Exploring the Design Space of User-System Communication for Smart-home Routine Assistants

Yi-Shyuan Chiang*
National Tsing Hua University
Hsinchu, Taiwan
yschiang@gapp.nthu.edu.tw

Ruei-Che Chang*
National Tsing Hua University
Hsinchu, Taiwan
Ruei-che.Chang.GR@dartmouth.edu

Yi-Lin Chuang
National Tsing Hua University
Hsinchu, Taiwan
irenechunag@gmail.com

Shih-Ya Chou
National Tsing Hua University
Hsinchu, Taiwan
108462010@g.nccu.edu.tw

Hao-Ping Lee
National Chiao Tung University
Hsinchu, Taiwan
dimension4.cs03@nctu.edu.tw

I-Ju Lin
National Tsing Hua University
Hsinchu, Taiwan
liniju@gapp.nthu.edu.tw

Jian-Hua Jiang Chen
National Chiao Tung University
Hsinchu, Taiwan
novel860329.cs08g@nctu.edu.tw

Yung-Ju Chang
National Chiao Tung University
Hsinchu, Taiwan
armuro@cs.nctu.edu.tw

ABSTRACT

AI-enabled smart-home agents that automate household routines are increasingly viable, but the design space of how and what such systems should communicate with users remains underexplored. Through a user-enactment study, we identified various interpretations of toward such a system's confidence in its automated acts. That confidence and their own mental models influenced what and how the participants wanted the system to communicate, and how they would assess, diagnose, and subsequently improve it. Automated acts resulted from false predictions were not generally considered improper, provided that they were perceived as reasonable or potentially useful. The participants' improvement strategies were of four general types. Factors affecting their preferred levels of involvement in automated acts and their interest in system confidence were also identified. We conclude by making design recommendations for the user-system communication design spaces of smart-home routine assistants.

Author Keywords

Smart-Home; routine assistant; intelligent agent; user enactment.

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
http://dx.doi.org/10.1145/3313831.3376501

CCS Concepts

•Human-centered computing → Empirical studies in HCI; Ubiquitous computing; •Computing methodologies → *Intelligent agents;*

*The two lead authors contributed equally to the work.

INTRODUCTION

The smart-home market is expected to witness double-digit growth in 2019, and 29 billion connected devices are forecast to be shipped by 2022 [17, 10]. Thanks to the rapid advancement of the Internet of Things (IoT) and machine learning, households can now preset their everyday routines with intelligent assistants [23], we may soon see prediction-based smart homes, which will leverage information collected from IoT devices to learn our intentions and then automate household routines based on our behaviors [54, 36]. Such prediction-based systems would require effective communication, not only because constant user feedback is crucial to making their predictions more accurate[28], but also because users who lack knowledge of machine learning are likely to have unrealistic expectations of what such systems can do and how long their learning phases will take [15]. This could lead them to stop utilizing such systems' predictive functions, or at worst, to abandon system use altogether [65, 47]. However, the design space of user-system communication for smart-home routine assistants (SHRAs) remains underexplored. We define SHRAs as "agents that predict residents' intended actions on connected home appliances based on their former daily routines/behaviors and automatically perform those actions for them in households." This paper is intended to study what users want from such communication and how systems should communicate with them, focusing on SHRA that are designed to automate households' routines based on predictions of their members intention, and proceeding from the assumption that effective communication is key to SHRA adoption.

Our two principal research questions are: 1) What do household members want to know about their SHRAs' automated actions when communicating with them? 2) How do household members want SHRAs to communicate such content with them? Additionally, in light of the fact that frictions caused by false predictions appeared to be unavoidable in an earlier stage of smart-home rollout, we added a third research question: 3) How do household members perceive and assess incorrect automated acts caused by false prediction, and how would they choose to correct them and/or change SHRAs to make such errors less common?

We adopted user enactment [52, 53, 68] to explore these aspects of user-system communication. User enactment allows researchers to understand potential users' behaviors through observing how they interact with study objects. In this exploration, we adopted the term *confidence* in the user-system communication content. Confidence, defined as "faith or belief that one will act in a right, proper, or effective way" [44], does not apply only to humans in this context. Several well-known AI-assisted products, including those focused on understanding language [39] and on activity recognition [18], have utilized the notion of confidence in their prediction outcomes.

This paper's main contributions are as follows. First, we show that SHRA' own levels of confidence and their users' mental models both affect how the users would assess, diagnose, and would seek to improve SHRAs. Second, it identifies four main strategies users adopt to improve their SHRAs' prediction accuracy, including configuring, demonstrating, simplifying, and complying, along with factors affecting their preferred levels of involvement in automated acts, and their interest in knowing about their systems' confidence levels. Third, we suggest four design spaces of user-SHRA communication, including onboarding, routine prediction and automation, occasional tips and quick facts, and configuration panels.

RELATED WORKS

To understand user-smart home communication better, we investigated related bodies of work including research on smart homes in general, on the importance of communication, and on how users interact with intelligent systems.

Smart homes, appliances and management systems

Smart homes have been referred to as "an application of ubiquitous or pervasive computing" [46] covering ventilation, air conditioning, lighting and home-security systems, among others [63]. Researchers have focused on single devices such as thermostats [1, 24, 65] or on general management systems [55]. With conversational agents gaining in popularity, a growing body of literature has studied the usage of Alexa in households [57]. Users' routines, privacy preferences, and relationship types, as well as long-term changes such as the arrival or departure of original occupants, have been considered crucial to the design of smart homes [12, 19, 66]. Previous studies have also discussed the adoption and appropriation of intelligent systems by households [64] and how to design proactive or reactive services for inhabitants that are tailored to their expectations [7, 25, 41]. Multiple interface designs for smart-home management have been explored [9, 5, 43], leading some scholars to urge the adoption of central control systems as essential to the integration of multiple agents [67]. And, in addition to the technology itself, the concept of digital housekeeping has been introduced as a framework for understanding how smart homes are maintained by their respective households [21, 26, 56].

Communication as Fostering Trust and Understanding

Trust is key to successful user-system collaboration. As demonstrated by research in the field of human-robotic interaction, users are more willing to use automated systems when they have more trust in them [61]. In household settings, trust is even more critical, due to homes' intimate and complex character, and trustworthiness has been ranked at the top of lists of desired qualities for smart homes [42, 43]. Conversely, unreliable automation can lead to frustration among users [7]. To foster trust, it has been argued that systems' communication with their users should include rationales for how they have been designed, and/or explanations for their specific actions [2, 58, 61, 62]. As well as building trust, such explanations could render users' understanding of such systems more realistic, which in turn could enhance the overall quality of their interactions with them [8]. In terms of specific communication techniques, "confidence" discourse has been proposed as a key means of improving the communicative effectiveness of system outputs [20]. However, "confidence" in this context has been assigned two divergent meanings: either a system's level of certainty that its own actions are appropriate [40, 61], or users' confidence that the outcomes of system actions will be desirable ones [16].

Mental Models and Expectations of Intelligent Systems

Past studies of users' mental models of intelligent systems have reported that such models can be shaped both by users' technical backgrounds and by their previous experiences [35, 50]. Nevertheless, without any assistance, prior research participants have been able to ascribe basic machine-learning concepts to intelligent systems [59], and to form plausible – albeit fuzzy – mental models of them quickly [8, 29]. This fuzziness could be addressed via direct explanations of system behavior [28], though when dealing with lightweight intelligent systems, such an approach might not be as useful as in relatively complex ones [58]. Because mental models are gradually formulated from subjective experience and knowledge, which *ipso facto* vary across individuals, it is reasonable to assume that there are numerous possible mental models of any given system. Kim and Lim, for example, identified two main mental models of interaction with intelligent agents, namely the *Getting-Things-Done (GTD) Agent* model and the *Companion Agent* model [27]. This categorization reflects that some individuals long for control over these systems, whereas others are more willing to cooperate with them [6, 27]. Jensen et al., meanwhile, identified three personas for smart homes – the helper, the optimizer and the hedonist [25] – each with its

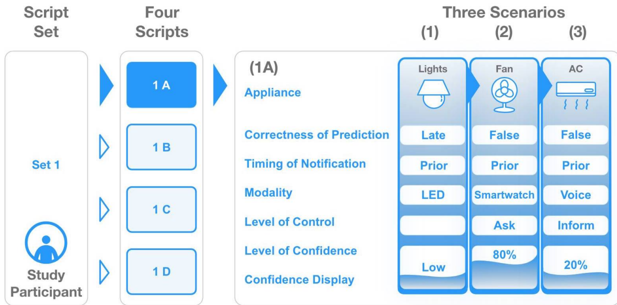

Figure 1. Sample combinations of factors in our user-enactment scripts

own distinctive combination of three desired characteristics, control, automation and uniqueness.

Users' Tactics When Systems Err

Myers et al. found that the tactics people adopted when facing obstacles to their interaction with voice-user interfaces generally included articulation, simplification, or adding more information to their utterances. These tactics were sometimes accompanied by frustration or led to task abandonment [47]. Lahoual and Frejus noted that users of Alexa would not only rectify their language but engage the device physically, such as by gazing at or approaching it, if hyper-articulation failed. Interestingly, those who persisted in their attempts to communicate were more likely than those who abandoned them to forgive the system and accept its errors [30].

Despite the abundance of prior research on smart homes, however, there have been few explorations of either how an intelligent agent that automates routines for users should communicate about its own behavior, or specifically *what* it should communicate. We aim to fill this research gap, with the wider aim of improving smart-home design.

METHODOLOGY

Participants

The 20 participants (10 females, 10 males) were found via a Facebook group dedicated to recruiting participants for HCI studies in Taiwan. All were aged between 20 and 33. Four had studied engineering-related subjects at university, and two of those had prior experience of smart-home devices such as Google Home because they worked in related fields. However, the majority were not tech-savvy. Further particulars of all participants' background can be found in Appendix 1.

User Enactments

Since SHRAs are still at the forefront of technology and have limited consumer visibility, we utilized a technique called user enactment to identify households' patterns of and preferences for interacting with such system [68]. Specifically, in user-enactment participants undergo a series of structured engagements with a given technology that vary in multiple dimensions that the researchers are interested in. This facilitates exploration of potential users' values and expectations regarding that technology, via both observing and directly querying their responses and reflections on different designs and settings, thereby establishing possible directions or concepts for future iterations [52, 53].

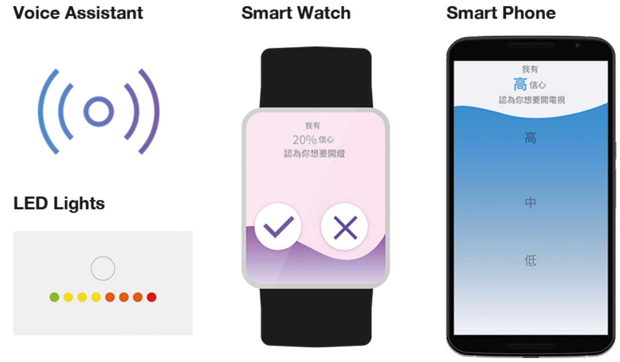

Figure 2. The four user interfaces used in this study. The message on the smartwatch reads, "I have 20% confidence that you want to turn on the lights", and the one on the smartphone says, "I have high confidence that you want to turn on the TV". The former was printed on cardboard while the later was an interactive prototype

Participants were asked to imagine that they were in a smart home where SHRA could automatically turn on appliances according to its predictions of their intentions. We designed three sets of scripts, each containing four scripts featuring three home-automation scenarios apiece (i.e., 36 scenarios in all). Each participant was assigned randomly to enact the 12 scenarios in one script set. Each scenario varied according to three primary and four secondary dimensions. The former were *home appliances*, *timeliness of automated actions*, and *notification timing*, and the latter, *communication interface*, *level of user involvement*, *level of confidence*, and *confidence-presentation format* (see Figure 1). Due to the existing literature's inattention to how SHRA should be explained to its users, we chose confidence as our communication prompt to elicit the participants' thoughts and desires for SHRAs.

The home appliances in the scenarios included a television (TV), an air conditioner (AC), a light, and a fan, which are popular devices for smart homes to control [57, 60]. Since light is fundamental in modern households, it was selected as the default appliance that participants were required to enact, resulting in three appliance combinations: 1) light, AC and fan, 2) light, TV and fan, and 3) light, AC and TV. Each script focused on one combination. The timeliness of automated actions had three facets: false, early, and late. The timings of notification also had three facets: prior to the automated action, after the automated action, and no notification. Thus, there were nine combinations of timeliness and timing (3 x 3) in any given set of 12 scenarios. The remaining three were filled with false-prior, false-after, and late-no, which we were interested in, as eliciting participants' reactions to SHRAs making mistakes or delaying automated actions without any notification. We randomized the order of these 12 combinations.

Regarding the secondary factors, the communication interface through which the SHRA communicated with the participant included a smartphone, a smartwatch, a voice-user interface (VUI), and ambient light (represented via LED light strings in front of home appliances), which have been explored in previous smart home studies [37, 22, 32] (See Figure 2). User involvement referred to the extent to which the participant was involved in the SHRA's three types of automated actions:

asking permission to perform an action; notifying the user that an action had occurred; and not providing any such information. To test confidence, we classified it into two levels, low and high, representing how sure the SHRA was about its prediction of the participant's intention to turn on a particular appliance. The above-mentioned dimensions were randomly and evenly assigned across all 36 scenarios. We randomly assigned a presentation format (categorical or numeric) to each confidence level, i.e., 80% or high confidence, and 20% or low confidence. This was done because LED strings could not display characters, and thus were incapable of displaying confidence levels numerically. Also, it should be noted the LED lights could not ask for, or inform users about the decisions made by SHRAs; thus, levels of user involvement were not tested for this interface.

Study Procedure

Participants were invited to our laboratory, which was set up to resemble a living room, with a set of sofas, a bookshelf, several home appliances (including the four target ones), and a panoramic camera attached to the ceiling, as adopted in prior research [38] (Figure 3). A member of the research team serving as a moderator explained the study procedure and introduced the SHRA configuration, including that its behavior-detection and intention-recognition efforts were based on what the panoramic camera captured. Each participant was then given the scripts for his/her enactment session, which was designed to last around 90 minutes.

We adopted the Wizard of Oz technique [11] to stimulate automation in each scenario. At the start of each script, the moderator explained its background and then ran the three scenarios in a predetermined order. Within each scenario, the participant was told the context and the appliance he/she intended to turn on, such as waking up and intending to turn on the light. The contextual factors included *busy*, *lazy*, *tired*, *energetic* in combination with *waking up*, *going to work*, *off work* and *on holiday*. The moderator asked the participant to imagine him- or herself in that scenario and enact it. According to the assigned combination of home appliance, the timeliness of automated act, and the notification timing, the wizard and the moderator determined the timing of both when the interface was shown and when the home appliance was controlled. Participants were presented with SHRA notifications such as "I have 20% confidence that you want to turn on the AC" or "I have high confidence that you want to turn on the lights"(for a full list, see Appendix 2 in supplementary materials). The participants were asked to use a think-aloud approach to describe their experiences and feelings about automated actions. Then, after the enactment was finished, the moderator immediately used a semi-structured interview approach to capture the participants' feelings regarding the appropriateness of the automated actions; the SHRA's communication style and content; and how, if at all, they would correct or improve the system (for the questions, see Appendix 3 in supplementary materials). In addition to the initially assigned interface, the moderator provided participants successively with other interfaces and/or designs of the same interface, and asked them to explain which they preferred and why. Upon the completion of all 12 scenarios, each participant was paid a cash honorarium of NTD 500 (USD 16).

Figure 3. The laboratory set up as a living space

Qualitative Analysis

For qualitative analysis, we used affinity diagramming [34]. After transcribing audio recordings of the interviews into affinity notes, we convened five organizing sessions to group and label them iteratively. Between labeling sessions, the researchers discussed notes they were unsure about, walked the affinity wall together and regrouped whenever necessary. This process resulted in several high-level themes that will be presented in the following section.

FINDINGS

Our findings fall into four main categories: 1) factors affecting participants' desire to be involved with the system's automated actions; 2) communication content, focusing on the participants' interpretations of confidence, factors affecting their interest in knowing about confidence, and other content they desired to know about when communicating with the SHRA; 3) how the level of confidence and the participants' own mental models affected their assessment and diagnosis of automated actions, and how they said they would improve the system's predictions; and 4) how the participants personified the SHRA and what they expected of it.

Should SHRA Ask for Rermission, or Only Notify?

How much a system should involve users in its decisions and actions has been a recurring theme in work on intelligent systems [3, 7, 42, 51]. In the present SHRA context, we found that participants generally wanted to stay on top of the system's automated actions, and wanted to be asked for permission or notified, rather than having the SHRA execute such actions "silently" or "secretly." As P8 commented, *"It just turned it on without showing anything. It just suddenly did it. That's it? Nothing else? I didn't even know that 'beep' sound was from the AC."* Other participants, including P20, commented that automated actions without notifications frightened them, and considered it especially inappropriate when such actions were not what they intended. However, whether participants preferred the SHRA to ask for permission, or simply to notify them about automated actions, was dependent on several factors that will be discussed below.

Perceived Possibility of An Action Being Wrong

The most commonly mentioned factor was the participants' subjective assessment of how likely an automated action was to be incorrect. System confidence provided an important signal of such a possibility. That is, most of the participants wanted the SHRA to automate without asking when it had high confidence in its prediction, but preferred it to ask for permission if it had low confidence. As P10 commented, *"If it has high confidence in its action, based on an analysis of my past behaviors, then it should do it without asking. Low confidence means it is not sure, and it should ask me."* And conversely, some participants felt that it would be redundant for the SHRA to ask for permission despite having high confidence. However, if they knew the SHRA was "still learning" a specific pattern, the participants preferred it to ask for permissions. And, regardless of confidence, some participants doubted that an SHRA could ever accurately recognize their specific intentions, from among the wide array of possibilities. As P6 commented, *"No automation for this. When I sit here I could do other things [than watching TV], like use my laptop, do some reading, and so on. It's very likely to judge it wrong. [...] [P]eople have their own habits"*. Similarly, P3 commented, *"It's too hard to predict one's mind. So, I'd go with it asking me what I want to do when I raise my hand and asking me another thing when I raise my leg."*

Desired Level of User Control: Be Respectful

The second important factor affecting the participants' permission-vs.-notification preferences was their desired level of user control. Despite most appreciating the convenience of automated actions in theory, a few disliked it in practice. As P7 complained, the SHRA *"just turned on the TV without asking me, [and] I found it quite disrespectful"*. P16 said, *"if it turned on the TV for me and then told me that it did it [without asking], that is assertive."* Some, like P11, told us that they *"preferred to have certain control instead of machines making every decision."* And P20 said, *"I'm more used to me telling it what to do. I'd be scared if its voice suddenly came out."*

Cost of Automated Acts: Automate When The Cost is Low

Whether the participants felt the SHRA should ask for permission also depended on the perceived cost of its actions: with these being deemed especially inappropriate for specific types of home appliances and in risky situations (*"systems should ask before turning on gas or something dangerous"* [P16]). Less serious, but still troublesome, was manually turning off appliances that had been turned on due to incorrect prediction: *"If the system turns on the wrong device, I have to go there just to shut it down"* (P12). The SHRA asking permission could not only prevent such trouble, but also give users the opportunity to coordinate while answering, as P6 explained: *"If the system told me it was turning the AC on for me, then I could go and close the window"*. On the other hand, if the cost of the incorrect action was seen as low, participants were more open to SHRA not asking but just giving it a try: *"Though it was not what I intended to do, I am heading to that direction [where the switch is at] and could just turn it off by the way. So I think it's OK"* (P16).

Other Contextual Factors

Participants also mentioned that contextual factors such as their own whereabouts and emotional states, the physical environment, and the presence of other people could influence their permission-vs.-notification preferences. As P3 noted, *"It depends on my mood, if I want to relax during the weekend, I would like the system to do it for me."* P4 said, *"It is about the weather. If it was hot outside, it would be nice to have it automatically turn on AC for me, but if it were cold, it would be annoying."* He also preferred the SHRA to ask for permission when kids were present, instead of acting on its own, *"If kids are also in the living room, when the TV is suddenly turned on, they might get distracted and abandon their original reading plan, and go watch TV instead"*.

Interpretation/Perception of and Interest in Confidence

Interpreted Meanings and Feelings About Confidence

When prompted to describe what confidence meant, the participants used an array of broadly similar terms such as *probability*, *possibility*, *likelihood*, *accuracy*, *certainty*, and *strength of association*. Other interesting interpretations included *"how sensitive the system is"* (P6), *"what the system knows/thinks I want to do"* (P13, P18), *"prediction of my own confidence in doing it"* (P10), and *"a kind of desire or urge to do something"* (P19). Despite positive comments about how confidence boosted their awareness of the system's status, the participants did report some negative feelings that we did not expect, including an impression of the system's incompetence, and being reminded that the SHRA was only a machine. For example, P16 told us the confidence model made *"the system seems like it is still under development"*, because he expected *"an actual product to be fine-tuned already [...] it wouldn't need to be supported by showcasing confidence."* P9 said that the confidence elements of user-SHRA communication made it seem *"very machine-to-human, instead of human-to-human. If an action is right, it is right; there is no 'confidence' in human interaction."* Also unexpectedly, P19 noted that confidence implied a sense of reluctance: *"it sounds a little bit involuntary. It's like a secretary who takes my input and does things for me but is unwilling"*. And P8 provided another interesting perception when seeing 20% confidence, *"At first sight, I felt it was low. But then I was thinking, is it actually telling me that it has high confidence? When you have one side of 20%, you have the other side being 80%."*

How Much Did Participants Want to Know About Confidence?

The participants did not always think confidence information was necessary. The four main factors that influenced their desire to learn about confidence are discussed in turn below.

Curiosity about the system. Simple curiosity about the SHRA's confidence had strong impact on whether the participants wished to be told about it. Curiosity operated on a spectrum, with a few participants uninterested in knowing about the system's confidence at any point in the user enactment, and at the other extreme, several wanting it to be stated for all automated actions. Curiosity about the system, in turn, was influenced by their level of understanding of how SHRAs work. Participants who believed that the SHRA's acts could only be right or wrong, with no room in between, often had low curiosity about confidence, since they only cared what was being done instead of why it was being done: P14, for example, said *"as long as it turns [the appliance] on for me, it is enough, I don't care much about how confident it is."* Similarly, P9 commented, *"to* *achieve my goal [turning on specific appliances], I don't need it to say what it has predicted"*. On the other hand, those who understood that intelligent systems were not as definitive were more curious about confidence regardless of the correctness of the automated actions, since they would like to know what went into the automation decisions. P12 said, *"I feel better if I know the system at least is confident in what it does. If it does not show confidence, I would wonder why it even automates it in the first place."*

Perceived opportunities to interfere with automated acts. Between the two ends of the above-mentioned spectrum were participants who cared about confidence only in certain circumstances. Mostly, this depended on whether they were allowed to modify the SHRA's decision in the moment, such as by correcting its appliance selection or stopping its action altogether. P1 noted, *"If the system had already done it, it was unnecessary to know its confidence level."* Others shared this view that confidence information was unhelpful in cases where they could not intervene in the SHRA's actions.

The correctness of automated actions. Participants' desire to know confidence information was also influenced by whether they perceived a given SHRA prediction as correct. If they saw it as incorrect, they often expressed interest in why the SHRA had been confident in its action, to help them make sense of the system and how best to use it.

Participants' availability in the moment. Finally, the participants' interest in confidence information was impacted by their own perceived availability, physically or mentally. When participants were in a rush or simply 'not in the mood,' they were less interested. As P15 said, *"I don't want to learn about confidence now, since I just got off work. All I want to do is relax."* Yet, some – including P15 – mentioned that once they felt available, they would choose to look such information up.

No Unified Preferences About The Presentation of Confidence
Nine participants were in favor of numeric presentation; seven preferred categorical presentation; two liked both equally; and two offered no opinion. P6 mentioned that with the numeric format, *"it was easier to customize the threshold for automation"*. Precision was another reason some people preferred the numeric approach: *"I like 80% better because it is more precise. Above the middle could be considered 'high', therefore 'high confidence' could range from 60% to 90%, and to me, there is a huge difference"* (P5). Those who preferred the categorical approach argued that a rough idea about confidence was good enough: *"If given too much information, I might overthink the system"* (P8). P19 likewise commented, *"it is hard to describe what I want to do now with numbers. The idea of high, medium and low is more abstract"*.

Assessment, Diagnosis, Correction, and Improvement
As frictions caused by false predictions may be unavoidable, especially in the early stage of adopting an SHRA. Below, we present how our participants assessed and corrected automated actions caused by the SHRA's false predictions, and their suggestions for system improvement.

Assessment of Automated Acts: Incorrect vs. Inappropriate
We observed that the participants' assessments of whether automated actions were correct or appropriate were based on three main factors: 1) the timing for turning on specific home appliances; 2) the confidence-level information they were given; and 3) how they made sense of the SHRA. With regard to the first, most participants preferred the SHRA to ask for permission to turn on the TV and AC before they reached the remote control. That is, they considered the system response to be too late if they were already reaching for the remotes, because such actions were obvious signs of their intentions. For appliances without remotes, on the other hand, the participants mostly felt that the ideal time to be notified was when they were walking toward the switches. However, some thought that this was too late, and preferred to be notified as soon as they stood up.

Confidence level also played a vital role in the participants' assessments of automated actions. A low-confidence action was not necessarily considered a success even if it was both correct and timely, as some participants ascribed it to luck, while others perceived it simply as problematic. As P4 commented, *"It did it right but only had 20% confidence. I think it's still a wrong prediction. I think it's too low, it should have been higher."* False or late automated actions performed on the basis of low confidence, on the other hand, were generally considered understandable and forgivable, though some people expressed curiosity about why the SHRA had performed them despite its confidence being low. Their greatest dissatisfaction was with false/late automated actions performed with high confidence. As P8 said, *"I would scold the system when it made wrong predictions and told me that it had high confidence. If it had low confidence and got it wrong, that's more understandable"*. Similarly, P17 said, *"If it had low confidence and made mistakes, I could accept it. But if it had high confidence and still made mistakes, I probably wouldn't"*.

We found it particularly interesting that a participant's assessment of whether an act was a failure also depended on its fit with their mental models and whether the false predictions made sense in the wider context. For example, P15 told us, *"Since the two remotes were placed right next to each other, it was reasonable for the system to automate the act, I think it's quite okay."* Similarly, P20 said, *"I would assess multiple automated acts together. If I wanted to turn on the lights yet it turned on the TV, I would be okay with it, since both of the things were what I would possibly do."* He also commented on a specific wrong action by the SHRA, i.e., turning on the TV when he intended to turn on the AC: *"I didn't feel it's a wrong judgment. It asked me about TV possibly because I might turn on the TV after I turned on the AC. [...] So it was predicting my next move."* As for the later factor, P8 mentioned that as long as the false prediction made sense in the wider context of whether he foresaw himself turning the same appliance on later, he felt comfortable with it being wrong.

Diagnosis: Figuring Out What Went Wrong and Why
Except for those (n=2) who cared only about prediction outcomes, the participants tended to diagnose SHRA actions' correctness by speculating about what information the system

had collected, how it had predicted their intentions, and why the automated action was chosen. All such participants also expressed a desire for the SHRA to offer information beyond confidence. P20 reported that he would know how to adjust the system based after learning "knowing how it decides", while P11 said *"Confidence allowed me to check what kinds of judgement it makes and what kinds of problems it has."*

The participants' desire for more information was especially strong when false predictions were made with high confidence. As P8 said, *"Which part of my action led it to have such high confidence that it made this mistake? If I knew the clues it used, I'd figure out how to weaken the connection. [...] I think this solution would work."*

While the surveillance camera was introduced to the participants only briefly, without an explanation of how a vision-based machine-learning model could recognize their behavior, several tried to identify the information SHRA might have leveraged to make its predictions. P2 speculated that the TV had been automatically turned on *"because I was facing the TV when reading"*. A few participants even speculated that the SHRA had collected extra, non-visual information from the surveillance camera, such as body temperature, postures, eye contact, and the status of other appliances: *"Could it tell me why it thought I wanted to turn on the AC? Was it because I entered a certain zone or I had high body temperature?"* (P13)

Correcting: Giving Feedback to The SHRA In The Moment.

The participants were shown a variety of interfaces with different modalities, and asked to compare how they liked each one as a means of receiving communication and providing feedback to the SHRA in the moment. Participants' availability was again a key factor here. While P2 said, *"Since the prediction was wrong, I would like to fix it now."*, many perceived themselves as pressed for time wished to provide feedback later; as P15 explained, *"It is annoying to have multiple prompts for feedback in a short period of time [...]. I'd rather provide feedback all at once later."*

The participants generally preferred interfaces that could be accessed quickly and directly. However, for this reason, they had diverse preferences vis-a-vis smartwatches, smartphones, and VUI. Smartwatches were considered the most convenient by those who already wore them on a daily basis: *"When at home, I might leave my smartphone somewhere and go do other things, but I would always have my watch with me"* (P6).

Other participants favored either smartphones or VUI, depending on how often they carried their phones with them at home, and how easy they found it to communicate via VUI. The advocates of the latter reported that "talking" to the SHRA from anywhere in the home was more convenient than communicating via a tangible interface. As P12 commented, *"I think voice assistant is better because I don't need to carry a device around."* Some mentioned other advantages, such as that smartphones allowed more fine-grained adjustments to room temperature or notification time, due to their relatively large screen size. Some participants felt that VUI absolved them from learning how to use a tangible interface, i.e., that they could "just talk" to give feedback; in that context, P19 stated, *"I don't want to learn another interface just for a new device."* However, a few participants with experience of using VUIs questioned how easy they would be to use as one' primary interface with an SHRA. Some mentioned that inputs might be recognized incorrectly, especially when there were other noises, which made them likely to repeat commands; and others thought that listening to VUI was too time-consuming and slow. *"I need to listen to the last word to find out that it has low confidence. I hope it can either talk faster or rearrange the important line to the front"* (P8). Another concern was that the voice the SHRA generated might disturb family members that were not engaging with it: *"I would prefer voice assistants when being alone, yet if there were other people doing their things in the same area, I would choose smartwatches or smartphones."* (P10).

Lastly, a few participants said they wanted to communicate via other methods to correct the SHRA in the moment: e.g., directly via the camera; moving their body/making gestures; or suddenly picking up the TV remote, in the hope that SHRA would recognize their intentions and correct itself.

Strategies for Improving The SHRA's Predictions

As well as wanting to give the SHRA real-time feedback, many participants mentioned that they would like to improve its prediction retrospectively, using one or more of the following strategies: configuring, teaching, simplifying, and complying. Their choices regarding such strategies depended both on their mental models of the SHRA, and how much effort they were willing to put in to improving it.

The configuration approach could be subdivided into two types. First, despite confidence being computed by the SHRA itself, many participants perceived it as modifiable, and wanted to suggest the 'correct' level of confidence the system should have computed. For example, P1 said, *"it said that it had 20% confidence. [...] the confidence level should have been higher. I wish there was some way to adjust the confidence level upward."* The second type was to reconfigure the measurements and thresholds the participants thought the SHRA had adopted to estimate their intentions. Whereas some participants suggested distance (in meters or step count), others suggested time (in seconds). Some also proposed allowing directly specification of a location on a map of their home.

The second strategy they mentioned was directly teaching the SHRA their behaviors. For example, P16 said that if the system erred, *"I would want to start re-training it. I could only train it by repeating my current gestures and giving the same feedback."*

The third strategy, advocated by participants who perceived that certain associations between their behaviors and their intentions were challenging for SHRA to accurately recognize (e.g., due to the complexity of the physical setting, or behaviors that might reasonably signal multiple possible intentions), was to simplify the patterns the SHRA needed to recognize and learn. To achieve this, they rearranged the physical configuration of the area, so that the relation between home appliances and furniture would be more distinguishable. As P5 commented, *"I should place the AC remote further [from* *the TV remote], so the system can detect it better."* Similarly, P8 said that if the SHRA made wrong predictions, she *"might move the TV a little bit away from the AC."*

Rather than attempting to improve the SHRA's learning, the fourth strategy consisted of the participants' compliance with the existing patterns that they perceived the system had learned from them. Several mentioned that they would try to recall which of their postures and gestures had triggered correct automated actions, and then deliberately perform them when they wanted the same effect. Therefore, those participants who thought the system recognized their intentions chiefly via the camera reported staring at it: e.g., *"I wanted to turn on the lights, so I gave the light an extra few stares. Even after I left, I still kept staring at the switch, thinking maybe that would increase the system's confidence [about turning on the light]"* (P5). Other participants waved at the AC or picked up the TV remote deliberately for similar reasons.

However, some participants did not expend effort on improving the system, either because they thought doing so was the sole responsibility of product developers, or because they regarded themselves as unable to improve it. P1, for instance, said: *"I suppose these errors should be resolved by the engineers [...]. Users can't do anything about it even though they know it had errors. They can't just read the manual and revise the codes"*. Participants who expected the system to have already been well-trained at the time of purchase reported that when the system errs they would just call customer service to fix it (e.g., P11). Although these coping strategies were inherently incapable of improving the system, they should be anticipated by anyone selling SHRAs to the public.

Personification and Expectations

Many previous studies have indicated that users would personify computers and conversational agents, such as by showing them politeness or otherwise responding to them as if they are human [33, 48]. Thus, it was not surprising that some of our participants scolded, encouraged, and/or praised the SHRA. Also, the adjectives used to describe it included "respectful", "considerate" and "warm". Taking P10's comment as an example, *"It is trying to interact with me. I find it quite sweet!"* When the system failed, they would reciprocate its friendliness; as P7 said, *"If it told me it would try harder to make better predictions next time, I would forgive it."*

The participants also mentioned numerous expectations of the SHRA. Essentially, they expected it to be like an intelligent housekeeper, who would ask the members of the household for permission to execute actions when it had not yet fully learned their needs and preferences, but who would then learn them rapidly. P20, for example, said, *"I want it to really understand me, and its understanding should be built on me telling it what I really need"*. Similarly, P11 noted that *"Just like fighting bosses in videogames, you have to reach certain levels to gain certain trust. [...] Winning the owners' trust should be based on the correctness of its actions. This is what this kind of relationship should look like, instead of doing whatever it wants. So, it's like 'mindsync'. If it can't operate at that level, it won't be 'smart' to me."* One participant (P8) also expected the SHRA to be able to justify why it did certain things through interactive Q&A.

Many participants also expected that the SHRA would be equipped with multiple capabilities related to personalization and contextual awareness, such as identifying seasonal routines. For example, P14 said, *"once the season changes, since I won't need AC during wintertime, smart homes could have a different setting that asks me whether I would like to set turning on the TV as the primary task."* Also, members of households with pets told us that they were worried the SHRA might automate actions based on pets' behaviors, and hoped it would eliminate such detection 'noise'.

Surprisingly, nine out of eleven participants who mentioned how fast systems should learn expected that an SHRA would learn their patterns well within just one week, due to their routines being broadly weekly; and of these, two even anticipated it would reach this level in three days. Just two were willing to give it at least a month. Yet if the SHRA failed to learn within the anticipated timeframe, it might lead to frustration. P7, for example, said, *"I would be pissed since I think grace period shouldn't really be longer than that!"* P19, who was only willing to give the system three days, said *"If it has to take up to a week to adapt, it would be problematic and I probably would start distrusting it, I would not use it"*. In fairness, given that not many of our participants had been exposed to machine learning or artificial intelligence before, they might not have been able to develop realistic expectations about how fast an SHRA could learn their pattern reliably. Given these circumstances, developers of such agents should anticipate such expectations and manage their customers' expectations about how long they took to train. As P9 suggested, *"If upon purchasing the system you have told the purchasers that it requires learning, and told the purchasers that it might take up to multiple weeks to achieve good results, then I would accept its errors and teach it slowly."* Similarly, P8 noted that if users understood that the system could gradually improve though their efforts, they might be more willing to "grow" with it: *"knowing that it is learning is enough for me, since I know it has put my feedback into consideration; though I don't know whether or not it is actually improving, knowing that it acknowledges it requires improvements is acceptable"*.

DISCUSSION

Confidence Mattered, but Sometimes Not Enough

It has been acknowledged that system status is vital to smooth human-computer interaction, regardless of whether AI is embedded in the system [49, 2]. We observed that the inclusion of confidence and its specific level impacted our participants' assessments and diagnoses of the SHRA's automated actions. Despite their different interpretations of what confidence was, the SHRA's expressions of its confidence would impact their assessments and perceptions both of whether an automated action was problematic, and how they would adjust the SHRA to correct problems. Their wide array of interpretations of and feelings about confidence included unexpected notions such as *machine-like*, *immature*, and *reluctant*. Miller [45] suggested the possibility that confidence is not an ideal way to explain intelligent systems to users, and Wang et al.[61]

also indicated that confidence models fall short when compared to natural-language explanations. Though no-one in the present study stated outright that confidence was an unacceptable term, our findings tend to support the idea that confidence may not be the most suitable or understandable way of describing an SHRA's status. This may be because ordinary users are unfamiliar with the use of confidence in such contexts. It could also be because our enactments did not include any other explanations for the SHRA's automated acts. If we had had included any, the participants might have comprehended confidence differently. But given that users make sense of, assess, and diagnose SHRAs' behavior based on the content of user-system communication, it is crucial for future research to explore other alternative means of describing and explaining such systems' prediction outcomes and decisions. It might also be also worthwhile to study quantitatively, with larger samples, how users' interpretations of SHRA's status affect their sense-making and subsequent system adoption.

Mental Models' Impacts on User-system Interaction

Myers et al. suggested that users' tactics when systems err mostly fall into the categories of guessing and exploration [47]; but in our SHRA case, we found that participants' assessment, diagnosis, and improvement strategies were influenced by their mental models, and varied dramatically. For example, those who perceived prediction outcomes as either right or wrong tended to care only about what was being done, and not about the reasons. This group also cared little about confidence, and often attributed incorrect predictions to internal errors that needed to be resolved by product developers, rather than to the system's learning curve. In contrast, those curious about the system react differently. They were eager to learn more about how the SHRA worked and how automated acts were made.

Participants' mental models also seemed to affect how they planned to improve the SHRA's prediction capability. Those who attempted to simplify the learning setting appeared to develop a more advanced mental model of SHRA based on their observations of the holistic relationship between the elements in the detection area, as well as on their articulation of the difficulty of recognizing patterns within a complex configuration. The teaching approach, meanwhile, was adopted by those who, at minimum, reasoned that prediction was related to the association between their behaviors and their subsequent actions involving specific home appliances. Participants adopting the compliance approach, in contrast, were relatively unmotivated to improve the SHRA's learning, but were generally comfortable with fitting themselves into what it had learned already. However, despite all participants' ability to diagnose the SHRA's problems and develop ways to address them, their understandings and the resulting strategies were not always accurate; and at worst, this could lead the SHRA's learning and performance to deteriorate. For example, while it was interesting to observe that false predictions were not considered inappropriate as long as they seemed reasonable, participants complying with the system and not correcting the acts was likely to adversely affect the system's confidence computations in the long run. More critically, most participants thought that the SHRA should start to perform well within a week, implying that the building of more accurate mental models and the fostering of more realistic expectations should be incorporated into SHRAs' user-system communication approaches.

DESIGN SPACES FOR SHRAS

Based on this study's main findings, we propose the following four design spaces for SHRAs' user-system communication.

At Onboarding

Three essential categories of information should be shared between an SHRA and first-time users. These are:

1) *Assessment* consists of checking the household's existing AI-related knowledge and mental models, i.e., their understanding of the AI embedded in some widely known services, along with their knowledge of machine training and learning. We have found that users might not share the same mental models, and that it is vital to identify where the household is on a spectrum from caring only about *what* is done vs. *why* it is done, since such mental models would affect their future interactions with, assessments of, and expectations about SHRAs. In the past, it has also been suggested that repairing strategies should be matched to people's different orientations [31].

2) *Customization* means allowing households to tailor user-system communication. At least, based on our study data, this will include their preferences about system-notification interfaces and responses, and their preferred default communication content regarding automated actions. SHRAs should provide details of their confidence (or a similar construct), with explanations, in a clear presentation format; and it should be made clear from the start which home appliances the SHRA must ask for permission to turn on or off. Additionally, for homes that are co-living spaces, SHRAs need to be able to identify their different inhabitants and tailor their automated actions to their respective needs and levels of authority.

3) *Setup* comprises providing households with tutorials on how to appropriately give feedback and improve their SHRAs, how to manage their expectations regarding these systems' performance, and how much training they need. Our study shows households with sufficient baseline understanding of SHRAs tend to be more willing to give such systems adequate time to learn.

At Routing Prediction and Execution

We learned from our findings that during its initial stage of learning, an SHRA should ask the members of its household for their permission to perform actions. In part, this is because receiving regular feedback over a lengthy period is important to enhancing the quality of its predictions. Moreover, as well as serving to remind its users that it is still learning, asking for permission shows the SHRA's respect for them. We learned that the key is to make households more comfortable with the system's learning curve. As its number of successful learning experiences increases, the SHRA may gain the trust of the members of its household. Once this stage is reached, it can ask the household members if it should continue asking for their permission before performing automated actions, or only notify them when its confidence is above a certain level.

As the above discussion implies, we feel that an SHRA should ask for permission when it has low confidence that its action is correct. However, if the above-mentioned onboarding assessment indicates that the household desires relatively little involvement with the system, and is more tolerant of incorrect predictions, the SHRA could occasionally automate without asking, at least in cases where the cost of the action is low (both financially and in terms of effort). The SHRA also ought to discourage the household from disabling its notifications, by giving a proper explanation of the potential negative consequences of choosing this option. We argue that this is vital not only because it maintains the household members' awareness of the SHRA's important functions, but also because it affords them opportunities to interfere with its automated actions whenever they wish to. On the other hand, as our findings indicate that many factors can affect a household's desired level of involvement with their SHRA as well as the content of their communication with it, a more advanced service should be context-aware, not only of when to turn on certain home appliances, but also when to involve users; what communication content to show; and when to use certain interfaces to communicate. Factors that can be considered include availability, mood, which people and pets are present, ambient noise, the cost of incorrect actions, etc.

Finally, the SHRA should present communication content preset by the household, but also offer an option for them to request further explanation of its predictions, as we found that occasionally they were interested in these details. For example, when the household interferes with an automated action, especially one executed on the basis of high confidence, a friendly message should be presented first, followed by the SHRA's prediction status and an explanation that will allow more accurate assessment and diagnosis of whether its learned patterns should be revised.

Occasional Tips, Tricks and Quick Facts

Our research has shown different mental models for SHRAs. Prior research has shown that people are more likely to form accurate mental models of autonomous intelligence systems when such systems provide explanations [13]. However, users do not always pay close attention to onboarding tutorials and have the tendency to skip those materials [4, 14]. The education of households who are uninterested in learning about their SHRAs is therefore a critical challenge. To prevent early system abandonment due to unrealistic expectations caused by inappropriate mental models of what SHRAs are capable of, such systems should occasionally sneak in brief tips, tricks, and quick facts about what information they detect, how they learn, and how users can help them improve their predictions. These messages could be placed on the user interface when households are waiting, e.g., during system loading or transitions. Links to external pages containing explanations could also be placed next to false and/or low-confidence predictions.

Administration

Finally, our data suggest that an SHRA should allow the members of its household to review all records of its automated actions and predictions, along with explanations for them, in an administration dashboard. This dashboard should also allow the users to modify all preferences they had previously set, and allow more advanced configurations such as parameters for prediction and for deciding whether to execute an act.

LIMITATIONS

The current study was subject to a number of limitations. First, its qualitative findings were obtained from user enactments, which took place in a mocked-up living room during a limited period of time. Thus, we were not able to observe the natural experience of households as we would have been able to in an in-the-wild study. The scenarios our participants enacted were also limited in terms of quantity and diversity, with the result that we could not observe or hear our participants reflecting on other scenarios. In addition, in all enacted scenarios, a lone participant was the only person interacting with the system, whereas in real life a smart-home system might be required to interact with several people and animals simultaneously.

The average age of our 20 participants was 24.5, with the oldest being 33, and all were students or white-collar workers with relatively high levels of technology acceptance. This would tend to limit the generalizability of the results. Also, smartwatch prototypes were printed out on cardboard, and the participants were asked to imagine that they were digital. Thus, it might not have captured nuanced differences in interactions across different types of screens. And lastly, given that the goal of the study was to explore the design space and to generate insights, our data were qualitative, and thus we could not draw any firm conclusions about the relationships among mental modes, the inclusion of confidence, and user-system interaction. Nevertheless, we believe that this study has generated numerous useful findings and practical design recommendations in the four proposed design spaces of SHRA user-system communication.

CONCLUSION

As various IoT appliances mature, and all-in-one smart-home ecosystems are on their way to the market, this exploratory study is an important step toward understanding both what and how to communicate with potential SHRA-owning households. Knowing how our participants perceived the SHRA and expected it to function should help practitioners tailor such systems. Of particular usefulness in this regard will be our findings regarding how confidence and users' mental models affected their assessment, diagnosis, and improvement strategies for the SHRA; the factors that affected their desired involvement in the system's decision-making; and their levels of interest in the system's confidence levels and the reasons for those levels. We hope that these findings and our design recommendations will help researchers and practitioners interested in SHRA to develop smart-home automation systems with smooth user experiences and long-term user loyalty.

ACKNOWLEDGMENT

We thank team CarePlus for their help with our study. We also thank the Ministry of Science and Technology, R.O.C. for their support (MOST 108-2634-F-007-006, MOST 108-2218-E-009-050).

REFERENCES

[1] Alper Turan Alan, Mike Shann, Enrico Costanza, Sarvapali Dyanand Ramchurn, and Sven Seuken. 2016. It is Too Hot: An In-Situ Study of Three Designs for Heating. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 5262–5273. DOI: http://dx.doi.org/10.1145/2858036.2858222

[2] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 3, 13 pages. DOI: http://dx.doi.org/10.1145/3290605.3300233

[3] Louise Barkhuus and Anind Dey. 2003. Is Context-Aware Computing Taking Control away from the User? Three Levels of Interactivity Examined. In *UbiComp 2003: Ubiquitous Computing*, Anind K. Dey, Albrecht Schmidt, and Joseph F. McCarthy (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 149–156.

[4] Fred Beisse. 2009. *A Guide to Computer User Support for Help Desk and Support Specialists* (4th ed.). Course Technology Press, Boston, MA, United States.

[5] Julia Brich, Marcel Walch, Michael Rietzler, Michael Weber, and Florian Schaub. 2017. Exploring End User Programming Needs in Home Automation. *ACM Trans. Comput.-Hum. Interact.* 24, 2, Article 11 (April 2017), 35 pages. DOI:http://dx.doi.org/10.1145/3057858

[6] Ryan Brotman, Winslow Burleson, Jodi Forlizzi, William Heywood, and Jisoo Lee. 2015. Building Change: Constructive Design of Smart Domestic Environments for Goal Achievement. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3083–3092. DOI: http://dx.doi.org/10.1145/2702123.2702602

[7] A.J. Bernheim Brush, Bongshin Lee, Ratul Mahajan, Sharad Agarwal, Stefan Saroiu, and Colin Dixon. 2011. Home Automation in the Wild: Challenges and Opportunities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 2115–2124. DOI:http://dx.doi.org/10.1145/1978942.1979249

[8] Andrea Bunt, Matthew Lount, and Catherine Lauzon. 2012. Are Explanations Always Important?: A Study of Deployed, Low-cost Intelligent Interactive Systems. In *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces (IUI '12)*. ACM, New York, NY, USA, 169–178. DOI: http://dx.doi.org/10.1145/2166966.2166996

[9] Nico Castelli, Corinna Ogonowski, Timo Jakobi, Martin Stein, Gunnar Stevens, and Volker Wulf. 2017. What Happened in My Home?: An End-User Development Approach for Smart Home Data Visualization. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 853–866. DOI: http://dx.doi.org/10.1145/3025453.3025485

[10] International Data Corporation. 2019. Double-Digit Growth Expected in the Smart Home Market, Says IDC. (2019). https://www.idc.com/getdoc.jsp?containerId=prUS44971219.

[11] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. 1993. Wizard of Oz Studies: Why and How. In *Proceedings of the 1st International Conference on Intelligent User Interfaces (IUI '93)*. ACM, New York, NY, USA, 193–200. DOI: http://dx.doi.org/10.1145/169891.169968

[12] Scott Davidoff, Min Kyung Lee, Anind K Dey, and John Zimmerman. 2007. Rapidly Exploring Application Design Through Speed Dating. In *Proceedings of the 9th International Conference on Ubiquitous Computing (UbiComp '07)*. Springer-Verlag, Berlin, Heidelberg, 429–446. http://dl.acm.org/citation.cfm?id=1771592.1771617

[13] Maartje de Graaf and Bertram Malle. 2017. How People Explain Action (and Autonomous Intelligent Systems Should Too). (2017). https://aaai.org/ocs/index.php/FSS/FSS17/paper/view/16009

[14] Jennifer deWinter and Ryan M Moeller. 2016. *Computer Games and Technical Communication: Critical Methods and Applications at the Intersection*. Taylor & Francis. https://books.google.com.tw/books?id=N0s3DAAAQBAJ

[15] Jillian D'Onfro. 2019. AI 50 Founders Say This Is What People Get Wrong About Artificial Intelligence. (2019). Retrieved November 9, 2019 from https://shorturl.at/dgqr3

[16] Upol Ehsan, Pradyumna Tambwekar, Larry Chan, Brent Harrison, and Mark O Riedl. 2019. Automated Rationale Generation: A Technique for Explainable AI and Its Effects on Human Perceptions. In *Proceedings of the 24th International Conference on Intelligent User Interfaces (IUI '19)*. ACM, New York, NY, USA, 263–274. DOI: http://dx.doi.org/10.1145/3301275.3302316

[17] Telefonaktiebolaget L. M. Ericsson. 2018. Internet of Things forecast. (2018). https://www.ericsson.com/en/mobility-report/internet-of-things-forecast.

[18] Google APIs for Android. 2019. Detected Activity. (2019). Retrieved September 12, 2019 from https://developers.google.com/android/reference/com/google/android/gms/location/DetectedActivity

[19] Christine Geeng and Franziska Roesner. 2019. Who's In Control?: Interactions In Multi-User Smart Homes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 268, 13 pages. DOI: http://dx.doi.org/10.1145/3290605.3300498

[20] Google. 2019. People+AI Guidebook: Explainability + Trust. (2019). Retrieved June 3, 2019 from https://pair.withgoogle.com/chapter/explainability-trust/

[21] Mateusz Halawa and Marta Olcoń-Kubicka. 2018. Digital householding: calculating and moralizing domestic life through homemade spreadsheets. *Journal of Cultural Economy* 11, 6 (2018), 514–534.

[22] Chris Harrison, John Horstman, Gary Hsieh, and Scott Hudson. 2012. Unlocking the Expressivity of Point Lights. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 1683–1692. DOI: http://dx.doi.org/10.1145/2207676.2208296

[23] Google Nest Help. 2019. Set up manage Routines. (2019). Retrieved September 12, 2019 from https://support.google.com/googlenest/answer/7029585?co=GENIE.Platform%3DAndroid&hl=en

[24] Chuan-Che (Jeff) Huang, Sheng-Yuan Liang, Bing-Hsun Wu, and Mark Webster Newman. 2017. Reef: Exploring the Design Opportunity of Comfort-Aware Eco-Coaching Thermostats. In *Proceedings of the 2017 Conference on Designing Interactive Systems (DIS '17)*. ACM, New York, NY, USA, 191–202. DOI: http://dx.doi.org/10.1145/3064663.3064685

[25] Rikke Hagensby Jensen, Yolande Strengers, Jesper Kjeldskov, Larissa Nicholls, and Mikael B Skov. 2018. Designing the Desirable Smart Home: A Study of Household Experiences and Energy Consumption Impacts. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 4, 14 pages. DOI: http://dx.doi.org/10.1145/3173574.3173578

[26] Jenny Kennedy, Bjorn Nansen, Michael Arnold, Rowan Wilken, and Martin Gibbs. 2015. Digital housekeepers and domestic expertise in the networked home. *Convergence* 21, 4 (2015), 408–422.

[27] Da-jung Kim and Youn-kyung Lim. 2019. Co-Performing Agent: Design for Building User-Agent Partnership in Learning and Adaptive Services. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 484, 14 pages. DOI: http://dx.doi.org/10.1145/3290605.3300714

[28] Todd Kulesza, Margaret Burnett, Weng-Keen Wong, and Simone Stumpf. 2015. Principles of Explanatory Debugging to Personalize Interactive Machine Learning. In *Proceedings of the 20th International Conference on Intelligent User Interfaces (IUI '15)*. ACM, New York, NY, USA, 126–137. DOI: http://dx.doi.org/10.1145/2678025.2701399

[29] Todd Kulesza, Simone Stumpf, Margaret Burnett, and Irwin Kwan. 2012. Tell Me More?: The Effects of Mental Model Soundness on Personalizing an Intelligent Agent. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 1–10. DOI: http://dx.doi.org/10.1145/2207676.2207678

[30] Dounia Lahoual and Myriam Frejus. 2019. When Users Assist the Voice Assistants: From Supervision to Failure Resolution. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, New York, NY, USA, Article CS08, 8 pages. DOI: http://dx.doi.org/10.1145/3290607.3299053

[31] Min Kyung Lee, Sara Kielser, Jodi Forlizzi, Siddhartha Srinivasa, and Paul Rybski. 2010. Gracefully Mitigating Breakdowns in Robotic Services. In *Proceedings of the 5th ACM/IEEE International Conference on Human-robot Interaction (HRI '10)*. IEEE Press, Piscataway, NJ, USA, 203–210. http://dl.acm.org/citation.cfm?id=1734454.1734544

[32] Yoga Liu, Ya-Han Lee, Yaliang Chuang, Rung-Huei Liang, and Lin-Lin Chen. 2017. Designing the Expressive Point Lights to Enhance User's Situated Awareness of Smart Systems. In *Proceedings of the 2017 ACM Conference Companion Publication on Designing Interactive Systems (DIS '17 Companion)*. ACM, New York, NY, USA, 333–336. DOI: http://dx.doi.org/10.1145/3064857.3079187

[33] Irene Lopatovska and Harriet Williams. 2018. Personification of the Amazon Alexa: BFF or a Mindless Companion. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval (CHIIR '18)*. ACM, New York, NY, USA, 265–268. DOI: http://dx.doi.org/10.1145/3176349.3176868

[34] Andrés Lucero. 2015. Using Affinity Diagrams to Evaluate Interactive Prototypes. In *Human-Computer Interaction – INTERACT 2015*, Julio Abascal, Simone Barbosa, Mirko Fetter, Tom Gross, Philippe Palanque, and Marco Winckler (Eds.). Springer International Publishing, Cham, 231–248.

[35] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf Between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 5286–5297. DOI: http://dx.doi.org/10.1145/2858036.2858288

[36] Gabriel Machado Lunardi, Fadi Al Machot, Vladimir A. Shekhovtsov, Vinícius Maran, Guilherme Medeiros Machado, Alencar Machado, Heinrich C. Mayr, and José Palazzo M. de Oliveira. 2018. IoT-based human action prediction and support. *Internet of Things* 3-4 (2018), 52 – 68. DOI: http://dx.doi.org/https://doi.org/10.1016/j.iot.2018.09.007

[37] Michal Luria, Guy Hoffman, and Oren Zuckerman. 2017. Comparing Social Robot, Screen and Voice Interfaces for Smart-Home Control. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 580–628. DOI: http://dx.doi.org/10.1145/3025453.3025786

[38] Ilias Maglogiannis, Kostas Delibasis, Dimitrios Kosmopoulos, Theodosios Goudas, and Charalampos Doukas. 2014. Activity Recognition in Assistive Environments: The STHENOS Approach. In *Universal Access in Human-Computer Interaction. Aging and Assistive Environments*, Constantine Stephanidis and Margherita Antona (Eds.). Springer International Publishing, Cham, 525–536.

[39] Mitchell Mason. 2019. IBM Watson: How To Set the Optimal Confidence Threshold for Your Assistant. (2019). Retrieved September 12, 2019 from https://medium.com/ibm-watson/how-to-set-the-optimal-confidence-threshold-for_your-assistant-bbb1be6f294f

[40] Jon McCormack, Toby Gifford, Patrick Hutchings, Maria Teresa Llano Rodriguez, Matthew Yee-King, and Mark d'Inverno. 2019. In a Silent Way: Communication Between AI and Improvising Musicians Beyond Sound. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 38, 11 pages. DOI: http://dx.doi.org/10.1145/3290605.3300268

[41] Sarah Mennicken, Jo Vermeulen, and Elaine M Huang. 2014a. From Today's Augmented Houses to Tomorrow's Smart Homes: New Directions for Home Automation Research. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*. ACM, New York, NY, USA, 105–115. DOI: http://dx.doi.org/10.1145/2632048.2636076

[42] Sarah Mennicken, Jo Vermeulen, and Elaine M Huang. 2014b. From Today's Augmented Houses to Tomorrow's Smart Homes: New Directions for Home Automation Research. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*. ACM, New York, NY, USA, 105–115. DOI: http://dx.doi.org/10.1145/2632048.2636076

[43] Sarah Mennicken, Oliver Zihler, Frida Juldaschewa, Veronika Molnar, David Aggeler, and Elaine May Huang. 2016. "It's Like Living with a Friendly Stranger": Perceptions of Personality Traits in a Smart Home. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 120–131. DOI:http://dx.doi.org/10.1145/2971648.2971757

[44] Merriam-Webster. 2019. Confidence. (2019). https://www.merriam-webster.com/dictionary/confidence.

[45] Tim Miller. 2017. Explanation in Artificial Intelligence: Insights from the Social Sciences. *CoRR* abs/1706.07269 (2017). http://arxiv.org/abs/1706.07269

[46] Mamun Bin Ibne Reaz Muhammad Raisul Alam and Mohd Alauddin Mohd Al. 2012. A Review of Smart Homes—Past, Present, and Future. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42, 6 (Nov 2012), 1190–1203. DOI: http://dx.doi.org/10.1109/TSMCC.2012.2189204

[47] Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for How Users Overcome Obstacles in Voice User Interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 6, 7 pages. DOI: http://dx.doi.org/10.1145/3173574.3173580

[48] Clifford Nass and Youngme Moon. 2000. Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues* 56 (03 2000), 81–103. DOI: http://dx.doi.org/10.1111/0022-4537.00153

[49] Jakob Nielsen and Rolf Molich. 1990. Heuristic Evaluation of User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '90)*. ACM, New York, NY, USA, 249–256. DOI:http://dx.doi.org/10.1145/97243.97281

[50] Don A Norman. 1987. Human-computer Interaction. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, Chapter Some Observations on Mental Models, 241–244. http://dl.acm.org/citation.cfm?id=58076.58097

[51] Donald A Norman. 1990. The 'problem'with automation: inappropriate feedback and interaction, not 'over-automation'. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 327, 1241 (1990), 585–593.

[52] William Odom, John Zimmerman, Scott Davidoff, Jodi Forlizzi, Anind K. Dey, and Min Kyung Lee. 2012. A Fieldwork of the Future with User Enactments. In *Proceedings of the Designing Interactive Systems Conference (DIS '12)*. ACM, New York, NY, USA, 338–347. DOI: http://dx.doi.org/10.1145/2317956.2318008

[53] William Odom, John Zimmerman, Jodi Forlizzi, Hajin Choi, Stephanie Meier, and Angela Park. 2014. Unpacking the Thinking and Making Behind a User Enactments Project. In *Proceedings of the 2014 Conference on Designing Interactive Systems (DIS '14)*. ACM, New York, NY, USA, 513–522. DOI: http://dx.doi.org/10.1145/2598510.2602960

[54] J. Rafferty, C. D. Nugent, J. Liu, and L. Chen. 2017. From Activity Recognition to Intention Recognition for Assisted Living Within Smart Homes. *IEEE Transactions on Human-Machine Systems* 47, 3 (June 2017), 368–379. DOI: http://dx.doi.org/10.1109/THMS.2016.2641388

[55] Tom A Rodden, Joel E. Fischer, Nadia Pantidi, Khaled Bachour, and Stuart Moran. 2013. At Home with Agents: Exploring Attitudes Towards Future Smart Energy Infrastructures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 1173–1182. DOI:http://dx.doi.org/10.1145/2470654.2466152

[56] Jennifer A Rode and Erika Shehan Poole. 2018. Putting the gender back in digital housekeeping. In *Proceedings of the 4th Conference on Gender & IT*. ACM, 79–90.

[57] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I-An Hong. 2018. "Hey Alexa, What's Up?": A Mixed-Methods Studies of In-Home Conversational Agent Usage. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. ACM, New York, NY, USA, 857–868. DOI: http://dx.doi.org/10.1145/3196709.3196772

[58] Simone Stumpf. 2019. Horses For Courses: Making The Case For Persuasive Engagement In Smart Systems. In *Joint Proceedings of the ACM IUI 2019 Workshops co-located with the 24th ACM Conference on Intelligent User Interfaces (ACM IUI 2019), Los Angeles, USA, March 20, 2019. (CEUR Workshop Proceedings)*, Christoph Trattner, Denis Parra, and Nathalie Riche (Eds.), Vol. 2327. CEUR-WS.org. http://ceur-ws.org/Vol-2327/IUI19WS-ExSS2019-4.pdf

[59] Joe Tullio, Anind K. Dey, Jason Chalecki, and James Fogarty. 2007. How It Works: A Field Study of Non-technical Users Interacting with an Intelligent System. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 31–40. DOI: http://dx.doi.org/10.1145/1240624.1240630

[60] Michele Turner. 2019. Lights, TV, even the backyard: how the Assistant helps you at home. (2019). Retrieved September 19, 2019 from https://www.blog.google/products/assistant/lights-tv-even-backyard-how-assistant-helps-you-home/

[61] Ning Wang, David Varghese Pynadath, and Susan G Hill. 2016. Trust Calibration Within a Human-Robot Team: Comparing Automatically Generated Explanations. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction (HRI '16)*. IEEE Press, Piscataway, NJ, USA, 109–116. http://dl.acm.org/citation.cfm?id=2906831.2906852

[62] Robert H Wortham, Andreas Theodorou, and Joanna J Bryson. 2016. What Does the Robot Think? Transparency as A Fundamental Design Requirement for Intelligent Systems. In *Ijcai-2016 ethics for artificial intelligence workshop*.

[63] Chu-Yao Wu. 2018. *A Study of Timing Issues for Multimodal Human Machine Interface for Smart Home*. Master's thesis. Japan Advanced Institute of Science and Technology, Stockholm(JAIST), Ishikawa, Japan.

[64] Heetae Yang, Wonji Lee, and Hwansoo Lee. 2018. IoT Smart Home Adoption: The Importance of Proper Level Automation. *Journal of Sensors* 2018 (2018).

[65] Rayoung Yang and Mark Webster Newman. 2013. Learning from a Learning Thermostat: Lessons for Intelligent Systems for the Home. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13)*. ACM, New York, NY, USA, 93–102. DOI: http://dx.doi.org/10.1145/2493432.2493489

[66] Yaxing Yao, Justin Reed Basdeo, Smirity Kaushik, and Yang Wang. 2019. Defending My Castle: A Co-Design Study of Privacy Mechanisms for Smart Homes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 198, 12 pages. DOI: http://dx.doi.org/10.1145/3290605.3300428

[67] Yu Zhao, Xihui Zhang, and John Crabtree. 2016. Human-Computer Interaction and User Experience in Smart Home Research: A Critical Analysis. *Issues in Information Systems* 17, 3 (2016).

[68] John Zimmerman and Jodi Forlizzi. 2017. Speed Dating: Providing a Menu of Possible Futures. *She Ji: The Journal of Design, Economics, and Innovation* 3, 1 (2017), 430–450.