10.1145/3173574.3173956

Clear, A. K., Morley, J., Hazas, M., Friday, A., & Bates, O. (2013). Understanding Adaptive Thermal Comfort: New Directions for UbiComp. *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 113–122. https://doi.org/10.1145/2493432.2493451

Cohen, R., Allaby, C., Cumbaa, C., Fitzgerald, M., Ho, K., Hui, B., … Siddiqi, S. (1998). What is Initiative? *User Modeling and User-Adapted Interaction*, *8*(3), 171–214. https://doi.org/10.1023/A:1008398023083

Cook, R., & Kay, J. (1994). The Justified User Model: A Viewable, Explained User Model. *In Proceedings of the Fourth International Conference on User Modeling*, 145–150.

Crabtree, A., Rodden, T., Hemmings, T., & Benford, S. (2003). Finding a Place for UbiComp in the Home. In A. K. Dey, A. Schmidt, & J. F. McCarthy (Eds.), *UbiComp 2003: Ubiquitous Computing* (pp. 208–226). Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-39653-6_17

Cramer, H., Evers, V., Ramlal, S., van Someren, M., Rutledge, L., Stash, N., … Wielinga, B. (2008). The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction*, *18*(5), 455. https://doi.org/10.1007/s11257-008-9051-3

Davidoff, S., Lee, M. K., Dey, A. K., & Zimmerman, J. (2007). Rapidly Exploring Application Design Through Speed Dating. In J. Krumm, G. D. Abowd, A. Seneviratne, & T. Strang (Eds.), *UbiComp 2007: Ubiquitous Computing* (pp. 429–446). https://doi.org/10.1007/978-3-540-74853-3_25

Davidoff, S., Lee, M. K., Yiu, C., Zimmerman, J., & Dey, A. K. (2006). Principles of Smart Home Control. In P. Dourish & A. Friday (Eds.), *UbiComp 2006: Ubiquitous Computing* (pp. 19–34). Retrieved from http://link.springer.com/chapter/10.1007/11853565_2

de Dear, R., & Brager, G. S. (1998). Developing an adaptive model of thermal comfort and preference. *Center for the Built Environment*. Retrieved from https://escholarship.org/uc/item/4qq2p9c6

Dillahunt, T. R., & Mankoff, J. (2014). *Understanding factors of successful engagement around energy consumption between and among households*. 1246–1257. https://doi.org/10.1145/2531602.2531626

Dong, T., Newman, M. W., Ackerman, M. S., & Schoenebeck, S. (2015). *Supporting reflection through play: Field testing the home trivia system*. 251–262. https://doi.org/10.1145/2750858.2804294

Edwards, W. K., & Grinter, R. E. (2001). At Home with Ubiquitous Computing: Seven Challenges. In *Lecture Notes in Computer Science*: *Vol. 2201*. *Ubicomp 2001: Ubiquitous Computing* (pp. 256–272). Retrieved from http://link.springer.com/chapter/10.1007/3-540-45427-6_22

EIA. (2015). Residential Energy Consumption Survey (RECS)—Data—U.S. Energy Information Administration (EIA). Retrieved August 24, 2018, from https://www.eia.gov/consumption/residential/data/2015/

Erickson, T., Li, M., Kim, Y., Deshpande, A., Sahu, S., Chao, T., … Naphade, M. (2013). The Dubuque Electricity Portal: Evaluation of a City-scale Residential Electricity Consumption Feedback System. *Proceedings of the SIGCHI Conference on Human*

Factors in Computing Systems, 1203–1212. https://doi.org/10.1145/2470654.2466155

Erickson, T., Podlaseck, M., Sahu, S., Dai, J. D., Chao, T., & Naphade, M. (2012). The Dubuque Water Portal: Evaluation of the Uptake, Use and Impact of Residential Water Consumption Feedback. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 675–684. https://doi.org/10.1145/2207676.2207772

Fanger, P. O. (1970). *Thermal comfort. Analysis and applications in environmental engineering*. Danish Technical Press.

Faragher, R., & Harle, R. (2015). Location Fingerprinting With Bluetooth Low Energy Beacons. *IEEE Journal on Selected Areas in Communications*, *33*(11), 2418–2428. https://doi.org/10.1109/JSAC.2015.2430281

Feldmeier, M. C. (2009). *Personalized building comfort control* (Massachusetts Institute of Technology). Retrieved from http://dspace.mit.edu/handle/1721.1/55185

Feldmeier, M., & Paradiso, J. A. (2010). Personalized HVAC control system. *2010 Internet of Things (IOT)*, 1–8. https://doi.org/10.1109/IOT.2010.5678444

Ferguson, G., Allen, J., & Miller, B. (1996, January 1). *TRAINS-95: Towards a mixed-initiative planning assistant*. 70–77.

Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *The Journal of Machine Learning Research*, *15*(1), 3133–3181.

Fogg, B. J. (2002). Persuasive Technology: Using Computers to Change What We Think and Do. *Ubiquity*, *2002*(December). https://doi.org/10.1145/764008.763957

Frank, E., & Hall, M. (2001). A Simple Approach to Ordinal Classification. *Proceedings of the 12th European Conference on Machine Learning*, 145–156. Retrieved from http://dl.acm.org/citation.cfm?id=645328.649997

Froehlich, J., Findlater, L., & Landay, J. (2010). The Design of Eco-feedback Technology. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1999–2008. https://doi.org/10.1145/1753326.1753629

Gagge, A. P., Stolwijk, J. A. J., & Hardy, J. D. (1967). Comfort and thermal sensations and associated physiological responses at various ambient temperatures. *Environmental Research*, *1*(1), 1–20. https://doi.org/10.1016/0013-9351(67)90002-3

Gao, P. X., & Keshav, S. (2013a). Optimal Personal Comfort Management Using SPOT+. *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*, 22:1–22:8. https://doi.org/10.1145/2528282.2528297

Gao, P. X., & Keshav, S. (2013b). SPOT: A Smart Personalized Office Thermal Control System. *Proceedings of the Fourth International Conference on Future Energy Systems*, 237–246. https://doi.org/10.1145/2487166.2487193

Goodrich, M. A., & Schultz, A. C. (2008). Human–Robot Interaction: A Survey. *Foundations and Trends® in Human–Computer Interaction*, *1*(3), 203–275. https://doi.org/10.1561/1100000005

Grinter, R. E., Edwards, W. K., Newman, M. W., & Ducheneaut, N. (2005). The Work to Make a Home Network Work. *Proceedings of the Ninth Conference on European Conference on Computer Supported Cooperative Work*, 469–488. Retrieved from http://dl.acm.org/citation.cfm?id=1242029.1242053

Gupta, M., Intille, S. S., & Larson, K. (2009). Adding GPS-Control to Traditional Thermostats: An Exploration of Potential Energy Savings and Design Challenges. In H. Tokuda, M. Beigl, A. Friday, A. J. B. Brush, & Y. Tobe (Eds.), *Proceedings of the 7th International Conference on Pervasive Computing* (pp. 95–114). https://doi.org/10.1007/978-3-642-01516-8_8

Gustafsson, A., & Gyllenswärd, M. (2005). The Power-aware Cord: Energy Awareness Through Ambient Information Display. *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, 1423–1426. https://doi.org/10.1145/1056808.1056932

Halawa, E., & van Hoof, J. (2012). The adaptive approach to thermal comfort: A critical overview. *Energy and Buildings*, *51*, 101–110. https://doi.org/10.1016/j.enbuild.2012.04.011

He, H. A., Greenberg, S., & Huang, E. M. (2010). One Size Does Not Fit All: Applying the Transtheoretical Model to Energy Feedback Technology Design. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 927–936. https://doi.org/10.1145/1753326.1753464

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience Sampling Method: Measuring the Quality of Everyday Life*. SAGE.

Horvitz, E. (1999). Principles of Mixed-initiative User Interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 159–166. https://doi.org/10.1145/302979.303030

Houwelingen, J. H. van, & Raaij, W. F. van. (1989). The Effect of Goal-Setting and Daily Electronic Feedback on In-Home Energy Use. *Journal of Consumer Research*, *16*(1), 98–105.

Huang, C.-C. (Jeff), Liang, S.-Y., Wu, B.-H., & Newman, M. W. (2017). Reef: Exploring the Design Opportunity of Comfort-Aware Eco-Coaching Thermostats. *Proceedings of the 2017 Conference on Designing Interactive Systems*, 191–202. https://doi.org/10.1145/3064663.3064685

Huang, C.-C. (Jeff), Yang, R., & Newman, M. W. (2015). The Potential and Challenges of Inferring Thermal Comfort at Home Using Commodity Sensors. *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 1089–1100. https://doi.org/10.1145/2750858.2805831

Humphreys, M. A., & Fergus Nicol, J. (2002). The validity of ISO-PMV for predicting comfort votes in every-day thermal environments. *Energy and Buildings*, *34*(6), 667–684. https://doi.org/10.1016/S0378-7788(02)00018-X

Humphreys, M. A., Nicol, J. F., & Raja, I. A. (2007). Field studies of indoor thermal comfort and the progress of the adaptive approach. *Advances in Building Energy Research*, *1*(1), 55–88.

Hutton, R. B., Mauser, G. A., Filiatrault, P., & Ahtola, O. T. (1986). Effects of Cost-Related Feedback on Consumer Knowledge and Consumption Behavior: A Field Experimental Approach. *Journal of Consumer Research*, *13*(3), 327–336.

Intille, S., Haynes, C., Maniar, D., Ponnada, A., & Manjourides, J. (2016). *μEMA: Microinteraction-based ecological momentary assessment (EMA) using a smartwatch*. 1124–1128. https://doi.org/10.1145/2971648.2971717

Intille, S.S. (2002). Designing a home of the future. *IEEE Pervasive Computing*, *1*(2), 76–82. https://doi.org/10.1109/MPRV.2002.1012340

Intille, Stephen S., Rondoni, J., Kukla, C., Ancona, I., & Bao, L. (2003). A Context-aware Experience Sampling Tool. *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, 972–973. https://doi.org/10.1145/765891.766101

Jones, B. W. (2002). Capabilities and limitations of thermal models for use in thermal comfort standards. *Energy and Buildings*, *34*(6), 653–659. https://doi.org/10.1016/S0378-7788(02)00016-6

Juang, B. H., & Rabiner, L. R. (1991). Hidden Markov Models for Speech Recognition. *Technometrics*, *33*(3), 251–272. https://doi.org/10.1080/00401706.1991.10484833

Kappel, K., & Grechenig, T. (2009). “Show-me”: Water Consumption at a Glance to Promote Water Conservation in the Shower. *Proceedings of the 4th International Conference on Persuasive Technology*, 26:1–26:6. https://doi.org/10.1145/1541948.1541984

Kay, J., & Kummerfeld, B. (2013). Creating Personalized Systems That People Can Scrutinize and Control: Drivers, Principles and Experience. *ACM Trans. Interact. Intell. Syst.*, *2*(4), 24:1–24:42. https://doi.org/10.1145/2395123.2395129

Kay, M., Choe, E. K., Shepherd, J., Greenstein, B., Watson, N., Consolvo, S., & Kientz, J. A. (2012). Lullaby: A Capture & Access System for Understanding the Sleep Environment. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, 226–234. https://doi.org/10.1145/2370216.2370253

Kidd, C. D., Orr, R., Abowd, G. D., Atkeson, C. G., Essa, I. A., MacIntyre, B., … Newstetter, W. (1999). The Aware Home: A Living Laboratory for Ubiquitous Computing Research. In N. A. Streitz, J. Siegel, V. Hartkopf, & S. Konomi (Eds.), *Cooperative Buildings. Integrating Information, Organizations, and Architecture* (pp. 191–198). Retrieved from http://link.springer.com/chapter/10.1007/10705432_17

Kim, D., & Lim, Y. (2019). Co-Performing Agent: Design for Building User-Agent Partnership in Learning and Adaptive Services. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI ’19*, 1–14. https://doi.org/10.1145/3290605.3300714

Kinect. (2015). In *Wikipedia, the free encyclopedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Kinect&oldid=670483329

Koehler, C., Ziebart, B. D., Mankoff, J., & Dey, A. K. (2013). TherML: Occupancy Prediction for Thermostat Control. *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 103–112. https://doi.org/10.1145/2493432.2493441

Kotsiantis, S., Kanellopoulos, D., & Pintelas, P. (n.d.). *Handling imbalanced datasets: A review*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.96.9248

Kuijer, L., & Giaccardi, E. (2018, April 21). *Co-performance: Conceptualizing the Role of Artificial Agency in the Design of Everyday Life*. 125. https://doi.org/10.1145/3173574.3173699

Kummerfeld, B., Tang, L. M., Kay, J., & Yekeh, F. (2015). *SAL: A small, simple, situated, ambient logger*. 403–406. https://doi.org/10.1145/2800835.2800929

Kuznetsov, S., & Paulos, E. (2010). UpStream: Motivating Water Conservation with Low-cost Water Flow Sensing and Persuasive Displays. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1851–1860. https://doi.org/10.1145/1753326.1753604

Lane, N. D., Xu, Y., Lu, H., Hu, S., Choudhury, T., Campbell, A. T., & Zhao, F. (2014). Community Similarity Networks. *Personal and Ubiquitous Computing*, *18*(2), 355–368. https://doi.org/10.1007/s00779-013-0655-1

Lee, J.-M., Kim, Y., & Welk, G. J. (2014). Validity of consumer-based physical activity monitors. *Medicine and Science in Sports and Exercise*, *46*(9), 1840–1848. https://doi.org/10.1249/MSS.0000000000000287

Leech, J. A., Nelson, W. C., Burnett, R. T., Aaron, S., & Raizenne, M. E. (2002). It's about time: A comparison of Canadian and American time–activity patterns. *Journal of Exposure Analysis and Environmental Epidemiology*, *12*(6), 427–432. https://doi.org/10.1038/sj.jea.7500244

Lim, B. Y., Dey, A. K., & Avrahami, D. (2009). Why and Why Not Explanations Improve the Intelligibility of Context-aware Intelligent Systems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2119–2128. https://doi.org/10.1145/1518701.1519023

Lu, J., Sookoor, T., Srinivasan, V., Gao, G., Holben, B., Stankovic, J., … Whitehouse, K. (2010). The Smart Thermostat: Using Occupancy Sensors to Save Energy in Homes. *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, 211–224. https://doi.org/10.1145/1869983.1870005

Lymberopoulos, D., Liu, J., Yang, X., Choudhury, R. R., Handziski, V., & Sen, S. (2015). A Realistic Evaluation and Comparison of Indoor Location Technologies: Experiences and Lessons Learned. *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*, 178–189. https://doi.org/10.1145/2737095.2737726

Madden, L. (1967). *1999 A.D.* Retrieved from http://www.imdb.com/title/tt6294456/

Matsukawa, I. (2004). The Effects of Information on Residential Demand for Electricity. *The Energy Journal*, *25*(1), 1–17.

McCullough, E. A., Olesen, B. W., & Hong, S. (n.d.). Thermal insulation provided by chairs. *ASHRAE Transactions-American Society of Heating Refrigerating Airconditioning Engin, 100(1), 795–804*.

Megri, A. C., Naqa, I. E., & Haghighat, F. (2005). A Learning Machine Approach for Predicting Thermal Comfort Indices. *International Journal of Ventilation*, *3*(4), 363–376. https://doi.org/10.5555/ijov.2005.3.4.363

Meier, A. (2012). How people actually use thermostats. *ACEEE Summer Study on Energy Efficiency in Buildings. Pacific Grove, Calif.: American Council for an Energy Efficient Economy*. Retrieved from http://escholarship.org/uc/item/3vd5q0cp.pdf

Mennicken, S., Vermeulen, J., & Huang, E. M. (2014). From Today's Augmented Houses to Tomorrow's Smart Homes: New Directions for Home Automation Research. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 105–115. https://doi.org/10.1145/2632048.2636076

Microsoft Band 2. (2019). In *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Microsoft_Band_2&oldid=885865219

Mozer, M. C. (1998). The neural network house: An environment that adapts to its inhabitants. *Proc. AAAI Spring Symp. Intelligent Environments*, 110–114. Retrieved from http://www.aaai.org/Papers/Symposia/Spring/1998/SS-98-02/SS98-02-017.pdf

Mozer, M. C. (2005). Lessons from an Adaptive Home. In D. J. Cook & S. K. Das (Eds.), *Smart Environments* (pp. 271–294). Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/047168659X.ch12/summary

Mozer, M., Vidmar, L., & Dodier, R. H. (1996). The Neurothermostat: Predictive Optimal Control of Residential Heating Systems. *Neural Information Processing Systems*, 953–959.

Mynatt, E. D., Rowan, J., Craighill, S., & Jacobs, A. (2001). Digital Family Portraits: Supporting Peace of Mind for Extended Family Members. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 333–340. https://doi.org/10.1145/365024.365126

Nest. (n.d.). Real Savings. Retrieved September 4, 2019, from Nest website: https://www.nest.com/thermostats/real-savings/

Nicol, J. F., & Humphreys, M. A. (2009). New standards for comfort and energy use in buildings. *Building Research & Information*, *37*(1), 68–73. https://doi.org/10.1080/09613210802611041

Nicol, J. Fergus, & Humphreys, M. A. (2002). Adaptive thermal comfort and sustainable thermal standards for buildings. *Energy and Buildings*, *34*(6), 563–572. https://doi.org/10.1016/S0378-7788(02)00006-3

Nouvel, R., & Alessi, F. (2012). A novel personalized thermal comfort control, responding to user sensation feedbacks. *Building Simulation*, *5*(3), 191–202. https://doi.org/10.1007/s12273-012-0076-5

Odom, W., Zimmerman, J., Davidoff, S., Forlizzi, J., Dey, A. K., & Lee, M. K. (2012). A Fieldwork of the Future with User Enactments. *Proceedings of the Designing Interactive Systems Conference*, 338–347. https://doi.org/10.1145/2317956.2318008

Odom, W., Zimmerman, J., Forlizzi, J., Choi, H., Meier, S., & Park, A. (2014). Unpacking the Thinking and Making Behind a User Enactments Project. *Proceedings of the 2014 Conference on Designing Interactive Systems*, 513–522. https://doi.org/10.1145/2598510.2602960

Paco. (n.d.). Retrieved September 3, 2016, from https://www.pacoapp.com/

Parker, D. (2013). *Determining Appropriate Heating and Cooling Thermostat Set Points for Building Energy Simulations for Residential Buildings in North America*. 21.

Paruthi, G., Raj, S., Baek, S., Wang, C., Huang, C., Chang, Y.-J., & Newman, M. W. (2018). Heed: Exploring the Design of Situated Self-Reporting Devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *2*(3), 132. https://doi.org/10.1145/3264942

Paruthi, G., Raj, S., Colabianchi, N., Klasnja, P., & Newman, M. W. (2018). Finding the Sweet Spot(s): Understanding Context to Support Physical Activity Plans. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, *2*(1), 29:1–29:17. https://doi.org/10.1145/3191761

Peffer, T., Pritoni, M., Meier, A., Aragon, C., & Perry, D. (2011). How people use thermostats in homes: A review. *Building and Environment*, *46*(12), 2529–2541. https://doi.org/10.1016/j.buildenv.2011.06.002

Pierce, J., Fan, C., Lomas, D., Marcu, G., & Paulos, E. (2010, August 16). *Some consideration on the (in)effectiveness of residential energy feedback systems*. 244–247. https://doi.org/10.1145/1858171.1858215

Pierce, J., & Paulos, E. (2012). Beyond Energy Monitors: Interaction, Energy, and Emerging Energy Systems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 665–674. https://doi.org/10.1145/2207676.2207771

Pierce, J., Schiano, D. J., & Paulos, E. (2010). Home, Habits, and Energy: Examining Domestic Interactions and Energy Consumption. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1985–1994. https://doi.org/10.1145/1753326.1753627

Pisharoty, D., Yang, R., Newman, M. W., & Whitehouse, K. (2015). ThermoCoach: Reducing Home Energy Consumption with Personalized Thermostat Recommendations. *Proceedings of the 2nd ACM International Conference on*

Embedded Systems for Energy-Efficient Built Environments, 201–210. https://doi.org/10.1145/2821650.2821671

Plourde, A. (2003). Programmable thermostats as means of generating energy savings: Some pros and cons. *Canadian Building Energy End-Use Data and Analysis Centre, Technical Report CBEEDAC 2003-RP*, *1*. Retrieved from http://www.ualberta.ca/~cbeedac/publications/documents/progtherm1_000.pdf

Pommeranz, A., Broekens, J., Wiggers, P., Brinkman, W.-P., & Jonker, C. M. (2012). Designing interfaces for explicit preference elicitation: A user-centered investigation of preference representation and elicitation process. *User Modeling and User-Adapted Interaction*, *22*(4), 357–397. https://doi.org/10.1007/s11257-011-9116-6

Pousman, Z., Stasko, J. T., & Mateas, M. (2007). Casual information visualization: Depictions of data in everyday life. *Visualization and Computer Graphics, IEEE Transactions On*, *13*(6), 1145–1152. https://doi.org/10.1109/TVCG.2007.70541

Pu, P., & Chen, L. (2008). User-Involved Preference Elicitation for Product Search and Recommender Systems. *AI Magazine*, *29*(4), 93–93. https://doi.org/10.1609/aimag.v29i4.2200

Ranjan, J., & Scott, J. (2016). *ThermalSense: Determining dynamic thermal comfort preferences using thermographic imaging*. 1212–1222. https://doi.org/10.1145/2971648.2971659

Rodden, T. A., Fischer, J. E., Pantidi, N., Bachour, K., & Moran, S. (2013). At Home with Agents: Exploring Attitudes Towards Future Smart Energy Infrastructures. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1173–1182. https://doi.org/10.1145/2470654.2466152

Rogers, Y. (2006). Moving on from Weiser's Vision of Calm Computing: Engaging UbiComp Experiences. In P. Dourish & A. Friday (Eds.), *UbiComp 2006: Ubiquitous Computing* (pp. 404–421). Retrieved from http://link.springer.com/chapter/10.1007/11853565_24

Sak, H., Senior, A., & Beaufays, F. (2014). Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition. *ArXiv:1402.1128 [Cs, Stat]*. Retrieved from http://arxiv.org/abs/1402.1128

Scott, J., Bernheim Brush, A. J., Krumm, J., Meyers, B., Hazas, M., Hodges, S., & Villar, N. (2011). PreHeat: Controlling Home Heating Using Occupancy Prediction. *Proceedings of the 13th International Conference on Ubiquitous Computing*, 281–290. https://doi.org/10.1145/2030112.2030151

Seligman, C., & Darley, J. M. (1977). Feedback as a means of decreasing residential energy consumption. *Journal of Applied Psychology*, *62*(4), 363–368. https://doi.org/10.1037/0021-9010.62.4.363

Sheridan, T. B., & Verplank, W. L. (1978). *Human and Computer Control of Undersea Teleoperators*. Retrieved from MASSACHUSETTS INST OF TECH CAMBRIDGE MAN-MACHINE SYSTEMS LAB website: https://apps.dtic.mil/docs/citations/ADA057655

Shove, E. (2003). *Comfort, Cleanliness and Convenience: The Social Organization of Normality*. Retrieved from http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10060563

Snow, S., Auffenberg, F., & Schraefel, M. C. (2017, May 2). *Log it While it's Hot: Designing Human Interaction with Smart Thermostats for Shared Work Environments*. 1595–1606. https://doi.org/10.1145/3025453.3025578

Sonderegger, R. C. (1978). Movers and stayers: The resident's contribution to variation across houses in energy consumption for space heating. *Energy and Buildings*, *1*(3), 313–324. https://doi.org/10.1016/0378-7788(78)90011-7

Sookoor, T., & Whitehouse, K. (2013). RoomZoner: Occupancy-based Room-level Zoning of a Centralized HVAC System. *Proceedings of the ACM/IEEE 4th International Conference on Cyber-Physical Systems*, 209–218. https://doi.org/10.1145/2502524.2502553

Srinivasan, V., & Takayama, L. (2016). Help Me Please: Robot Politeness Strategies for Soliciting Help From Humans. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 4945–4955. https://doi.org/10.1145/2858036.2858217

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. A Bradford Book.

The Nest Thermostat. (n.d.). Retrieved from https://nest.com/thermostat/

Thorshauge, J. (1982). Air-velocity fluctuations in the occupied zone of ventilated spaces. *ASHRAE Transactions*, *88*, 753–764.

Total Energy. (n.d.). Retrieved from http://www.eia.gov/totalenergy/

Turner, L. D., Allen, S. M., & Whitaker, R. M. (2015). Interruptibility Prediction for Ubiquitous Systems: Conventions and New Directions from a Growing Field. *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 801–812. https://doi.org/10.1145/2750858.2807514

U. S. Census Bureau, American Community Survey, 5-Year Estimates. (n.d.). Retrieved from http://quickfacts.census.gov/qfd/meta/long_HSD310213.htm

Ugursal, A. (2010). *Thermal comfort under transient metabolic and dynamic localized airflow conditions combined with neutral and warm ambient temperatures*. Retrieved from http://gradworks.umi.com/34/46/3446576.html

U.S. Energy Information Administration (EIA)—Total Energy Monthly Data. (n.d.). Retrieved September 18, 2019, from https://www.eia.gov/totalenergy/data/monthly/

US EPA, O. (n.d.). Clean Power Plan [Collections and Lists]. Retrieved September 3, 2016, from https://www.epa.gov/cleanpowerplan

Vu, T. T., Song, D., Willis, A., Tran, S. N., & Li, J. (2014). *Improving search personalisation with dynamic group formation*. 951–954. https://doi.org/10.1145/2600428.2609482

Walker, M., & Whittaker, S. (1990). Mixed Initiative in Dialogue: An Investigation into Discourse Segmentation. *Proceedings of the 28th Annual Meeting on Association for Computational Linguistics*, 70–78. https://doi.org/10.3115/981823.981833

Winett, R. A., & Kagel, J. H. (1984). Effects of Information Presentation Format on Resource Use in Field Studies. *Journal of Consumer Research*, *11*(2), 655–667.

Yang, R., & Newman, M. W. (2013). Learning from a Learning Thermostat: Lessons for Intelligent Systems for the Home. *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 93–102. https://doi.org/10.1145/2493432.2493489

Yang, R., Newman, M. W., & Forlizzi, J. (2014). Making Sustainability Sustainable: Challenges in the Design of Eco-Interaction Technologies. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 823–832. https://doi.org/10.1145/2556288.2557380

Yang, R., Pisharoty, D., Montazeri, S., Whitehouse, K., & Newman, M. W. (2016). How Does Eco-coaching Help to Save Energy? Assessing a Recommendation System for Energy-efficient Thermostat Scheduling. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 1176–1187. https://doi.org/10.1145/2971648.2971698