Check for updates

A User-Centric Evaluation of Smart Home Resolution Approaches for Conflicts Between Routines

ALI ZAIDI*, University of Illinois at Urbana-Champaign, USA
RUI YANG*, University of Illinois at Urbana-Champaign, USA
VINAY KOSHY, University of Illinois at Urbana-Champaign, USA
CAMILLE COBB, University of Illinois at Urbana-Champaign, USA
INDRANIL GUPTA, University of Illinois at Urbana-Champaign, USA
KARRIE KARAHALIOS, University of Illinois at Urbana-Champaign, USA

With the increasing adoption of smart home devices, users rely on device automation to control their homes. This automation commonly comes in the form of smart home *routines*, an abstraction available via major vendors. Yet, questions remain about how a system should best handle *conflicts* in which different routines access the same devices simultaneously. In particular—among the myriad ways a smart home system could handle conflicts, which of them are currently utilized by existing systems, and which ones result in the highest user satisfaction? We investigate the first question via a survey of existing literature and find a set of conditions, modifications, and system strategies related to handling conflicts. We answer the second question via a scenario-based Mechanical-Turk survey of users interested in owning smart home devices and current smart home device owners (N=197). We find that: (i) there is no context-agnostic strategy that always results in high user satisfaction, and (ii) users' personal *values* frequently form the basis for shaping their expectations of how routines should execute.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing**; **User centered design**; *User studies.*

Additional Key Words and Phrases: smart home, routine, user satisfaction, execution

ACM Reference Format:
Ali Zaidi, Rui Yang, Vinay Koshy, Camille Cobb, Indranil Gupta, and Karrie Karahalios. 2023. A User-Centric Evaluation of Smart Home Resolution Approaches for Conflicts Between Routines. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 1, Article 45 (March 2023), 35 pages. https://doi.org/10.1145/3581997

45

1 INTRODUCTION

When automated actions in a smart home conflict with each other, a smart home routine management system must choose one among many possible outcomes to resolve the conflict. What strategies for choosing outcomes are currently used, which outcomes best meet user expectations, and what factors influence these expectations?

*Both authors contributed equally to the paper

Authors' addresses: Ali Zaidi, aliz2@illinois.edu, University of Illinois at Urbana-Champaign, Urbana, IL, USA; Rui Yang, ry2@illinois.edu, University of Illinois at Urbana-Champaign, Urbana, IL, USA; Vinay Koshy, vkoshy2@illinois.edu, University of Illinois at Urbana-Champaign, Urbana, IL, USA; Camille Cobb, camillec@illinois.edu, University of Illinois at Urbana-Champaign, Urbana, IL, USA; Indranil Gupta, indy@illinois.edu, University of Illinois at Urbana-Champaign, Urbana, IL, USA; Karrie Karahalios, kkarahal@illinois.edu, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

2474-9567/2023/3-ART45 $15.00
https://doi.org/10.1145/3581997

This paper seeks to answer these questions and help create design principles to aid developers of smart home management systems.

As smart homes grow in popularity [25, 90], so has the importance of dealing with these conflicts in a way that meets users' expectations. *Routines* serve as the primary abstraction of smart device actions in Amazon Alexa [6], Google Home [35], and other platforms [94], and can be used for a variety of purposes. A routine is essentially a series of commands that touches multiple devices. Routines may be triggered by time, sensors, or users. Thus, multiple routines may co-exist simultaneously within a smart home. While other platforms may refer to routines differently, such as scenes or automations [45, 46, 49, 101], they all provide similar functionalities and empower users to effectively operationalize smart devices to achieve goals at the scale of the smart home, beyond individual devices.

Conflicts among multiple routines, which occur when they happen to access overlapping sets of devices, can cause smart homes to act in unintended ways. Consider a home where users have installed two routines: *FireAlarm* and *LockTheDoor* [106]. *FireAlarm* sounds an alarm and unlocks the front door if smoke is detected. *LockTheDoor* automatically locks the front door at 9 pm. Most of the time, these two routines can co-exist within a home without any issue. However, consider a situation where smoke is detected at 8:59 pm, triggering the *Fire Alarm* routine. By the time the user reaches the front door, it becomes 9 pm, at which point the *LockTheDoor* routine is triggered, causing a conflict between these two routines on the state of the front door. This conflict can have significant consequences, such as the occupant of the house unexpectedly encountering a locked door while fleeing the fire because the lock receives conflicting commands. In isolation, each of these two routines functions correctly, yet real-life circumstances may lead to conflicts, which can lead to unintended consequences.

Consider another scenario: a user is going on vacation for one week, and so runs their *Vacation* routine, which keeps their smart camera at the front of the house on for the whole week, and schedules the lights to turn on at 6 pm, and turn off at 10 pm, to keep up the appearance that they are home. At the same time, this user has a *Who's there* routine installed, which when triggered by movement, starts recording on the smart camera attached to the front door for five minutes and then turns the camera off [4]. Again, in isolation, these routines are logical and can exist together in one home without any issues. However, consider a scenario where during this week-long vacation someone comes to the front door. There is now a routine conflict, as the camera is being told to turn off after five minutes by the *Who's there* routine, while the *Vacation* routine is telling the camera to stay on. If the resolution of this conflict is left purely to the routine management system of the smart home, a possible outcome the system may decide that the *Vacation* routine is overwritten by the *Who's there* routine, and after five minutes the camera is no longer recording for the rest of the user's vacation. This example highlights a situation where it is crucial to understand user preferences for resolving inter-routine conflicts so that these can be incorporated into future smart home management systems. Investigating smart home platforms rigorously from a human-centered perspective now is critical alongside the rise of consumer options.

Developers attempted to address conflicts via the introduction of smart home *management* systems that encapsulate routines, automate some of the user's preferences, and allow the system to *operate autonomously without user intervention.* These systems encode a variety of strategies for resolving conflicts in a systematic way [4, 24, 68, 79, 92, 97], treating the conflict like a distributed systems problem. Which of these strategies produces preferable outcomes with respect to user expectations, and how do users form these expectations in the first place?

We seek to understand the relationship between system strategies to resolve conflicts and user satisfaction with the corresponding outcomes of applying these strategies. Given that nearly 60% of smart home routines in a single home can access the same device [106], the high potential for routine conflicts must be considered when designing smart home routine management systems. It is crucial to consider user preferences regarding how systems should deal with routine conflicts. We evaluate how different system strategies affect user satisfaction, and identify potential factors that contribute to how users form their expectations. We accomplish this by conducting a survey on Amazon Mechanical Turk.

In doing so, we address the following research questions:

- **RQ1**: How do existing systems, including real-world systems and research systems, handle conflicts among routines?
- **RQ2**: How would users react to existing approaches for conflict handling?
- **RQ3**: What factors influence users' expectations for smart home routine conflict resolution?

This paper makes the following contributions:

- Presents a taxonomy of current strategies to address conflicts between routines in smart homes used in existing real-world and research systems (Sec. 3).
- Evaluates how well each strategy affects user satisfaction when applied to different scenarios in order to determine if there is a context-agnostic strategy that consistently addresses user expectations (Sec. 4).
- Reveals personal values that users consider when forming their expectations for how smart homes should handle routine conflicts (Sec. 5).
- Suggests design principles for more user-centric smart home routine management systems and future research directions (Sec. 6).

Smart homes have the potential to improve users' lives by increasing convenience, and feelings of safety, as well as creating a more personalized home experience based on user preferences. At the same time, smart homes also present several problems to users, as they can take away their feelings of autonomy, reduce their sense of privacy, over-complicate the day-to-day function of users' lives, and more. In this work, we focus on smart home routine conflicts, which can generate additional stress and risk that counteract the potential benefits.

Our work seeks to analyze the effect of existing approaches to resolving these conflicts from a *user-centric* point of view. Our taxonomy of current state-of-the-art strategies for resolving conflicts between routines reveals a diversity of approaches, underscoring the need for a user-centric evaluation (RQ1). The findings of our user-centric evaluation (RQ2) show a similarly diverse set of user opinions, and this motivates further analysis to uncover how users form their expectations of smart home routine conflict resolution systems (RQ3). The results of our user-centric evaluation provide a basis for how smart home designers should evaluate existing approaches, as well as what factors to consider to build more user-centric routine conflict resolution approaches. Through this work, we see where the current state-of-the-art falls short from a *user-centric* point of view, and provide design implications to give users more actionable power when using smart home systems.

2 RELATED WORK

There are three main areas of prior research that relate to this work: smart homes/devices and routines, existing smart home routine management systems, and user desires and expectations around smart homes.

2.1 Smart Homes, Device, and Routines

A *smart home* is a "generic descriptor for the introduction of enhanced monitoring and control functionality into homes"; it consists of data collection and analysis technology that are used to enhance traditional "domestic systems", such as heating, lighting, and entertainment. From this definition of a smart home, we define *smart devices* as "sensors, monitors, interfaces, appliances, and devices networked together to enable automation as well as localized and remote control" [38].

Prior research viewed smart homes as complex environments equipped with custom hardware tailored to the specific home [40, 59]. We instead focus on smart home environments created through the purchase of off-the-shelf technologies that support the creation of custom routines, as the user base of such smart homes is significantly larger. Within these smart homes, users will often create or use existing routines. A routine is composed of a list of actions to be completed *together* [20]. Each of the actions sets a smart device to a specific state, which we refer to as a *command* in this paper. For example, in the *FireAlarm* routine from Sec. 1, the two commands are turning on the alarm and locking the front door when smoke is detected in the house. Routines and commands can either be *long running*, meaning they execute over an extended period of time or *short running*, meaning they execute instantaneously. Smart home platforms such as Google Home [35], Alexa [6], and iRobot vacuums include support for *delayed* commands and routines [39, 71, 112], otherwise known as *long running* routines or commands [4]. Such routines with delayed commands are quite common, e.g. turn the water sprinkler ON at 8 am, and then turn the sprinkler OFF after 20 minutes.

Long routines or commands increase the probability of routine execution under non-ideal conditions, such as executing routines concurrently or with limited resources due to other routines using them. *Conflicts* between routines are more likely to happen when long routines are involved, as short running routines finish instantly, meaning there is less time for a second routine to run and conflict with it unless both routines start at the same time.

2.2 Smart Home Routine Management Systems

The increasing use of smart devices and routines as well as the resulting conflicts underscores the importance of researching routine management systems and abstraction designs. HomeOS [24] provides a PC-like abstraction so that all in-home devices appear to the user as peripherals connected to a single PC. Beam [98] optimizes resource utilization by partitioning applications/routines across devices. There are also several systems that focus on programming abstraction for sensors and actuators in the home [28, 109, 110].

Some routine management systems take the approach of detecting routine conflicts to inform the user or attempt to resolve them automatically. Kripke [116], IA-Graph [66, 67], and IoTCheck [105, 106] all take this approach. The evaluation of IoTCheck revealed that more than 90% of routines [1] interact with the physical environment and more than 60% of routine pairs can access the same types of device in the home, which can all be potential conflicts. User expectations motivate the adoption of routines in smart homes; when routines conflict with each other, so do these expectations. Yet, the authors suggested designing new APIs so that app developers could make more informed decisions during development, rather than a more user-centered solution.

Previous research has also integrated concurrency control schemes that attempt to adjust the execution of routines when conflicts arise and do not involve the user [24, 68, 79, 92]. In addition, abstracting smart homes as existing infrastructures or practices is another common approach in routine management. Previous work has used calendars to integrate smart home automation which provides users with a clear interface to spot conflicts [76], or used trigger-action programming paradigms in smart homes which can simplify user experiences with routines but do not help with conflict detection or resolution [107]. Similarly, researchers have suggested treating smart homes like databases rather than involving users during routine conflicts. *Transactuations* provides a consistent soft-state inside the smart home system, meaning that the states of smart devices can change without routine intervention, to reach a *consistent* state in the home [97]. *SafeHome* proposes visibility models that trade off atomicity, consistency, and isolation for routine execution, similar to the ACID properties of database transactions [18]. For example, it allows for the "rollback" of smart device commands when routines are in conflict, so that a "safe" state in the home is reached [4].

The systems described above provide options to the user, each of which ensures different system properties. However, they take a system-focused point of view, grounding the design choices in existing abstractions or distributed systems principles. This point of view assumes that user expectations of smart home devices and

[1] Most smart home applications are quite simple and can be expressed in routine format. For example, the `Lock-it-when-I-leave` application mentioned in IoTCheck [106] only has one command "locking the door" that will interact with the user. Thus, we will use the terms application, app, and routine interchangeably.

routines are all similar. However, we argue that smart home routine expectations differ among users. Therefore, resolving the conflict only from a system-centric point of view may not be sufficient to produce satisfactory outcomes for users. Previous work has identified components of user desires and expectations of smart homes, and we build upon this work by specifically investigating routine conflicts, and user expectations of their smart homes during them.

2.3 User Desires and Expectations around Smart Homes

Users who are not satisfied with how the technology operates have no incentive to adopt or continue using smart devices [37]. This satisfaction depends on several factors, such as compatibility with a user's life, how "easy" the device is to get started with and use, and the overall perceived usefulness of the device. Smart home systems must consider these factors in their design to gain new users (by establishing a positive perception of their product) and retain current users [99].

Previous research found that users weigh the potential downsides of smart homes (lack of privacy, device maintenance, loss of personal agency) against the potential additional comfort from using these devices when deciding to adopt them [11, 64, 113].

Smart homes should also be highly personalized to fit the needs of different users, or else they risk not being adopted. The smart home of a family with small children has different needs than one of a single adult, and without personalization smart home adoption would be severely limited [21]. Even within a single home, there can be different types of users, each with different sets of expectations of how smart devices can and should assist them [61]. Previous research has proposed technical solutions to increase the personalization of smart homes, such as using process mining, which involves analyzing device data to learn a user's daily routine [12], or user feedback loops to incrementally make smart homes more in tune with individual user needs [55]. In the past few years, smart device designers accepted that their systems must be flexible in the face of different users' needs and intentions [29], or else they risk becoming obsolete, or never being used at all by mainstream users [23]. Personal values have been shown to affect how users interact with products, such as smart homes, demonstrating their importance [111].

At the same time, users do not want to give up their agency in exchange for the added benefit of using smart devices in their homes. Davidoff et al. [22] found that people want to "control" their lives rather than just devices, meaning that they want to use smart homes as a tool to make their lives convenient, without feeling that this convenience comes at the cost of some personal agency. Users prefer smart devices that promote autonomy and remote control so that the "downside" of smart devices is no longer present while the "upside" remains [19, 96]. Rather than contorting their own behaviors to match smart devices and routines, users can use smart devices to achieve goals they already had [87], even if it means causing interpersonal tensions with other users in the same smart home [32, 61].

Many popular devices enable users to create their own new routines or use existing routines created by developers or other users to automatically control their smart home devices [6, 35, 94]. Discussions of comfort and control are, again, essential in the context of routines. One threat to comfort and control occurs when routines conflict; that is, when two routines attempt to control the same (at least one) smart device in an overlapping time duration.

Prior work has presented solutions to routine conflicts to users *before* these conflicts actually occur [13]. We extend the prior work by studying user expectations for how conflicts will be resolved by a routine management system *after* these conflicts occur.

3 EXISTING STRATEGIES ANALYSIS

3.1 Methodology

To answer **RQ1**, as well as to generate the list of strategies we would present in our survey to answer **RQ2** and **RQ3**, we first examined the set of conditions (or parameters) considered by existing systems for routine conflict resolutions. Once we understood the relevant dimensions in a routine conflict, we determined the set of possible *modifications* that could be made to a routine involved in a conflict, such as changing if and when an associated command of a routine would execute. Together, this allowed us to tabulate a comprehensive set of possible ways a system could deal with a routine conflict by modifying one or both routines, which we refer to as *system strategies* in this paper.

The process above consisted of several rounds of consensus building within the research group. "Consensus building" methods incorporate individuals' interests and thoughts into the final group decision. Through consensus building, we produced a set of conditions, modifications, and system strategies [103]. This process is further outlined in Section 4.1.2.

First, the research team analyzed the landscape of current smart home routine management systems. This analysis was conducted over two categories of systems: 1) Real world systems (Google Home [35], Amazon Alexa [6], Apple HomeKit [46], etc.), and 2) Research systems (SafeHome [4], IoTCheck [106], Krpike[117], etc.). Real-world systems were added to the analysis based on their market share of the smart home market [91], while research systems were added to the body of research until saturation had been reached, meaning that no new insights could be gathered from additional research [30]. The goal of this analysis was to find thematic intersections between these systems in terms of how they managed routines–especially in cases when two or more routines would provide conflicting instructions to the same device.

Members of the research team separately analyzed existing systems to extract sets of conditions. A condition refers to a single question relating to the routines or commands involved in a conflict that a system must answer when deciding how to resolve the conflict. For example, when resolving a conflict, the system must decide whether commands will be manipulated at a command or routine level, as this can drastically change how the routine conflict is resolved. These conditions encapsulate what a generic routine management system considered when deciding how to address a conflict between two routines. After working independently, the research team met jointly five times over a one-month period to discuss and reach a consensus on a single set of conditions.

From these conditions, the research team determined all possible modifications for a routine in conflict to reach a resolution. These modifications were on the individual command/routine level, and thus could be put together to generate full strategies for how a system can address a routine conflict. This process, similar to determining conditions, took place over several weeks, as the research team members worked individually, and then met to build consensus. The process resulted in a set of eight modifications (Table 2) that could be made to a routine/command in response to a routine conflict.

3.2 Results

Our consensus building process yielded three ***conditions*** (listed in Table 1) that existing systems addressed whenever proposing solutions to routine conflicts in smart homes. The possible values of these conditions then led to a list of ***modifications***. These became our ***system strategies***, some of which were used by actual smart home systems.

In order to address conditions C-1 and C-2 (Table 1), there were several ***modifications*** that a system could make to a whole routine or part of a routine (command) when there is a routine conflict. The research team built consensus over the set of these modifications, which are listed in Table 2.

A ***system strategy*** consists of the modifications to both routines (or commands, depending on the value of C-3) involved in the routine conflict. Thus, for two conflicting routines, multiplying the nine possible modifications to

Table 1. Conditions to be considered when choosing a strategy to deal with routine conflicts and the possible values of these conditions.

Id	Condition	Possible Values
C-1	If and when the unexecuted part of the routine/command will take place	[**Does not take place**, **Takes place right away**, **Takes place after the other routine (with delay)**]
C-2	Whether the system manipulates routines at a routine-level or command-level granularity	[**Routine-level**, **Command-level**]
C-3	Whether the system manipulates the schedule of the already running routine or the new routine that introduces the conflict	[**Already running routine**, **New routine**]

Table 2. Modifications that could be made to a routine/command in a routine in response to a conflict with another routine.

Id	Possible Modifications
M-1	The entire routine stops or does not start.
M-2	All associated commands that have not yet executed of the routine execute now.
M-3	All associated commands of the routine that have not yet executed execute with a delay to allow the other routine to execute.
M-4	The command that attempts to access the same device as the other routine (known as the *conflicting command*) does not execute.
M-5	The conflicting command executes with a delay, to allow the other conflicting command to execute.
M-6*	Any commands that have already executed are reverted (meaning that these commands are reversed), and all associated commands of the routine execute now.
M-7*	Any commands that have already executed are reverted, and all associated commands of the routine execute with a delay to allow the other routine to execute.
M-8*	Any commands that have already executed are reverted, and the conflicting command does not execute.
M-9*	Any commands that have already executed are reverted, and the conflicting command executes with a delay, to allow the other conflicting command to execute.

* Modifications 6-9 take the same approaches as modifications 2-5, but first revert any associated commands that have already executed for the routine, meaning that the commands that have "not yet executed" refer to the entire routine rather than only part of it.

routine 1 by the nine to routine 2 results in a set of 81 possible strategies. However, some of these strategies were duplicates of some others in this set or were impossible to execute in practice. For example, both routines cannot apply M-5 at the same time, otherwise, the two routines would be waiting for the other to execute indefinitely. The full list of strategy candidates is shown in Appendix A.

After removing these duplicates and impossible strategies, there were 27 total possible strategies that a system could use to address a routine conflict. In our survey however, we presented only 10 of these strategies, based on pilot survey responses, to reduce the cognitive load on participants while preserving our original goal of seeing how different system strategies met user expectations (detailed in Sec. 4.1.2).

4 SURVEY METHODOLOGY

We used the results of our system strategies analysis (Section 3) to inform the design of an online survey to answer **RQ2** and **RQ3**. We collected $N = 116$ valid survey responses via deployment on Amazon Mechanical Turk (MTurk) in September of 2021[2].

Participants first completed a short screening survey (median 1.6 minutes to complete), which elicited demographic information (including race, gender, education level and living situation) and details on their prior experience with smart home devices. The screening survey was available to all MTurk workers (MTurkers). Screening survey respondents who indicated that they were at least 18 years old and had experience with at least

[2]All study procedures were approved by our institution's Institutional Review Board (IRB) and piloted with other members of our lab.

Table 3. Summary of Participant Demographics.

Age	18-24 (3%), 25-34 (51%), 35-44 (27%), 45-54 (10%), 55-64 (7%), 65+ (1%)
Education	High School or GED (9%), Some College (5%), Associate's Degree (1%), Bachelor's Degree (70%), Post-Graduate Degree (15%)
Gender	Male (67%), Female (33%)
Race	Black/African American (15%), White (78%), Asian (5%), American Indian/Alaskan (2%)
Living Situation	House (62%), Apartment (38%)
Number of Smart Devices Owned	1-3 (59%), 4-6 (36%), 7+ (5%)

one smart device[3] were invited to participate in the main survey (median 15.6 minutes to complete). Participants were paid $0.50 USD for completing the demographic survey and $3.50 USD for completing the main survey, which is approximately $11 USD per hour. An overview of participant demographics is included in Table 3.

The first page of the main survey provided definitions and examples of routines, commands, and routines conflicts to participants. Participants were then asked to read scenarios involving a smart home routine conflict. For each scenario, they were presented with possible outcomes of the scenario and asked to rate their satisfaction with each outcome on a 5-point Likert-Scale. Additionally, participants were encouraged to explain their reasoning by answering an optional free response question. Based on pilot study feedback as well as previous work, we sought to reduce participant cognitive load by keeping the main survey length under 20 minutes [31, 70, 73, 93]. Thus, we limited the survey to include four randomly-selected scenarios, presented in a random order, out of the full set of 20 scenarios we used in our study. In addition, we controlled for participants' living situation (apartment vs house) to adjust the scenarios shown to ensure that participants were shown relevant scenarios. To prevent participants from seeing scenarios that were too similar to each other, such as only seeing scenarios related to certain smart devices, scenarios were grouped together by the smart devices they included. If a question from a group was randomly selected to be shown to a participant, then no other questions from the same group would be shown. Details on the 20 scenarios and their possible conflict resolution outcomes are described in Sections 4.1 and 4.1.2, respectively. Before beginning the hypothetical scenario portion of the survey, participants were also to determine if a pair of commands were necessary for a routine to accomplish a specified goal, and to rank their preferred end states of devices involved in a routine. These questions were simple and meant to reinforce participant understanding of routines and commands before answering the scenario-based questions. Examples of these questions are shown in Appendix D.

4.1 Scenario Design

Each scenario consisted of a description of two routines conflicting on one device in a smart home and included various relevant details about the scenario, such as the time of day, what the user is doing, and the cause of the conflict. The routines in each scenario were ordered, meaning the start of routine 2 always caused a conflict with an already-executing routine 1. Thus, one scenario could contain the same routines as another, but the order of

[3]Defined as Smart plugs/switches, Smart lights, Smart locks, Smart thermostats, Smart Vacuums, Smart Windows/Blinds, Motion Sensors, Smart smoke alarms, Smart security systems/cameras, Smart TV, Smart Speakers, or Other (participants could list other devices and the research team would determine if they were valid)

these routines could be flipped, which would make these two scenarios unique. The full list of scenarios used in our survey appears in Appendix B.

The research team iterated for several weeks, developing a set of scenarios to present to participants. We focused on building realistic scenarios that participants could imagine being in without significant cognitive load [60], while also presenting a diverse set of user experiences. We had planned to send surveys to around 300 participants, and with each user only seeing a subset of these scenarios, we concluded that a set of 20 scenarios would achieve our diverse and realistic scenario goals while also allowing each scenario to be seen by an adequate number of participants.

Table 4. Involved devices, product existence, and routine source. The products or devices we covered all support remote control such as through phone or smart hub) and/or are designed with potentially regular usage patterns (e.g., smart trash can moving outside every week)

Device type	**On market**	**Example**	**Main routine source**
Smart cooktop	Yes	Samsung Smart Cooktop[52]	SmartThings Community
Smart thermostat	Yes	Google Nest[81]	Google Nest Community[83], IoTBench[54]
Smart window	Somewhat	Fenestra[26] (Prototype), Air-Lux[5] (voice control)	Online posts/blogs/tutorials
Smart light	Yes	Philips Hue[48], Lutron[72]	IoTBench, Community posts
Smart speaker	Yes	Amazon Echo, Google Nest[81]	IoTBench, Community posts
Smart lock	Yes	Yale Lock[44], myQ[80]	IoTBench, Community posts
Smart garage door	Yes	myQ[80], Wyze[53]	IoTBench
Smart doorbell	Yes	Nest Doorbell, SimpliSafe[100]	IoTBench, Community posts
Smart water sprinkler	Yes	Rachio[88], Imolaza[50]	Online posts/blogs/tutorials
Smart Trash can	Yes	Rezzi [8]	Online posts/blogs/tutorials

4.1.1 Routine Selection. To identify candidate routines, we extracted routines from several data sources including IoTBench [54], Samsung SmartThings [94] forums, and other platforms [45, 63]. Table 4 describes the list of device types included in chosen routines and whether there are products in the market that already support such devices. We found that there was a wide range of devices that support remote control with people enthusiastically attempting to use and integrate them into their own smart homes. For example, Samsung smart cooktop [52] users can monitor the operating status and power level setting of the cooktop elements, as well as check and change cooktop timers. Users were actively using these [75, 85] and attempting to create routines with this device and its various features [42].

Routines were added to this list based on how often they were encountered in our analysis. Routines were also more likely to be added to the list if they could be expressed as a long running routine, as these types of routines are more likely to conflict with other routines (Sec. 2). This list was then used to generate scenarios presented in the survey.

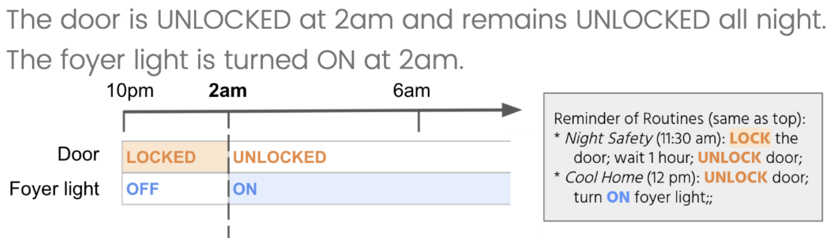


Fig. 1. An example of a textual description and visual timeline of a conflict between routines outcome in our main survey. Each smart device involved is listed on the left, with critical times marked on top. Each device is assigned a color: orange to the door, blue to the foyer light, with each of the two possible states (`LOCKED`/`UNLOCKED`, `ON`/`OFF`) of the device shown by whether the region is filled or not. This timeline details a scenario starting at 10 pm where the door will be `LOCKED` and the foyer light will be `OFF` until 2 am when the door is `UNLOCKED` and the foyer light turns `ON`.

Four of the 20 scenarios involved [situations/routines] were not applicable to apartment living (e.g., putting a trash bin out to the curb or turning a sprinkler system on or off). Thus, these scenarios were excluded before the random selection of four scenarios was made for the 75 participants (38%) who indicated in the screening survey that they lived in an apartment rather than a house, so no participant would have to imagine a scenario that would be impossible for their current living situation. Scenarios were presented with several potential outcomes, which we will detail in the next section.

4.1.2 Possible Outcomes. Each scenario in our survey had several ***outcomes***, which showed the device states over time after the routines are executed following some system ***strategy***. We presented each outcome to participants in both a written description and via a visual timeline, as shown by the example in Fig. 1.

Initially, we had 27 outcomes, based on the 27 system strategies identified in Sec. 3. The list of these strategies can be seen in Appendix A. However, only a subset of these strategies was seen in previous work. Furthermore, the cognitive load on participants to try and parse this long list of outcomes for each scenario when attempting to complete the survey would be extremely high. Therefore, the research team decided to include only a subset of 10 of these strategies in the survey (Table 5), which was the result of consensus building between the research team. The consensus building took place over two months with several meetings between the research team, where the leading two authors would present a set of strategies and discuss them in order to collect feedback. This feedback was used to improve the set of strategies, which was presented and evaluated again by the research team in future consensus building sessions. After several sessions, the research team converged on the set of strategies used in the survey.

The set of strategies we picked in Table 5 was also validated by our pilot study, whose feedback indicated that the outcomes presented to participants were almost always sufficient, meaning there was no other outcome that they felt could be satisfying. Thus, we concluded that our selected system strategies were sufficient for describing preferred outcomes in most cases. After eliminating duplicates, each of the 20 scenarios had between five and nine unique associated outcomes, as some strategies, when applied to scenarios led to the same outcome. Consider two routines: “Welcome Home”, which opens the blinds and turns on the coffee machine in the house when a user opens the front door, and “Night Time”, which closes the blinds and turns off the lights in the house at 8 pm. Imagine a routine conflict where a user enters the home at 8 pm, resulting in a situation where the two routines request that the blinds be both open and closed at the same time. The system could apply the *Cancel* and *Ignore* strategies in this situation, which cancels the entire “Night Time" routine and ignores the conflicting command on the blinds. However, the outcomes of both of these strategies are the blinds being open and the coffee machine running. To the user, these outcomes will be the same; in these cases, we show the outcome

Table 5. System strategies, along with their acronym, behavior, granularity, and the routine that they affect.

Strategy	Short	System Behavior	Granularity	Source*
Cancel	CC	Cancel R2 (the routine that introduces the conflict). Finish the rest of R1 (the routine that was running before R2).	Routine	Google Home
Delay	DL	Start executing R2 after R1 finishes.	Routine	SafeHome
Execute**	EX	Put all commands in time schedule and execute as it is.	-	Google Home
Ignore	IN	Cancel the CD (conflicting smart device) commands in R2. Then start executing the rest of R1 and R2 as time schedule.	Command	Other
Overwrite	OW	Cancel the CD commands in R1. Then start executing R2 and the rest of R1 as time schedule.	Command	Other
Pause	PZ	Pause R1. Resume R1 after R2 finishes.	Routine	Alexa
Resume	RZ	Put all commands in time schedule and keep executing. After R2 finishes, if there are commands for CD in R1, set CD back to the last state that R1 set it to in the last command. Continues the time schedule.	Command	Other
Revert	RV	Set all devices to the states before R1 started; Execute R2. Afterwards, re-execute R1.	Routine	SafeHome
Stretch	ST	Delay the CD command in R2 after R1 finishes.	Command	SafeHome
Truncate	TC	Start executing R2 directly. Cancel all un-executed commands in R1.	Routine	Other

*Example system that can support this strategy with user choice of system mode, user intervention, or user programming. "Other" refers to a strategy that was not observed in existing systems, but was a result of our consensus building.
**This is the status quo strategy that most commercial smart home system like Google Home is using.

only once and include the participant's satisfaction rating in our analysis for both strategies. For each scenario, participants were asked to rate their satisfaction with each outcome from 1 (unsatisfied) to 5 (satisfied).

4.1.3 Data Analysis. Our analysis is centered on: (1) Likert-scale responses for participants' reported satisfaction with a particular outcome in a particular scenario, and (2) optional free-response answers given to explain a given satisfaction rating.

The Likert-scale satisfaction ratings were used to weigh how well each outcome met a participant's expectations. Satisfaction has been defined as "the fulfillment of a specified desire or goal" [56]; the specified goal is to have smart homes meet user expectations. Studies of user satisfaction are done in several ways and are conducted in many contexts, including smart homes [17, 47, 115]. Thus, our data analysis looked at the satisfaction ratings as a way to measure the effectiveness of each routine conflict resolution strategy in a scenario.

Because satisfaction ratings are ordinal, we used an Ordinal Logistic Regression (OLR) model to examine relationships in the data. OLR is a statistical analysis method that models the relationship between an ordinal response variable and one or more explanatory variables [114]. In our study, the ordinal response is our five-point Likert-Scale satisfaction score, and explanatory variables are each system strategy and scenario.

We also calculated basic parametric summary statistics such as the mean, median, and standard deviation of these ratings to get an overview of the data. While such statistics are thought to be unnecessary for ordinal data, research has shown that robust parametric statistics can still be extracted from Likert-Scales [84].

We used a hierarchical Bayesian model with weakly informative priors to estimate the one-way effects of strategy and scenario, as well as the interaction effects between the two factors. Compared to a non-hierarchical regression, our model is more regularized, which is desirable in smaller-N studies, and addresses issues of multiple comparisons, since we are investigating reported satisfaction across system strategies as well as scenarios [34]. In addition, we used a mixed-effect model to reduce the random effect caused by the difference in rating dispositions among survey takers, ie. consistently rating outcomes high or low.

Free response question responses were coded using open coding for qualitative analysis [57]. During the first round of coding, we tried to extract overarching themes from the data; however, during this process, we found that many participants expressed personal *values* in their reasoning for rating outcomes. Thus, we investigated further by coding responses to uncover the underlying values the participant seemed to have when it came to their expectations of how smart devices in the home should respond to routine conflicts. One member of the research team generated these codes and received several rounds of feedback from the rest of the team, working to reach a consensus on a descriptive set of qualitative codes. These values were either 1) *moral* [2] or 2) *utilitarian* [7].

The quantitative analysis helped answer **RQ2** by looking at the relationships between different contexts of smart home routine conflicts, the strategies used to resolve those conflicts, and participants' satisfaction with the outcomes produced by these strategies. We then extracted themes relating to what participants consider when forming their expectations of smart home routines during conflicts through qualitative analysis of written responses, allowing us to answer **RQ3**. The results of the analysis are detailed in Sec. 5.

4.2 Data Quality

Amazon Mechanical Turk is a well-established source for collecting crowdsourced data [51], able to collect large amounts of potentially generalizable data [89]. In addition, we instituted several measures in our data collection to improve the overall quality of data used in our analysis (described in the previous section). Participants were compensated, but their data was not included in our analysis if they failed any of these measures. Two simple arithmetic questions were placed at different points in the survey to serve as "attention checks" [1]. In addition, participants' responses were discarded if they took less than five minutes, which was one-third of the mean completion time for pilot participants. Finally, at the end of the survey, we asked participants to verify demographic data from the screening survey; if there was a mismatch, the response was not included.

To ensure we included only high-quality responses in our analysis, we conducted validity checks and excluded 81 total survey responses from the data analysis (but still compensated the participant). The conditions for exclusion were if the participant: (1) completed the survey in under five minutes (one-third of the mean pilot survey completion time), (2) answered either of two attention check questions incorrectly, or (3) gave demographic responses in the main survey that did not match with those from the screening survey (this check was done manually to evaluate the plausibility of life changes of participants between the two surveys).

The main survey stage was sent to 297 potential participants after applying our inclusion criterion of having experience with a smart device. 197 participants responded to this solicitation and provided full responses to the survey, 116 of which were considered "valid", meaning they were used in our data analysis. For the invalid responses (responses not used in our analysis but were still compensated), 26 were completed in less than five minutes and the other 55 invalid responses were due to inconsistent demographic information between the screening and main survey. All participants passed the attention check.

4.3 Limitations and Tradeoffs

We made several tradeoffs in our main survey design and distribution. Survey questions set in hypothetical settings, such as the ones used in our survey, are frequently used and are deemed as a valued and accepted form of inquiry [10, 118]. While using a method that reported in-situ behavior would have higher ecological validity [14], such studies have a more limited set of potential participants to recruit from since they are limited by who is local to where the study is being conducted. Using the MTurk platform, our participant pool was not limited by proximity to the research team.

The specific routine pairs used in scenarios presented to participants, while realistic, were not directly derived from real-world observations of routine pairs. This limits some of the generalizability of the results, as the exact conditions outlined in the scenarios may not be very common. However, as described in Sec. 4.1, the individual device routines in each scenario were observed in real life and are commonly used (Table 4). Thus, even though the scenarios in our survey were hypothetical, they had a strong connection to real-world experiences that participants and users can have. In addition, the individual routines would be easy to imagine for participants who had some experience or interest in smart home devices, contributing to external validity. The research space of routine pairs is limited, meaning that there was no set of routine pairs that had been established as common amongst smart home users that we could use in our survey. We prioritized using diverse sets of routines in pairs to offer a wide range of experiences that participants could encounter in order to address external validity.

While our survey was primarily focused on user satisfaction with outcomes produced by routine conflict resolution strategies, this is not the only way to measure the effectiveness of these strategies. For example, the complexity of a conflict resolution strategy may contribute to user perceptions of that strategy. However, it is infeasible to capture all dimensions of user perception of smart home routine conflict resolution strategies, especially without an in-situ observational study.

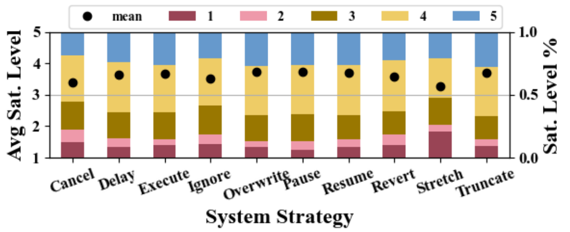
Since our free-response question involving participants explaining their reasoning behind their ratings was optional, the characteristics of participants who answered it may have been different from the overall participant population in important ways. To provide some insight into the possible biases this self-selection may have introduced, we report demographic information about the subset of participants who wrote valid free-response answers (Section 5.4). The purpose of the free response questions was not to get large numbers of quantitative generalizable results, but rather to understand the depth and nuance behind what shapes individual user satisfaction with different conflict resolution techniques.

The MTurk platform allows for the collection of large quantities of data, but not necessarily high-quality data [16]. Pre-screening participants based on platform performance (ie number of HITs completed and HIT acceptance rate) [41] is a common practice to achieve high data quality; instead of this, we included several validity checks as described in Sec. 4.2, which also let us reach a broader audience than pre-screening criteria like those in prior work [41].

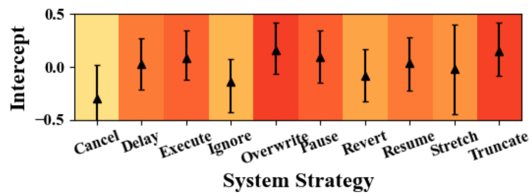
Another potential weakness of MTurk is the likely skewed demographic of its workers in terms of technological literacy. Given the large number of participants in our study with Bachelor's degrees (see Table 3 for participant demographics), we see that our sample may be skewed towards those with higher education and technological know-how [58], especially in the context of smart homes and devices [95, 99, 108]. However, this may be due to trends in smart device ownership rather than the MTurk platform. In general, smart device ownership is skewed in age, gender, and socioeconomic status [102]. In addition, participants had to either already own a smart device, or express an interest in smart devices if they did not already own one. This allowed us to capture a larger portion of the population, with results that could generalize to potential users, not only existing users. Smart device usage is not a prerequisite for a user to know what is satisfactory to them and what is not. Thus, there is value in soliciting feedback from both those with a broad interest in smart devices and actual device owners. While a more tech-savvy population would provide more data at this point in time, we acknowledge that design decisions that cater to only a fraction of the population may not capture the needs, budgets, and values of all future users, resulting in alienation [36].

5 RESULTS

We analyzed our results to help answer **RQ2** and **RQ3**, which investigated how different existing approaches to handling smart home routine conflicts addressed user expectations, and what factors influenced these expectations.



(a) *Average satisfaction levels across all scenarios. Each stacked bar shows the satisfaction level distribution for each strategy, and the black dot shows the mean value.*



(b) *The intercepts of system strategies; Error bars show 95% credible intervals.The darker colors correspond to a higher likelihood that participants are satisfied with outcomes produced by these system strategies in general.*

Fig. 2. Satisfaction level statistics for different system outcomes. In general, the satisfaction levels among different outcomes produced by strategies were not significantly different from each other. The top 5 outcomes based on average satisfaction level were strategies produced by: Truncate, Execute, Pause, Overwrite, and Resume. Outcomes produced by the Cancel strategy had the lowest average satisfaction level.

5.1 Demographic Distributions

Of the 116 valid responses (i.e., participants that answered the demographic survey as well as provided a response to the main survey that passed the validity checks described in Sec. 4), 71 were from participants whose living situation was classified as a "house," while the other 45 came from participants with "apartment" style living situations. Overall, the participant sample was similar to the US racial distribution after filtering participants who had failed at least one of the validity checks.. The 116 participants were primarily male (67%) and highly educated compared to the US population, as over 70% had obtained a Bachelor's degree.

5.2 Reported Satisfaction

As introduced in Sec. 4, each participant rated four scenarios and up to 36 outcomes for each scenario (corresponding to the 10 system strategies which sometimes led to overlapping outcomes). Due to the ordinal nature of our Likert-scale ratings, we conducted an Ordinal Linear Regression (OLR) analysis to try to identify the strategies that performed best within each scenario, and overall.

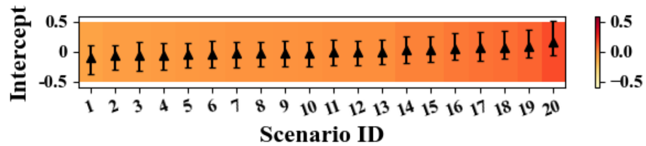
5.2.1 Strategy Preference across All Scenarios. From the aggregate of all 3,136 participant ratings across all scenarios and system outcomes, ratings tended to be positive; 63% were either *Slightly Satisfied* or *Satisfied*. However, it was also not unusual for participants to consider an outcome bad. This occurred in 17% of all ratings, where participants reported being *Unsatisfied* or *Slightly Unsatisfied* with certain system outcomes. This motivated further investigation into how users' expectations were formed, so that design suggestions could be made to help future systems avoid this high percentage of "bad" outcomes.

This general trend of positive ratings was reinforced by central tendency measures of outcome-specific ratings. Of the 118 possible outcomes that were rated by participants, only 22 (15%) had a median rating of less than *Slightly Satisfied* (4). The mean satisfaction levels for outcomes grouped by the system strategy used to create it were also all between *Neutral* (3) and *Slightly Satisfied* (4) (Fig. 2a).

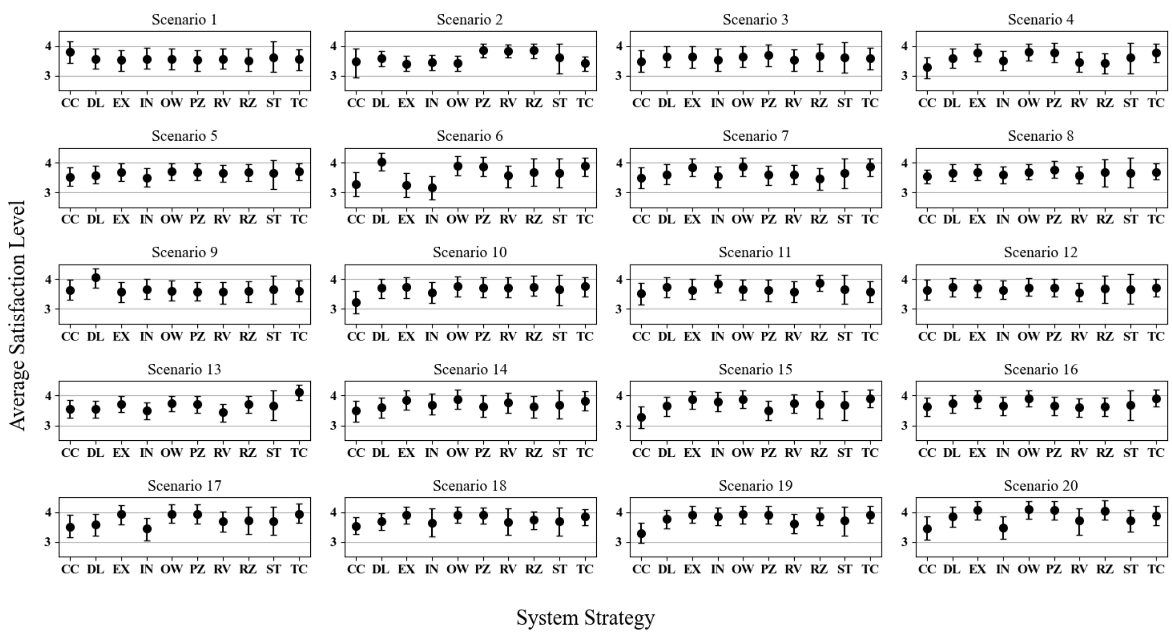
We also grouped results by the system strategy that produced each outcome. The Truncate and Overwrite strategies produced system outcomes with the highest average satisfaction levels and OLR intercepts. The Cancel strategy, in which the system cancels the newly triggered routine in favor of keeping the existing one running, produced outcomes that received the lowest average satisfaction ratings. While we cannot say with certainty, participants may have considered the most recently triggered routine more urgent. Pausing the running routine/command and resuming it after the most recently triggered routine finishes (Pause/Resume strategies) created systems outcomes that received positive satisfaction ratings compared to outcomes produced by the rest of the strategies. Given our results in Sec. 5.2.2, further investigation is needed, especially into the role that context plays in user satisfaction.

We found no single "best" strategy; the credible intervals of intercepts for most strategies overlapped (Fig. 2b). This means that for all strategies, the range of reported satisfaction with the produced outcome were similar to each other. The credible intervals based on the different values for each parameter (scenario and system strategy), can be found in Appendix C.

5.2.2 Scenario-Dependent Satisfaction Levels. The OLR model revealed the existence of a relationship between the scenario presented to the participant and their reported satisfaction with different outcomes (Fig. 3a). Intuitively, the scenarios presented to participants were different. They involved different devices, took place at different times of the day, placed the user in different situations, and more. The regression model revealed a statistically



(a) *The intercepts of the OLR model of scenarios versus reported satisfaction with an outcome. The scenarios are ordered based on the mean value of the intercept. If a scenario is in a darker color, its intercept is higher, meaning that in general, the scenario is more likely to have a satisfactory outcome for the user regardless of which outcomes are present.*



(b) *Estimated average satisfaction grouped by scenario. Within each scenario, we group by system strategy. The error bars represent 95% credible intervals.*

Fig. 3. Ordinal Linear Regression results for scenario intercepts and estimated satisfaction level under the interacted factors of system strategy and scenario.

significant difference between participants' satisfaction with outcomes produced by the same strategy depending on which scenario was presented.

As shown in Fig. 3b, participant satisfaction with an outcome produced by a given strategy varies across scenarios. For example, the Truncate strategy produces a highly satisfactory outcome in Scenario 13, but fails to do so in Scenario 9. Fig. 3b also shows other results from our analysis of the interaction effect between scenarios and outcomes produced by different system strategies.

Some of the scenarios included components that made some outcomes clearly more desirable than others. For example, Fig. 3b shows how the Delay strategy outperforms the other strategies presented in the scenario. In nearly half of the scenarios, participants expressed clear preferences with respect to outcomes, meaning there was a single outcome that received significantly more high ratings than others. For example, in Scenario 2 (Scenario IDs are listed in the Appendix), the outcome where participants could get into their home and have the door lock 5 minutes later (produced by the Revert, Resume, and Pause strategies) was rated as *Slightly Satisfied* (4) or higher more than 80% of the time (61/74 ratings). The outcome which kept the door locked overnight (produced by the Delay strategy) was the next closest outcome, receiving a *Slightly Satisfied* (4) or higher rating only 65% of the time (48/74 ratings).

5.3 Regression Model Implications

Our data (Fig. 3b) did not provide evidence for the existence of a universally acceptable strategy. Part of the motivation of **RQ2** was to investigate the existence of strategy patterns that produce favorable outcomes in a variety of scenarios. We observed relationships between the scenario presented to participants and their reported satisfaction, as well as the outcomes produced by system strategies presented to participants and their reported satisfaction (see Fig. 3a and Fig. 2b).

We also observed that individual participants had distinct preferences from one another that carried over between their ratings of different scenarios. For example, the model estimates a large difference in the estimated user satisfaction with applying the Overwrite strategy as compared to the Revert strategy in Scenario 2, with Revert having a much higher estimated user satisfaction. However, in Scenario 4, this is flipped, with the model estimating a much higher satisfaction with the Overwrite strategy compared to that estimated by the Revert strategy. The model's estimations revealed a lack of consistent preferences among strategies throughout scenarios.

While our analysis of ratings provided substantial insights such as the relationships between scenarios (contexts)/system strategies and reported satisfaction, much of the variation in reported participant satisfaction could not be explained solely by our model, which had an R^2 value of 0.03. Thus, to supplement these quantitative results, we also analyzed participants' free-response explanations in order to gain further insight into what factors were most relevant to their decisions.

5.4 Factors Influencing Participant Expectations

A subset of 37 participants answered our optional request to "Please explain your reasoning for your preferences" (148 total relevant responses, since participants could include explanations for more than one of the scenarios they saw). Scenarios 2 and 5, elicited the most participant responses with 17 unique participants writing in answers. Home safety was a repeated theme in both of these scenarios. The overall demographics of these participants were similar to that of the 116 valid respondents.

Table 6 shows example responses for each code. In half of these responses, participants indicated that they value feeling comfortable in their own home (coded as "Comfort" in Table 6). This aligned with previous research that rated comfort and/or convenience as the key factor users consider when deciding to use a smart home [9].

Other values that appeared in the qualitative coding of responses were: "Personal Schedule," which signaled that the participant values their daily regimen and wants their smart home to reflect this value, and "Safety,"

Table 6. Qualitative Codes

Code (Value)	Meaning	# Responses with this value	Example Response
Personal Schedule	Participant values their personal schedule/regimen	30 (20.27%)	"I don't like when the music plays before it is supposed to. I also want to make sure... that I can enjoy my coffee when I wake up."
Safety	Participant value feeling safe in their home	14 (9.46%)	"I feel that in this case safety is first... I would definitely want the door to be locked in the middle of the night."
Comfort	Participant values feeling comfortable in their home	78 (52.70%)	"On that time I feel more comfortable with music than news. So I choose [to] turn on [the] coffee maker and play music."
Conserving Resources	Participant value saving some kind of resources (electricity, water, money, etc.)	11 (7.43%)	"I prefer to have the AC turned off until all the windows are closed, it's a waste of energy if the AC stays on while the windows are open."
Responsibility	Participant values their career/responsibilities	15 (10.14%)	"I prefer the system to postpone the news until the end of the meeting time, so that there are no interruptions."

where participants emphasized the value of feeling safe in their home. Participants also wrote about the value of ensuring smart devices do not interfere with their obligations, reflecting the value of "Responsibilities/Work." The results indicated that a distinct set of values influenced participants' decision-making when choosing their satisfaction levels with outcomes. Previous work indicated that values influence smart home users, and we observed this influence in our study results [77].

The contexts or scenarios elicited these different values from participants. For example, in Scenario 1, P2 expressed higher levels of satisfaction with outcomes that limited the functionality of the second routine, stating that "*I would hate being interrupted especially if [the meeting] is work related.*" However, when a work meeting was interrupting an already running routine, as in Scenario 17, P12 rated outcomes that interrupted their running routine to make way for their work meeting highly, saying that "*I would insist that the speaker be muted during the duration of the meeting.*" Both participants expressed similar values and based on the specific scenario highly rated outcomes that achieved a similar result of the work meeting taking precedence.

Some participants seeing the same scenario mentioned the same value in their free responses. In these cases, the participants often gave similar ratings to the various outcomes. For example, in response to Scenario 20, two participants both described their value of conserving resources and avoiding energy waste in their homes. They wrote, "*I'm concerned with avoiding energy waste,*" (P49) and "*it's a waste of energy if the AC stays on while the windows are open*" (P71). Both P49 and P71 gave low ratings to outcomes that kept the window open while the AC was running.

However, participants who conveyed the same values in their written responses to the same scenario sometimes rated the outcomes of that scenario differently. For example, in scenario 18, P84 and P111 expressed the value of their "personal schedule." P84 wrote that the second routine in the scenario "*already aligns with my practiced [schedule].*" P111 stated they did not like when the first routine played music "*before it was supposed to,*" and that they wanted to ensure they "*can enjoy my coffee right when I wake up.*" They did not like outcomes that deviated from their schedules, yet the details of their existing schedules affected which outcomes they preferred. P84 and P111 expressed similar values but had significantly different outcome preferences. Both scenarios that P84 rated as *Satisfied* (5) and *Slightly Satisfied* (4) were rated as *Slightly Unsatisfied* (2) by P111. In contrast, the highest rated outcome by P111, rated as *Slightly Satisfied* (4), was only rated as *Neutral* (3) by P84.

6 DISCUSSION

The results from our survey and analysis of existing systems laid out several findings and promising directions for future research.

6.1 Finding 1: Both the Context of a Routine Conflict Scenario and the Strategy Used to Resolve the Conflict Affected Users' Satisfaction.

The results of the regression model indicated that both the scenario presented to the participant and the system strategy used to get to a specific outcome in the scenario had a statistically significant impact on the satisfaction level of the participant with that outcome (Section 5.2). Participants considered the situation where the conflict occurs, as well as how the system resolved the conflict when rating how the system met their expectations. This finding suggests that existing systems should move away from one-size-fits-all solutions.

Users encounter different situations with their smart devices on a day-to-day basis. Previous work has stated that users find it difficult to write effective routines when the environment of their smart home is changing or unknown. If users cannot express routines that maintain their desires in changing contexts, it is even more important for system designers to create flexible routines to meet user expectations at a satisfactory level. Furthermore, these routines should be easy to understand and validate from a user perspective, and not be black boxes [65]. Context-agnostic approaches (mentioned in Section 2) that handle routine conflicts in a specific manner regardless of the specifics of the situation, center around the system and offer system-centric approaches. Our results pointed out the necessity and benefit of shifting to a more user-centric approach in developing systems to handle smart home routine conflicts, and that a context-agnostic approach to routine conflicts is not sufficient.

In the *FireAlarm* and *LockTheDoor* routine examples described in Section 1, the Cancel strategy may be appropriate based on user expectations, but if the order of these routines is switched, then the Truncate strategy would likely be preferred. There are different variables (e.g., cultural, social, professional, medical, etc.) that affect a specific scenario, person, and family. While we have not investigated the role that each of these specific variables plays in outcomes meeting user expectations, the results indicated relationships that should be explored further in future work.

This shift to a more user-centered focus in routine conflict handling requires an understanding of what influences users' formation of these expectations upfront. This was the goal of answering **RQ3**, and the results pointed to one possible source of influence.

6.2 Finding 2: Users' Personal Values Shaped Their Expectations for Routine Conflict Resolution, Leading to Different Sets of System Strategy Preferences across Scenarios.

The results from coding participant responses (Section 5.4) brought to light several values that played a key role in their expectations and desires for smart device behavior in their home. We were not comparing the frequency of different values, but rather looking at them individually and explaining how they influenced participants in their expectations of smart devices.

Across many scenarios, participants expressed the value of "comfort". The promise of additional convenience or comfort in the home is what drives many users to purchase smart devices in the first place [15]. Some participants were not willing to compromise on comfort and thus rated strategies that prioritized their comfort when dealing with conflicts.

Participants indicated a desire for a smart home to integrate into their existing life and automate tasks that are already a part of their routine (rather than them shaping their life into the smart home ecosystem). This connects back to Davidoff et al.'s [22] work stating that users want to control "their lives" rather than devices. Even in the face of unanticipated situations such as the routine conflicts presented in the survey, participants preferred not to disrupt their original schedule. System strategies that preserved participant control by maintaining their personal schedule were looked at favorably.

Our results also indicated that the value of fulfilling participants' work or other responsibilities plays a role in their expectations. If the scenario involved the user virtually attending a meeting, participants leaned heavily towards system outcomes that would not jeopardize their meeting attendance or co-worker perceptions. With the recent rise in "working-from-home" [86], smart home devices should understand users' professional priorities [74].

Some of our scenarios challenged the participant's safety, in the form of conflicts between routines that affected whether their front door was locked. In these scenarios, recall that nearly all participants expressed the value of safety in their justifications. When routine conflicts threaten their safety, our participants preferred strategies that prioritize keeping the smart home in a state that will maximize their safety, such as a strategy that ensures the front door is locked. While these scenarios involved an exterior door needing to be locked to keep potential invaders out, there are many other types of smart home devices and routines not included in our study but could have implications for safety if and when they are involved in conflicts. For example, smart smoke/carbon monoxide detectors are becoming increasingly common and can be referenced in routines [82]. Based on our results, if a situation arises where there is a routine conflict that can potentially affect how the smoke detector would function, users would likely prefer strategies that would not affect the smoke detector. Follow-up studies focusing primarily on house safety and user preferences for smart routine conflicts are needed to make definitive statements, but our results signaled the importance of safety as a user value.

Although privacy did not surface as a value in our study, substantial prior work demonstrates its importance [69, 119]. This work has laid out designer guidelines with respect to privacy for smart home system designers [69]. There is a "trade-off" between privacy and autonomy in smart devices, as users are willing to give up some level of their privacy in exchange for additional autonomy via a smart device [104, 119]. We observed values related to convenience ("Comfort") and personal control ("Routine" or "Responsibilities") in the open coding of responses, which can sometimes come at the cost of privacy.

Several of the values observed in user responses, such as safety, may be considered as "norms", which are widely acceptable across groups of users [43]. Examples include socially-conscious values such as energy savings, security, etc. On the other hand, values such as "personal schedule" are more user-specific. System designers can shift to a more user-centric design by prioritizing norms and additionally cater to user-specific values, as these values are the result of users' lived experiences, and smart devices can acknowledge these values to become more *user-centric*. However, this comes with the caveat that two users with the same values may choose differently in the same scenario, meaning future research should look at this trade-off between context-agnostic and user-centric system strategies [78]. One such direction is to allow users to directly express their most important values to the system, giving the system information on how to proceed in the face of routine conflicts to stay in accordance with a user's values.

Designers of operating systems (e.g., Linux, MacOS, etc.) attempt to design generalizable systems over which a variety of systems can be built. Designers of smart home management systems (e.g., Alexa, Google Home, Samsung Smart Things, etc.) have been steadily moving towards an analogous goal. Our study's result is important in showing which specific kinds of human factors need to be taken into account while designing these smart home management systems of the future.

6.3 Finding 3: There Are Two Significant Design Principles Smart Home System Designed Should Consider to Ensure a Quality User Experience.

From the results of our study which evaluated user satisfaction levels of different system approaches to conflict handling, we propose two main design principles:

First, *a smart home routine management system should attempt to embed user values because they directly affect user preferences.* In Sec. 6.2, we discussed how users often apply "general" values to their situation, making them "domain-specific" [111]. In this case, we saw how the general value of comfort or safety, which is something a person would want regardless of whether they live in a smart home or not, was applied to the context of smart home routines and formed people's expectations of how the system should respond to routine conflicts.

A smart home routine management system that allows users to express these values has the potential to better accommodate user expectations, resulting in higher user satisfaction. Such a system would place people at the center, provide users with potentially more control, less labor, and result in a system with a better understanding of how to execute routines so that they reflect people's values [27].

This design principle helps mark a shift to a more personalized smart home routine management system. Different user expectations shaped by similar values do not necessarily lead to similar views on which strategies best met those expectations, as shown in Section 5.4. However, systems that capture user values and attempt to preserve them in their handling of routine conflicts can make meaningful *progress* towards choosing strategies that handle these conflicts. While our research points out this direction, the question of *how* to best capture these values is still an active research question and one that should be explored in future work.

Second, *a smart home routine management system should be flexible in order to handle diverse user preferences.* The values of one user can differ significantly from that of another user. The results of our qualitative coding revealed that within a single scenario, different values were elicited from participants. These different values further led to multiple outcomes in a scenario being rated highly. This suggests that a smart home should be able to manage routines by implementing multiple strategies, rather than trying to apply one single strategy across situations in order to satisfy all users. A smart home routine management system should be flexible enough to accommodate multiple, diverse sets of co-existing preferences, and apply them to conflict scenarios to achieve a satisfactory outcome for the multiple users involved. From our survey results, we observed that users indicated valuing comfort, safety, adherence to a personal schedule, and more.

6.4 Future Work

The focus of this paper was to uncover user expectations around smart device routine execution systems in the face of conflicts. In the future, an in-person study placing people in a smart home would be a worthwhile extension of this work. Understanding how participants in a real environment behave around routine conflicts and their outcomes, ideally in their own homes would add ecological validity to a study and its findings.

Our scenarios varied along many dimensions, but this variance was not systematic, as we were mainly focused on providing a diverse set of realistic experiences in smart homes. The possible variations in time of day, smart devices involved, and what the user is currently doing for each scenario all have the potential to impact user expectations for that situation. However, given that no formal conclusions can be made, a larger study that controls these dimensions closely, such as presenting scenarios that only use certain smart devices, may help to uncover if there are "scenario-agnostic" strategies that users may universally prefer. These future studies can help to understand the impact that each of these, and other dimensions can have on user expectations for smart devices and routines when conflicts occur.

Our survey focused on a single person living in a smart home. However, homes with multiple users present an even larger challenge, as multiple members of the same smart home usually have different values and priorities. Consider a household where one person prefers the AC to be on, while the other person has asthma and prefers the AC to be off more often than on. These values clearly conflict with each other and can create a dynamic that is dependent on factors such as relationships, age, who set up the smart devices in the home, and other factors not explored in our survey [61].

The survey also focused on using satisfaction ratings as a proxy for how well smart home routine conflict resolution strategies met user expectations. Future work should look at other aspects of user perceptions, such as how well a strategy can map to a user's mental model. Previous research has done this in the context of intelligent agents and found that overly-simple as well as overly-complex solutions can impact perceptions of the agent overall [62]. Such a study would complement this work and provide a more complete understanding of how smart home designers can build more user-centric solutions.

Finally, our work also indicated the importance of user values in the formation of user desires and expectations as it relates to smart devices. We believe that future smart home routine management systems should allow for the expression of these values. At the same time, these future systems should also be designed to promote transparency and flexibility. Future work can seek to try different approaches of eliciting and storing user values and reflect them in the system's approach to handling routine conflicts.

The implementation of a system that allows for user value expression, followed by a user study of this system to evaluate its robustness and user experience is a significant next step from this work. For example, a system could be implemented to allow "tagging" [33], where users could tag routines with tags relating to the routine's importance to the user which can then influence how the system resolves a conflict between two routines. In addition, previous research has looked at using guaranteed home safety properties to avoid scenarios such as the one described in Sec. 1 where a user is locked in the house during a fire [3, 68]. Future work can use this body of research as well as this work to build upon and present smart homes that are more "context-aware."

In addition, existing systems like Google Nest [81] attempt to use machine learning to learn behaviors and values from people without explicit expression. Future work can examine the ethics of inferring values from users as well as the overall effectiveness of this approach.

7 CONCLUSION

Smart devices, and their associated routine management systems, are often designed with an emphasis on robustness or technical issues. We argue in this work that addressing the actual expectations and desires of the people using these systems is just as, if not more important. We designed a survey presenting hypothetical conflicts between routines scenarios in a smart home to participants as part of an Amazon Mechanical Turk survey to understand user expectations of how a routine management system should handle conflicts between routines. We found that there is no universally satisfactory system strategy. However, there are strategies that most participants seemed to prefer in our scenarios based on the context of the scenario. In addition, we observed the effects of user values on their expectations, seeing how users take general values they have and re-purpose them in the context of smart home routines, forming their expectations for how the system should handle routine conflicts. Our work highlights the importance that future system designers should place on user expectations and desires in their designs and corresponding models to better integrate such systems into users' lives.

ACKNOWLEDGMENTS

We would like to thank all our survey participants for contributing their time and valuable data. We would also like to thank the Social Systems Group, the Social Spaces Group, The Center for Just Infrastructures, and the Distributed Protocols Research Group (DPRG) for their feedback on the early drafts of our surveys as well as this paper. Finally, we would like to thank the Illinois Statistics Office for providing guidance on how to conduct and interpret complex quantitative analyses of our data. This project was supported in part by NSF CNS 1908888, a University of Illinois College of Engineering Strategic Initiatives Program grant, and a gift grant from Capital One.

REFERENCES

[1] James D Abbey and Margaret G Meloy. 2017. Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management* 53 (2017), 63–70.

[2] Critical Thinker Academy. [n. d.]. What are Moral Values? https://criticalthinkeracademy.com/courses/45150/lectures/659294. Last accessed January 2023.

[3] Shegufta Bakht Ahsan, Rui Yang, Shadi Abdollahian Noghabi, and Indranil Gupta. 2019. Home,{SafeHome}: Ensuring a Safe and Reliable Home Using the Edge. In *2nd USENIX Workshop on Hot Topics in Edge Computing (HotEdge 19)*.

[4] Shegufta B Ahsan, Rui Yang, Shadi A Noghabi, and Indranil Gupta. 2021. Home, safehome: smart home reliability with visibility and atomicity. In *Proceedings of the Sixteenth European Conference on Computer Systems*. 590–605.

[5] Air-Lux. [n. d.]. More than just a window. https://www.air-lux.com/smart-window/en/. Last accessed January 2023.

[6] Amazon Alexa 2014. Amazon Alexa. https://developer.amazon.com/alexa. Last accessed January 2023.

[7] Barry J Babin, William R Darden, and Mitch Griffin. 1994. Work and/or fun: measuring hedonic and utilitarian shopping value. *Journal of consumer research* 20, 4 (1994), 644–656.

[8] Nick Balboni. 2019. Rezzi: Bringing the Waste Collection Industry Into the Digital Future. https://www.rezzicompany.com/. Last accessed January 2023.

[9] Patricia Baudier, Chantal Ammi, and Matthieu Deboeuf-Rouchon. 2020. Smart home: Highly-educated students' acceptance. *Technological Forecasting and Social Change* 153 (2020), 119355.

[10] Sadie Boniface and Nicola Shelton. 2013. How is alcohol consumption affected if we account for under-reporting? A hypothetical scenario. *The European Journal of Public Health* 23, 6 (2013), 1076–1081.

[11] Åse Brandt, Kersti Samuelsson, Outi Töytäri, and Anna-Liisa Salminen. 2011. Activity and participation, quality of life and user satisfaction outcomes of environmental control systems and smart home technology: a systematic review. *Disability and Rehabilitation: Assistive Technology* 6, 3 (2011), 189–206.

[12] Berardina De Carolis, Stefano Ferilli, and Domenico Redavid. 2015. Incremental learning of daily routines as workflows in a smart home environment. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 4, 4 (2015), 1–23.

[13] Dipankar Chaki, Athman Bouguettaya, and Sajib Mistry. 2020. A conflict detection framework for iot services in multi-resident smart homes. In *2020 IEEE International Conference on Web Services (ICWS)*. IEEE, 224–231.

[14] Jae Bong Chang, Jayson L Lusk, and F Bailey Norwood. 2009. How closely do hypothetical surveys and laboratory experiments predict field behavior? *American Journal of Agricultural Economics* 91, 2 (2009), 518–534.

[15] Marketing Charts. 2020. Why do people buy smart home devices? https://www.marketingcharts.com/digital/non-mobile-connected-devices-114654.

[16] Michael Chmielewski and Sarah C Kucker. 2020. An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science* 11, 4 (2020), 464–473.

[17] Minjee Chung, Eunju Ko, Heerim Joung, and Sang Jin Kim. 2020. Chatbot e-service and customer satisfaction regarding luxury brands. *Journal of Business Research* 117 (2020), 587–595.

[18] IBM Corporation. [n. d.]. ACID properties of transactions. https://www.ibm.com/docs/en/cics-ts/5.4?topic=processing-acid-properties-transactions. Last accessed January 2023.

[19] Aykut Coskun, Gül Kaner, and İdil Bostan. 2018. Is smart home a necessity or a fantasy for the mainstream user? A study on users' expectations of smart household appliances. *International Journal of Design* 12, 1 (2018), 7–20.

[20] Rodney Cowled. 2022. Smart Hone Routine: Automate Your Day. https://rodneycowled.com/smart-home-routines/. Last accessed January 2023.

[21] Kari Dahlgren, Sarah Pink, Yolande Strengers, Larissa Nicholls, and Jathan Sadowski. 2021. Personalization and the Smart Home: questioning techno-hedonist imaginaries. *Convergence* 27, 5 (2021), 1155–1169.

[22] Scott Davidoff, Min Kyung Lee, Charles Yiu, John Zimmerman, and Anind K Dey. 2006. Principles of smart home control. In *International conference on ubiquitous computing*. Springer, 19–34.

[23] Guy Dewsbury, Bruce Taylor, and Martin Edge. 2001. The process of designing appropriate smart homes: including the user in the design. In *Equator IRC Workshop on Ubiquitous Computing in Domestic Environments*. Citeseer, 131–46.

[24] Colin Dixon, Ratul Mahajan, Sharad Agarwal, AJ Brush, Bongshin Lee, Stefan Saroiu, and Paramvir Bahl. 2012. An operating system for the home. In *9th USENIX Symposium on Networked Systems Design and Implementation (NSDI'12)*. USENIX Association, 337–352.

[25] CE Pro Editors. 2021. Smart Home Market expands as lower-income households embrace technology. https://www.cepro.com/news/smart-home-market-expands-lower-income-households/.

[26] Fenestra. [n. d.]. Make your windows smart. https://www.smartfenestra.com/home. Last accessed January 2023.

[27] Batya Friedman. 1996. Value-sensitive design. *interactions* 3, 6 (1996), 16–23.

[28] Silvery Fu and Sylvia Ratnasamy. 2021. dSpace: Composable Abstractions for Smart Spaces. In *Proceedings of the ACM SIGOPS 28th Symposium on Operating Systems Principles*. 295–310.

[29] Mathias Funk, Lin-Lin Chen, Shao-Wen Yang, and Yen-Kuang Chen. 2018. Addressing the need to capture scenarios, intentions and preferences: Interactive intentional programming in the smart home. *International Journal of Design* 12, 1 (2018), 53–66.
[30] Patricia I Fusch and Lawrence R Ness. 2015. Are we there yet? Data saturation in qualitative research. *The qualitative report* 20, 9 (2015), 1408.
[31] Mirta Galesic and Michael Bosnjak. 2009. Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public opinion quarterly* 73, 2 (2009), 349–360.
[32] 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*. 1–13.
[33] Gary Geisler and Sam Burns. 2007. Tagging video: Conventions and strategies of the YouTube community. In *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*. 480–480.
[34] Andrew Gelman, Jennifer Hill, and Masanao Yajima. 2012. Why we (usually) don't have to worry about multiple comparisons. *Journal of research on educational effectiveness* 5, 2 (2012), 189–211.
[35] Google Home 2016. Curious about smart home devices? Start here. https://store.google.com/us/product/google_home. Last accessed January 2023.
[36] Mary Gray. 2021. Dump the Dogfood: Towards a New Theory of Power in Computing. (2021). https://programs.sigchi.org/cscw/2021/program/content/66080 ACM Conference on Computer Supported Cooperative Work – Keynote Talk.
[37] Wei Gu, Peng Bao, Wenyuan Hao, and Jaewoong Kim. 2019. Empirical examination of intention to continue to use smart home services. *Sustainability* 11, 19 (2019), 5213.
[38] Tom Hargreaves and Charlie Wilson. 2017. Introduction: smart homes and their users. In *Smart homes and their users*. Springer, 1–14.
[39] Andrew Heinzman. 2021. Google Assistant Now Lets You Delay Actions Within Your Routines. https://www.reviewgeek.com/96057/google-assistant-now-lets-you-delay-actions-within-your-routines/. Last accessed January 2023.
[40] Sumi Helal, William Mann, Hicham El-Zabadani, Jeffrey King, Youssef Kaddoura, and Erwin Jansen. 2005. The gator tech smart house: A programmable pervasive space. *Computer* 38, 3 (2005), 50–60.
[41] Arto Hellas, Albina Zavgorodniaia, and Juha Sorva. 2020. Crowdsourcing in Computing Education Research: Case Amazon MTurk. In *Koli Calling'20: Proceedings of the 20th Koli Calling International Conference on Computing Education Research*. 1–5.
[42] Zac Henson. 2021. Turn on Fan (switch) when Samsung WiFi Cooktop is turned on? https://community.smartthings.com/t/turn-on-fan-switch-when-samsung-wifi-cooktop-is-turned-on/223445. Last accessed January 2023.
[43] Denis Hilton. 2017. Social attribution and explanation. (2017).
[44] Yale Home. [n. d.]. Welcome to Yale - Trusted every day. https://www.yalehome.com/. Last accessed January 2023.
[45] Home Assistant 2013. HomeAssistant: Awaken your home. https://www.home-assistant.io/. Last accessed January 2023.
[46] HomeKit 2014. HomeKit: Developing Apps and Accessories for the Home. https://developer.apple.com/homekit/. Last accessed January 2023.
[47] Matthew B Hoy. 2018. Alexa, Siri, Cortana, and more: an introduction to voice assistants. *Medical reference services quarterly* 37, 1 (2018), 81–88.
[48] Philips Hue. [n. d.]. Philips Hue: Smart lighting. www.philips-hue.com/. Last accessed January 2023.
[49] Ifttt. [n. d.]. https://ifttt.com/. Last accessed January 2023.
[50] Imolaza. [n. d.]. The Best Smart Sprinkler Controller. https://imolazaonline.com/. Last accessed January 2023.
[51] Amazon Mechanical Turk Inc. [n. d.]. Amazon Mechanical Turk: Access a global, on-demand, 24x7 workforce. https://www.mturk.com/.
[52] Samsung Electronics America Inc. [n. d.]. Connect and use SmartThings with your Samsungcooktop. https://www.samsung.com/us/support/answer/ANS00060835/. Last accessed January 2023.
[53] Wyze Labs Inc. [n. d.]. Wyze Cam v3 Garage Door Controller. https://www.wyze.com/products/wyze-garage-door-controller?related_selling_plan=41618559008930. Last accessed January 2023.
[54] IoTBench. [n. d.]. IoTBench/iotbench-test-suite: A micro-benchmark suite to assess the effectiveness of tools designed for IOT Apps. https://github.com/IoTBench/IoTBench-test-suite.
[55] Abir B Karami, Anthony Fleury, Jacques Boonaert, and Stéphane Lecoeuche. 2016. User in the loop: Adaptive smart homes exploiting user feedback—state of the art and future directions. *Information* 7, 2 (2016), 35.
[56] Diane Kelly. 2009. *Methods for evaluating interactive information retrieval systems with users*. Now Publishers Inc.
[57] Shahedul Huq Khandkar. 2009. Open coding. *University of Calgary* 23 (2009), 2009.
[58] Shashank Khanna, Aishwarya Ratan, James Davis, and William Thies. 2010. Evaluating and improving the usability of Mechanical Turk for low-income workers in India. In *Proceedings of the first ACM symposium on computing for development*. 1–10.
[59] Cory D Kidd, Robert Orr, Gregory D Abowd, Christopher G Atkeson, Irfan A Essa, Blair MacIntyre, Elizabeth Mynatt, Thad E Starner, and Wendy Newstetter. 1999. The aware home: A living laboratory for ubiquitous computing research. In *International workshop on cooperative buildings*. Springer, 191–198.
[60] Julia Kiseleva, Kyle Williams, Jiepu Jiang, Ahmed Hassan Awadallah, Aidan C Crook, Imed Zitouni, and Tasos Anastasakos. 2016. Understanding user satisfaction with intelligent assistants. In *Proceedings of the 2016 ACM on Conference on Human Information*

Interaction and Retrieval. 121–130.
[61] Vinay Koshy, Joon Sung Sung Park, Ti-Chung Cheng, and Karrie Karahalios. 2021. "We Just Use What They Give Us": Understanding Passenger User Perspectives in Smart Homes. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.
[62] Todd Kulesza, Simone Stumpf, Margaret Burnett, Sherry Yang, Irwin Kwan, and Weng-Keen Wong. 2013. Too much, too little, or just right? Ways explanations impact end users' mental models. In *2013 IEEE Symposium on visual languages and human centric computing*. IEEE, 3–10.
[63] Fabienne Lang. 2019. New Automated Trash Can Drives Itself to the Curb on Trash Day. https://interestingengineering.com/smartcan-new-automated-trash-can-drive-itself-to-the-curb-on-trash-day.
[64] Åsa Larsson Ranada and Helene Lidström. 2019. Satisfaction with assistive technology device in relation to the service delivery process—A systematic review. *Assistive Technology* 31, 2 (2019), 82–97.
[65] Thibaut Le Guilly, Michael K Nielsen, Thomas Pedersen, Arne Skou, Jesper Kjeldskov, and Mikael Skov. 2016. User constraints for reliable user-defined smart home scenarios. *Journal of Reliable Intelligent Environments* 2, 2 (2016), 75–91.
[66] Xinyi Li, Lei Zhang, and Xipeng Shen. 2020. DIAC: An Inter-app Conflicts Detector for Open IoT Systems. *ACM Transactions on Embedded Computing Systems (TECS)* 19, 6 (2020), 1–25.
[67] Xinyi Li, Lei Zhang, Xipeng Shen, and Yong Qi. 2017. *A Systematic Examination of Inter-App Conflicts Detections in Open IoT Systems*. Technical Report. Technical Report TR-2017-1. North Carolina State University, Dept. of Computer Science.
[68] Chieh-Jan Mike Liang, Börje F Karlsson, Nicholas D Lane, Feng Zhao, Junbei Zhang, Zheyi Pan, Zhao Li, and Yong Yu. 2015. SIFT: Building an Internet of Safe Things. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks* (Seattle, Washington) *(IPSN '15)*. ACM, 298–309. https://doi.org/10.1145/2737095.2737115
[69] Huichen Lin and Neil W Bergmann. 2016. IoT privacy and security challenges for smart home environments. *Information* 7, 3 (2016), 44.
[70] Mingnan Liu and Laura Wronski. 2018. Examining completion rates in web surveys via over 25,000 real-world surveys. *Social Science Computer Review* 36, 1 (2018), 116–124.
[71] Craig Lloyd. 2018. You Can Now Set a Delay in Alexa Routines. https://www.howtogeek.com/fyi/you-can-now-set-a-delay-in-alexa-routines/. Last accessed January 2023.
[72] Lutron. [n. d.]. Lutron: Beautiful light. Intelligent Shades. Powerful Controls. www.lutron.com. Last accessed January 2023.
[73] Bernd Marcus, Michael Bosnjak, Steffen Lindner, Stanislav Pilischenko, and Astrid Schütz. 2007. Compensating for low topic interest and long surveys: a field experiment on nonresponse in web surveys. *Social Science Computer Review* 25, 3 (2007), 372–383.
[74] Davit Marikyan, Savvas Papagiannidis, Rajiv Ranjan, and Omer Rana. 2021. Working in a Smart Home-office: Exploring the Impacts on Productivity and Wellbeing. In *17th International Conference on Web Information Systems and Technologies (WEBIST 2021)*. SciTePress, 275–282.
[75] SmartThings Community Member. 2016. Crock Pot / Slow Cooker - Smart App / Device Handler replacement? https://community.smartthings.com/t/crock-pot-slow-cooker-smart-app-device-handler-replacement/67898. Last accessed January 2023.
[76] Sarah Mennicken, David Kim, and Elaine May Huang. 2016. Integrating the smart home into the digital calendar. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 5958–5969.
[77] Sarah Mennicken, Jo Vermeulen, and Elaine M Huang. 2014. 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*. 105–115.
[78] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence* 267 (2019), 1–38.
[79] Sirajum Munir and John A Stankovic. 2014. Depsys: Dependency aware integration of cyber-physical systems for smart homes. In *2014 ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS '14)*. IEEE, 127–138.
[80] myQ. [n. d.]. Look UP – myQ turns any garage into a smart garage. https://www.myq.com/. Last accessed January 2023.
[81] Google Nest. [n. d.]. Google Nest Smart Home. Build your helpful home. https://store.google.com/category/connected_home?hl=en-US. Last accessed January 2023.
[82] Google Nest. [n. d.]. Nest Protect smoke and CO alarm. https://store.google.com/product/nest_protect_2nd_gen?hl=en-US&pli=1
[83] Google Nest. [n. d.]. Welcome to the Google Nest Community! https://www.googlenestcommunity.com/. Last accessed January 2023.
[84] Geoff Norman. 2010. Likert scales, levels of measurement and the "laws" of statistics. *Advances in health sciences education* 15, 5 (2010), 625–632.
[85] Dan P. Parker. 2019. Samsung Gas Cooktop Device. https://community.smartthings.com/t/samsung-gas-cooktop-device/166651. Last accessed January 2023.
[86] Kim Parker, Juliana Menasce Horowitz, and Rachel Minkin. 2021. How coronavirus has changed the way Americans work. https://www.pewresearch.org/social-trends/2020/12/09/how-the-coronavirus-outbreak-has-and-hasnt-changed-the-way-americans-work/.
[87] Ed.D. Posted byKristen Mattson. 2016. App-dependent or app-enabled? https://drkmattson.wordpress.com/2015/02/02/app-dependent-or-app-enabled/.

[88] Rachio. [n. d.]. Connect Outside – Elevate your yard from the ground up with Rachio's smart watering products. https://rachio.com/. Last accessed January 2023.

[89] Elissa M Redmiles, Sean Kross, and Michelle L Mazurek. 2019. How well do my results generalize? Comparing security and privacy survey results from mturk, web, and telephone samples. In *2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, 1326–1343.

[90] Statista Digital Market Outlook Market Report. 2022. Smart home - United States: Statista market forecast. https://www.statista.com/outlook/dmo/smart-home/united-states. Last accessed January 2023.

[91] Grand View Research. [n. d.]. Smart Home Platforms Market Size, Share & Trends Analysis Report By Type (IoT Platform for Smart Appliances, IoT Platform for Security & Surveillance Systems), By Region, And Segment Forecasts, 2022 - 2028. https://www.grandviewresearch.com/industry-analysis/smart-home-platforms-market-report. Last accessed January 2023.

[92] D. Retkowitz and S. Kulle. 2009. Dependency management in smart homes. In *IFIP International Conference on Distributed Applications and Interoperable Systems (DAIS'09)*. Springer, 143–156.

[93] Melanie Revilla and Carlos Ochoa. 2017. Ideal and maximum length for a web survey. *International Journal of Market Research* 59, 5 (2017), 557–565.

[94] Samsung SmartThings 2012. Samsung SmartThings. https://www.smartthings.com/. Last accessed January 2023.

[95] Anique Scheerder, Alexander Van Deursen, and Jan Van Dijk. 2017. Determinants of Internet skills, uses and outcomes. A systematic review of the second-and third-level digital divide. *Telematics and informatics* 34, 8 (2017), 1607–1624.

[96] Birte Schiffhauer, Jasmin Bernotat, Friederike Eyssel, Rebecca Bröhl, and Jule Adriaans. 2016. Let the User Decide! User Preferences Regarding Functions, Apps, and Interfaces of a Smart Home and a Service Robot. In *International Conference on Social Robotics*. Springer, 971–981.

[97] Aritra Sengupta, Tanakorn Leesatapornwongsa, Masoud Saeida Ardekani, and Cesar A Stuardo. 2019. Transactuations: where transactions meet the physical world. In *2019 USENIX Annual Technical Conference (ATC'19)*. USENIX Association, 91–106.

[98] Chenguang Shen, Rayman Preet Singh, Amar Phanishayee, Aman Kansal, and Ratul Mahajan. 2016. Beam: Ending monolithic applications for connected devices. In *2016 USENIX Annual Technical Conference (ATC'16)*. USENIX Association, 143–157.

[99] Jungwoo Shin, Yuri Park, and Daeho Lee. 2018. Who will be smart home users? An analysis of adoption and diffusion of smart homes. *Technological Forecasting and Social Change* 134 (2018), 246–253.

[100] SimpliSafe. [n. d.]. Find the right protection for your home. https://simplisafe.com/security. Last accessed January 2023.

[101] Smart Life 2019. SmartLife: Build smart home with smart life. https://www.ismartlife.me/. Last accessed January 2023.

[102] Smiljanic Stasha. 2022. An In-Depth View Into Smart Home Statistics. https://policyadvice.net/insurance/insights/smart-home-statistics/ Last accessed January 2023.

[103] Lawrence E Susskind, Sarah McKearnen, and Jennifer Thomas-Lamar. 1999. *The consensus building handbook: A comprehensive guide to reaching agreement*. Sage publications.

[104] Daphne Townsend, Frank Knoefel, and Rafik Goubran. 2011. Privacy versus autonomy: a tradeoff model for smart home monitoring technologies. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 4749–4752.

[105] Rahmadi Trimananda, Seyed Amir Hossein Aqajari, Jason Chuang, Brian Demsky, and Guoqing Harry Xu. 2020. IoTCheck and manual study supporting materials.

[106] Rahmadi Trimananda, Seyed Amir Hossein Aqajari, Jason Chuang, Brian Demsky, Guoqing Harry Xu, and Shan Lu. 2020. Understanding and automatically detecting conflicting interactions between smart home iot applications. In *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. 1215–1227.

[107] Blase Ur, Elyse McManus, Melwyn Pak Yong Ho, and Michael L Littman. 2014. Practical trigger-action programming in the smart home. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 803–812.

[108] Alexander JAM van Deursen, Alex van der Zeeuw, Pia de Boer, Giedo Jansen, and Thomas van Rompay. 2021. Digital inequalities in the Internet of Things: differences in attitudes, material access, skills, and usage. *Information, Communication & Society* 24, 2 (2021), 258–276.

[109] Pascal A Vicaire, Enamul Hoque, Zhiheng Xie, and John A Stankovic. 2011. Bundle: A group-based programming abstraction for cyber-physical systems. *IEEE Transactions on Industrial Informatics* 8, 2 (2011), 379–392.

[110] Pascal A Vicaire, Zhiheng Xie, Enamul Hoque, and John A Stankovic. 2010. Physicalnet: A generic framework for managing and programming across pervasive computing networks. In *2010 16th IEEE Real-Time and Embedded Technology and Applications Symposium*. IEEE, 269–278.

[111] Donald E Vinson, Jerome E Scott, and Lawrence M Lamont. 1977. The role of personal values in marketing and consumer behavior. *Journal of marketing* 41, 2 (1977), 44–50.

[112] What is Imprint 2021. What is Imprint™ Link Technology? https://homesupport.irobot.com/app/answers/detail/a_id/21088/\sim/what-is-imprint%E2%84%A2-link-technology%3F. Last accessed January 2023.

[113] Charlie Wilson, Tom Hargreaves, and Richard Hauxwell-Baldwin. 2015. Smart homes and their users: a systematic analysis and key challenges. *Personal and Ubiquitous Computing* 19, 2 (2015), 463–476.

[114] Christopher Winship and Robert D Mare. 1984. Regression models with ordinal variables. *American sociological review* (1984), 512–525.

[115] Jong-bum Woo and Youn-kyung Lim. 2015. User experience in do-it-yourself-style smart homes. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 779–790.
[116] Miki Yagita, Fuyuki Ishikawa, and Shinichi Honiden. 2015. An application conflict detection and resolution system for smart homes. In *2015 IEEE/ACM 1st international workshop on software engineering for smart cyber-physical systems*. IEEE, 33–39.
[117] Yonomi 2013. Yonomi: A common framework for connecting smart devices to an app. https://www.yonomi.com/. Last accessed January 2023.
[118] Justin C Zhang, Katherine L Christensen, Richard K Leuchter, Sitaram Vangala, Maria Han, and Daniel M Croymans. 2022. Examining the role of COVID-19 testing availability on intention to isolate: A Randomized hypothetical scenario. *PLOS one* 17, 2 (2022), e0262659.
[119] Serena Zheng, Noah Apthorpe, Marshini Chetty, and Nick Feamster. 2018. User perceptions of smart home IoT privacy. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–20.

A STRATEGY DETAILS

Table 7. List of All Strategy Candidates

R1	**R2**	**Type**	**Description**	**# Combination**
Any	M-6 / M-7 / M-8 / M-9	Duplicated	R2 has not started yet when conflict happens. Thus, M-i = M-(i-4) for $i \geq 6$ for R2	36
M-1	M-1	Possible, but not selected.	R1 stops executing, and R2 does not start	1
M-1	M-2 / M-3 / M-5	Selected	**Truncate**	3
M-1	M-4	Possible, but not selected.	R1 stops executing, and R2 only execute non-conflicting command	1
M-2 / M-3 / M-5	M-1	Selected	**Cancel**	3
M-2	M-2	Selected	**Execute**	1
M-2	M-3	Selected	**Delay**	1
M-2 / M-5	M-4	Selected	**Ignore**	2
M-2	M-5	Selected	**Stretch**	1
M-3	M-2	Selected	**Pause**	1
M-3 / M-5	M-3 / M-5	Not possible	Deadlock. Both routines try to wait for the other one to finish.	4
M-3	M-4	Possible, but not selected.	R1 and R2 both execute non-conflicting commands only. After R2 finished, R1 continues the conflicting command.	1
M-4	M-1	Possible, but not selected.	R1 stops executing the conflicting command, but R2 does not start.	1
M-4	M-2 / M-5	Selected	**Overwrite**	2
M-4	M-3	Possible, but not selected.	R1 stops executing the conflicting command. R2 starts executing non-conflicting commands first, and will start conflicting command after R1 is done.	1
M-4	M-4	Possible, but not selected.	R1 and R2 both execute non-conflicting commands only.	1

Continued on next page

Table 7 – continued from previous page

R1	**R2**	**Type**	**Description**	**# Combination**
M-5	M-2	Selected	**Resume**	1
M-6 / M-7 / M-9	M-1	Possible, but not selected.	Reset all devices to the states before R1 happens. Then R1 is redone, and R2 never start.	3
M-6	M-2	Possible, but not selected.	Reset all devices to the states before R1 happens. Then R1 and R2 start to execute together. Conflict still exists.	1
M-6	M-3	Possible, but not selected.	Reset all devices to the states before R1 happens. Then execute the whole R1, then whole R2.	1
M-6 / M-9	M-4	Possible, but not selected.	Reset all devices to the states before R1 happens.Then execute the both R1 and R2 except for the conflicting command in R2.	2
M-6	M-5	Possible, but not selected.	Reset all devices to the states before R1 happens.Then execute the both R1 and R2 except for the conflicting command in R2. After R1 is done, execute the conflicting command in R2.	1
M-7	M-2	Selected	**Revert**	1
M-7 / M-9	M-3 / M-5	Not possible	Deadlock. Both routines try to wait for the other one to finish.	4
M-7	M-4	Possible, but not selected.	Reset all devices to the states before R1 happens. Then execute the non-conflicting commands in R2, then whole R1.	1
M-8	M-1	Possible, but not selected.	Reset all devices to the states before R1 happens. Then all non-conflicting commands in R1 is redone, and R2 never start.	1
M-8	M-2 / M-5	Possible, but not selected.	Reset all devices to the states before R1 happens.Then execute the both R1 and R2 except for the conflicting command in R1.	2
M-8	M-3	Possible, but not selected.	Reset all devices to the states before R1 happens.Then execute the non-conflicting commands in both R1 and R2. After R1 is done, execute the conflicting command in R2.	1
M-8	M-4	Possible, but not selected.	Reset all devices to the states before R1 happens.Then execute the non-conflicting commands in both R1 and R2.	1
M-9	M-2	Possible, but not selected.	Reset all devices to the states before R1 happens.Then execute the both R1 and R2 except for the conflicting command in R1. After R2 is done, execute the conflicting command in R1.	1
				Total: 81

B SCENARIO DETAILS

Table 8. Scenario Content

ID	**Short Name**	**Description**	**Trigger**
1	Doorbell during meeting	Today at 10 am you have a meeting, so you start your "meeting" routine. This routine turns on the lights in your room and mutes all the speakers for 1 hour, until 11 pm. However, at 10:20 am, the delivery driver presses your doorbell and triggers your "visitor comes" routine. This routine starts the doorbell camera and plays an alarm followed by the sound of what's going on outside on your speaker for 3 minutes until 10:23 am.	Human v.s. Human (other)

Continued on next page

Table 8 – continued from previous page

ID	Short Name	Description	Trigger
2	Late night back	Every night after 10 pm, your "night safety" routine runs. This routine locks the door until the next morning at 6 am. The door is unlocked at 6am. However, at 2 am, you're back from a night out with your friends, and you start the "back home" routine. This routine unlocks the door and turns on the foyer light.	Time v.s. Human
3	Baking when cooling	At 11:00am, you feel hot and start the "Cool Home" routine, which turns on the AC and closes all the windows in the house for 1 hour. The AC is supposed to turn off at 12:00pm. However, 30 minutes later at 11:30am, you want to start the "Baking" routine, which opens the windows in the house and turns on the oven to 375 degrees Fahrenheit for one hour. The oven is supposed to be off at the end of the "Baking" routine (12:30pm).	Human v.s. Human
4	Reading around bedtime	Today at 11 pm, you start your "bedside reading" routine. This routine turns off the roof light and turns on the bedside light for 1 hour, until 12 am. However, at 11:30 pm, your nightly "bedtime" routine starts. This routine turns off all the lights except for the night light in your room and starts playing deep-sleep-helping music on the speaker for 20 minutes, until 11:50 pm.	Human v.s. Time
5	Family visit during vacation	You are on a trip for 3 days, so you start your "vacation" routine. This routine randomly turns your lights on and off for 3 days. However, on the 2nd day, you need some information on a document in your home, so you ask your family to get the information for you. When your family gets to your place, they trigger the "welcome home" routine, which turns on the hallway lights for 5 minutes and plays a "welcome" sound on the speakers. Your family gets to your place at 3 pm and leaves soon after.	Human v.s. Human
6	Trash out at the end of leaving	On Tuesday at 7:52am, you drive your car out of the garage to get to work which triggers your "automated garage door" routine. This routine closes the garage door 10 minutes after you leave, at 8:02 am. However, every Tuesday morning at 8 am, your "trash out" routine starts. This routine opens the garage door, moves out the trash can from inside the garage to the curb, and closes the garage door after 5 minutes at 8:05 am.	Sensor v.s. Time
7	Work before bedtime	At 11 pm, you get a work call so you start your "work" routine. This routine turns the desk light on for 1 hour until 12 am, and then turns it off. However, at 11:30 pm, your "bedtime" routine starts. This routine turns off all the lights except for the night light in your room and then starts playing deep-sleep-helping music through your speaker for 20 minutes, until 11:50 pm.	Human v.s. Time
8	Get up early2	At 7:25am, you wake up and start your "Good Morning" routine. This routine starts the coffee maker and plays the news on your speaker for 15 mins, until 7:40am. However, 5 minutes later at 7:30am, your daily "Wake Up" routine starts, which plays soft music at a gradually increasing volume on the speaker for 30 minutes.	Time v.s. Human
9	News during meeting	At 9:50 am you have a meeting, so you start the "meeting" routine which lasts for 1 hour, which turns on the lights in your room and mutes the speakers until 10:50 am. However, every morning at 10 am your smart speaker plays the news for 5 minutes, until 10:05 am.	Human v.s. Time
10	Baking when cooling longer	At 11:00am, you feel hot and start your "Cool Home" routine, which turns on the AC and closes all the windows for 2 hours. The routine is supposed to turn off the AC and open the windows at 1:00pm. However, 30 minutes later at 11:30am, you want to start your "Baking" routine. This routine turns on the oven to 375 degrees Fahrenheit and opens the windows for one hour. The routine is supposed to turn the oven off and open the windows are supposed to be open at 12:30pm.	Human v.s. Human
11	Rain when baking	At 11:30 am, you start your "Baking" routine. This routine turns on the oven to 375 degrees Fahrenheit and opens the windows for 1 hour. The routine is supposed to turn off the oven and keep the windows open at 12:30 pm. Oven is supposed to turn off at 12:30 pm, but the windows will stay open. However, 30 minutes later at 12:00pm, it starts to rain, which triggers your "Raining Time" routine. This routine closes the windows and the balcony door.	Huaman v.s. Time
12	Meeting during nap	At 1:45 pm, you feel sleepy and decide to take a nap. You start your "naptime" routine, which dims the lights and shuts the blinds for 30 minutes, until 2:15 pm. However at 2 pm, you have a meeting scheduled, which will trigger your "meeting" routine, which turns on the lights to full brightness level in your room, and turns on your computer.	Human v.s. Time-triggered
13	Early back from vacation	You are on a trip for 3 days. When you left home at 5 pm, you started your "vacation" routine. This routine randomly turns your lights on and off for 3 days. However, on the 3rd day at 3 pm, you get back home early and trigger your "welcome home" routine. This routine turns on the hallway lights for 5 minutes and plays a "welcome home" sound on the speaker.	Human v.s. Human

Continued on next page

Table 8 – continued from previous page

ID	Short Name	Description	Trigger
14	Visit when watering	At 7 pm, you start the "Water Lawn" routine, which turns ON the sprinklers in the front lawn for 1 hour, until 8 pm. However, at 7:30 PM: someone walks up the walkway to the front door. This triggers the motion sensor and starts the "in garden" routine. This routine is supposed to turn ON the garden lights and turn OFF the sprinklers.	Human v.s. Sensor
15	Leaving when watering	Every Wednesday at 7 PM, your "Water Lawn" routine begins. This routine turns on the sprinklers in the front lawn for 1 hour, till 8 PM. However, at 7:30 PM, you go out for a dinner with friend and start your "leaving home" routine. This routine turns off the water sprinklers, and after 5 minutes turns off all the lights in the house and locks the door for 30 minutes. If people pass the garden when the water sprinkler is on, people may get wet.	Human v.s. Human
16	Doorbell during nap	Your "Nap time" routine runs every day at 3 pm. This routine dims all the lights in your room for 30 minutes until 3:30 pm. However, someone presses the doorbell today at 3:15 pm, which starts your "doorbell" routine. This routine flashes the lights in your room for 1 minute until 3:16 pm and then sets the lights to normal brightness level (full brightness).	Human v.s. Human (other)
17	Meeting during news time	Every morning at 10 am you get the news from your smart speaker, which takes 20 minutes, until 10:20 am. However, at 10:10 am you get a call from your co-worker and need to start a meeting, so you start your "meeting" routine. This routine turns on the lights in your room and mutes the speakers for 1 hour until 11:10 am.	Time v.s. Human
18	Get up early1	At 7:50am, you wake up and play music on your speaker for 30 mins to listen to in bed. However, at 8am, your daily "Good Morning" routine runs. This routine starts the coffee maker and plays the news on your speaker for 20 mins.	Human v.s. Time
19	Trash out right after Leaving	On Tuesday at 7:58 am, you drive your car out of the garage to get to work which triggers your "automated garage door" routine. This routine closes the garage door 10 minutes after you leave, at 8:08 am. However, every Tuesday morning at 8 am, your "trash out" routine starts. This routine opens the garage door, moves out the trash can from inside the garage and closes the garage door after 5 minutes, at 8:05 am.	Sensor v.s. Time
20	Hot when baking	At 11:30am, you start your "Baking" routine. This routine turns on the oven to 375 degrees Fahrenheit and opens the window for 1 hour until 12:30pm, when it turns the oven off and keeps the window open. However, at 12:00pm, you feel hot and want to start your "Cool Home" routine. This routine turns on the AC and closes all the windows in the house.	Human v.s. Human

C CREDIBLE INTERVALS OF REGRESSION MODEL

Table 9. 95% Credible Intervals of model slopes in the Regression model of the Interaction between Scenario and System Strategy Effects on user satisfaction. The range between the lower and upper bounds captures 95% of the predicted changes to user satisfaction when a specific system strategy is applied to a specific scenario. Values greater than 0 correspond generally to improvements in Likert ratings, and values less than 0 correspond generally to decreases in Likert ratings.

ID	Scenario Name	Strategy	Mean	Lower Bound (2.5%)	Upper Bound (97.5%)
1	Doorbell during meeting	Truncate	3.552	0.363	0.334
1	Doorbell during meeting	Delay	3.561	0.308	0.352
1	Doorbell during meeting	Cancel	3.802	0.367	0.357
1	Doorbell during meeting	Execute	3.524	0.36	0.332
1	Doorbell during meeting	Pause	3.53	0.372	0.333
1	Doorbell during meeting	Overwrite	3.554	0.338	0.357
1	Doorbell during meeting	Resume	3.513	0.352	0.397
1	Doorbell during meeting	Ignore	3.573	0.331	0.362
1	Doorbell during meeting	Stretch	3.614	0.491	0.542
1	Doorbell during meeting	Revert	3.563	0.315	0.339
2	Late night back	Truncate	3.42	0.263	0.234

Continued on next page

Table 9 – continued from previous page

ID	Scenario Name	Strategy	Mean	Lower Bound (2.5%)	Upper Bound (97.5%)
2	Late night back	Delay	3.583	0.257	0.251
2	Late night back	Cancel	3.478	0.525	0.431
2	Late night back	Execute	3.409	0.248	0.266
2	Late night back	Pause	3.855	0.223	0.213
2	Late night back	Overwrite	3.42	0.253	0.243
2	Late night back	Resume	3.849	0.245	0.216
2	Late night back	Ignore	3.465	0.275	0.235
2	Late night back	Stretch	3.623	0.536	0.461
2	Late night back	Revert	3.835	0.228	0.225
3	Baking when cooling	Truncate	3.589	0.386	0.356
3	Baking when cooling	Delay	3.651	0.352	0.349
3	Baking when cooling	Cancel	3.491	0.37	0.371
3	Baking when cooling	Execute	3.634	0.376	0.361
3	Baking when cooling	Pause	3.692	0.365	0.348
3	Baking when cooling	Overwrite	3.648	0.361	0.359
3	Baking when cooling	Resume	3.663	0.495	0.406
3	Baking when cooling	Ignore	3.528	0.358	0.374
3	Baking when cooling	Stretch	3.632	0.514	0.499
3	Baking when cooling	Revert	3.529	0.38	0.347
4	Reading around bedtime	Truncate	3.794	0.334	0.282
4	Reading around bedtime	Delay	3.601	0.328	0.304
4	Reading around bedtime	Cancel	3.282	0.358	0.338
4	Reading around bedtime	Execute	3.779	0.286	0.294
4	Reading around bedtime	Pause	3.778	0.323	0.32
4	Reading around bedtime	Overwrite	3.797	0.294	0.293
4	Reading around bedtime	Resume	3.43	0.358	0.315
4	Reading around bedtime	Ignore	3.518	0.344	0.314
4	Reading around bedtime	Stretch	3.632	0.559	0.464
4	Reading around bedtime	Revert	3.463	0.324	0.349
5	Family visit during vacation	Truncate	3.699	0.288	0.273
5	Family visit during vacation	Delay	3.583	0.289	0.32
5	Family visit during vacation	Cancel	3.514	0.293	0.32
5	Family visit during vacation	Execute	3.68	0.286	0.286
5	Family visit during vacation	Pause	3.684	0.276	0.281
5	Family visit during vacation	Overwrite	3.694	0.293	0.269
5	Family visit during vacation	Resume	3.668	0.284	0.277
5	Family visit during vacation	Ignore	3.5	0.294	0.324

Continued on next page

Table 9 – continued from previous page

ID	Scenario Name	Strategy	Mean	Lower Bound (2.5%)	Upper Bound (97.5%)
5	Family visit during vacation	Stretch	3.638	0.523	0.444
5	Family visit during vacation	Revert	3.647	0.299	0.277
6	Trash out at the end of leaving	Truncate	3.881	0.346	0.289
6	Trash out at the end of leaving	Delay	4.023	0.292	0.313
6	Trash out at the end of leaving	Cancel	3.268	0.4	0.404
6	Trash out at the end of leaving	Execute	3.24	0.393	0.406
6	Trash out at the end of leaving	Pause	3.863	0.323	0.318
6	Trash out at the end of leaving	Overwrite	3.886	0.317	0.326
6	Trash out at the end of leaving	Resume	3.671	0.452	0.479
6	Trash out at the end of leaving	Ignore	3.17	0.419	0.379
6	Trash out at the end of leaving	Stretch	3.641	0.487	0.506
6	Trash out at the end of leaving	Revert	3.576	0.412	0.328
7	Work before bedtime	Truncate	3.859	0.324	0.285
7	Work before bedtime	Delay	3.589	0.321	0.349
7	Work before bedtime	Cancel	3.497	0.357	0.337
7	Work before bedtime	Execute	3.841	0.301	0.297
7	Work before bedtime	Pause	3.586	0.336	0.306
7	Work before bedtime	Overwrite	3.862	0.309	0.295
7	Work before bedtime	Resume	3.465	0.389	0.362
7	Work before bedtime	Ignore	3.532	0.355	0.326
7	Work before bedtime	Stretch	3.648	0.508	0.483
7	Work before bedtime	Revert	3.603	0.327	0.319
8	Get up early2	Truncate	3.68	0.258	0.284
8	Get up early2	Delay	3.664	0.271	0.288
8	Get up early2	Cancel	3.543	0.249	0.227
8	Get up early2	Execute	3.673	0.269	0.266
8	Get up early2	Pause	3.771	0.269	0.271
8	Get up early2	Overwrite	3.688	0.257	0.25
8	Get up early2	Resume	3.681	0.487	0.441
8	Get up early2	Ignore	3.603	0.295	0.267
8	Get up early2	Stretch	3.646	0.482	0.526
8	Get up early2	Revert	3.568	0.275	0.286
9	News during meeting	Truncate	3.6	0.358	0.363
9	News during meeting	Delay	4.052	0.331	0.293
9	News during meeting	Cancel	3.623	0.324	0.347
9	News during meeting	Execute	3.586	0.358	0.32
9	News during meeting	Pause	3.568	0.291	0.323

Continued on next page

Table 9 – continued from previous page

ID	Scenario Name	Strategy	Mean	Lower Bound (2.5%)	Upper Bound (97.5%)
9	News during meeting	Overwrite	3.606	0.334	0.345
9	News during meeting	Resume	3.615	0.38	0.312
9	News during meeting	Ignore	3.663	0.319	0.341
9	News during meeting	Stretch	3.656	0.497	0.452
9	News during meeting	Revert	3.57	0.391	0.331
10	Baking when cooling longer	Truncate	3.766	0.355	0.293
10	Baking when cooling longer	Delay	3.707	0.334	0.299
10	Baking when cooling longer	Cancel	3.22	0.373	0.374
10	Baking when cooling longer	Execute	3.737	0.369	0.325
10	Baking when cooling longer	Pause	3.713	0.314	0.348
10	Baking when cooling longer	Overwrite	3.758	0.342	0.33
10	Baking when cooling longer	Resume	3.75	0.307	0.355
10	Baking when cooling longer	Ignore	3.546	0.357	0.343
10	Baking when cooling longer	Stretch	3.649	0.535	0.489
10	Baking when cooling longer	Revert	3.717	0.34	0.338
11	Rain when baking	Truncate	3.576	0.362	0.338
11	Rain when baking	Delay	3.728	0.348	0.322
11	Rain when baking	Cancel	3.513	0.36	0.367
11	Rain when baking	Execute	3.639	0.317	0.358
11	Rain when baking	Pause	3.636	0.379	0.339
11	Rain when baking	Overwrite	3.662	0.353	0.315
11	Rain when baking	Resume	3.876	0.284	0.271
11	Rain when baking	Ignore	3.842	0.286	0.3
11	Rain when baking	Stretch	3.655	0.489	0.478
11	Rain when baking	Revert	3.577	0.355	0.353
12	Meeting during nap	Truncate	3.721	0.301	0.28
12	Meeting during nap	Delay	3.737	0.32	0.309
12	Meeting during nap	Cancel	3.628	0.312	0.343
12	Meeting during nap	Execute	3.701	0.304	0.29
12	Meeting during nap	Pause	3.704	0.297	0.302
12	Meeting during nap	Overwrite	3.723	0.294	0.301
12	Meeting during nap	Resume	3.691	0.488	0.431
12	Meeting during nap	Ignore	3.623	0.301	0.328
12	Meeting during nap	Stretch	3.658	0.491	0.508
12	Meeting during nap	Revert	3.551	0.294	0.312
13	Early back from vacation	Truncate	4.112	0.268	0.241
13	Early back from vacation	Delay	3.548	0.301	0.272

Continued on next page

Table 9 – continued from previous page

ID	Scenario Name	Strategy	Mean	Lower Bound (2.5%)	Upper Bound (97.5%)
13	Early back from vacation	Cancel	3.556	0.293	0.295
13	Early back from vacation	Execute	3.724	0.272	0.273
13	Early back from vacation	Pause	3.724	0.305	0.266
13	Early back from vacation	Overwrite	3.736	0.263	0.254
13	Early back from vacation	Resume	3.714	0.301	0.263
13	Early back from vacation	Ignore	3.496	0.299	0.287
13	Early back from vacation	Stretch	3.66	0.49	0.527
13	Early back from vacation	Revert	3.438	0.325	0.289
14	Visit when watering	Truncate	3.82	0.333	0.32
14	Visit when watering	Delay	3.604	0.349	0.322
14	Visit when watering	Cancel	3.487	0.376	0.344
14	Visit when watering	Execute	3.844	0.323	0.32
14	Visit when watering	Pause	3.648	0.36	0.371
14	Visit when watering	Overwrite	3.867	0.32	0.333
14	Visit when watering	Resume	3.633	0.369	0.342
14	Visit when watering	Ignore	3.698	0.333	0.363
14	Visit when watering	Stretch	3.677	0.458	0.51
14	Visit when watering	Revert	3.774	0.351	0.325
15	Leaving when watering	Truncate	3.893	0.293	0.299
15	Leaving when watering	Delay	3.664	0.35	0.303
15	Leaving when watering	Cancel	3.287	0.391	0.34
15	Leaving when watering	Execute	3.874	0.322	0.283
15	Leaving when watering	Pause	3.506	0.328	0.33
15	Leaving when watering	Overwrite	3.89	0.31	0.294
15	Leaving when watering	Resume	3.71	0.487	0.434
15	Leaving when watering	Ignore	3.804	0.322	0.317
15	Leaving when watering	Stretch	3.678	0.505	0.468
15	Leaving when watering	Revert	3.746	0.338	0.293
16	Doorbell during nap	Truncate	3.91	0.274	0.286
16	Doorbell during nap	Delay	3.738	0.318	0.281
16	Doorbell during nap	Cancel	3.635	0.33	0.287
16	Doorbell during nap	Execute	3.893	0.309	0.274
16	Doorbell during nap	Pause	3.651	0.32	0.298
16	Doorbell during nap	Overwrite	3.916	0.279	0.271
16	Doorbell during nap	Resume	3.637	0.318	0.292
16	Doorbell during nap	Ignore	3.665	0.325	0.294
16	Doorbell during nap	Stretch	3.688	0.512	0.487

Continued on next page

Table 9 – continued from previous page

ID	Scenario Name	Strategy	Mean	Lower Bound (2.5%)	Upper Bound (97.5%)
16	Doorbell during nap	Revert	3.602	0.319	0.299
17	Meeting during news	Truncate	3.946	0.315	0.331
17	Meeting during news	Delay	3.598	0.382	0.331
17	Meeting during news	Cancel	3.515	0.35	0.389
17	Meeting during news	Execute	3.927	0.328	0.3
17	Meeting during news	Pause	3.93	0.327	0.321
17	Meeting during news	Overwrite	3.948	0.32	0.311
17	Meeting during news	Resume	3.726	0.455	0.466
17	Meeting during news	Ignore	3.443	0.402	0.346
17	Meeting during news	Stretch	3.691	0.469	0.482
17	Meeting during news	Revert	3.693	0.351	0.333
18	Get up early1	Truncate	3.858	0.307	0.246
18	Get up early1	Delay	3.684	0.29	0.291
18	Get up early1	Cancel	3.543	0.294	0.295
18	Get up early1	Execute	3.899	0.288	0.269
18	Get up early1	Pause	3.901	0.278	0.262
18	Get up early1	Overwrite	3.916	0.277	0.255
18	Get up early1	Resume	3.739	0.307	0.27
18	Get up early1	Ignore	3.637	0.451	0.487
18	Get up early1	Stretch	3.701	0.482	0.451
18	Get up early1	Revert	3.672	0.449	0.446
19	Trash out right after Leaving	Truncate	3.922	0.289	0.281
19	Trash out right after Leaving	Delay	3.772	0.325	0.288
19	Trash out right after Leaving	Cancel	3.29	0.335	0.358
19	Trash out right after Leaving	Execute	3.907	0.276	0.292
19	Trash out right after Leaving	Pause	3.903	0.295	0.292
19	Trash out right after Leaving	Overwrite	3.924	0.298	0.293
19	Trash out right after Leaving	Resume	3.854	0.294	0.302
19	Trash out right after Leaving	Ignore	3.851	0.296	0.305
19	Trash out right after Leaving	Stretch	3.713	0.512	0.47
19	Trash out right after Leaving	Revert	3.621	0.326	0.305
20	Hot when baking	Truncate	3.893	0.325	0.315
20	Hot when baking	Delay	3.863	0.37	0.318
20	Hot when baking	Cancel	3.459	0.373	0.388
20	Hot when baking	Execute	4.071	0.32	0.298
20	Hot when baking	Pause	4.074	0.326	0.286
20	Hot when baking	Overwrite	4.089	0.301	0.284

Continued on next page

Table 9 – continued from previous page

ID	Scenario Name	Strategy	Mean	Lower Bound (2.5%)	Upper Bound (97.5%)
20	Hot when baking	Resume	4.056	0.299	0.347
20	Hot when baking	Ignore	3.478	0.386	0.382
20	Hot when baking	Stretch	3.711	0.358	0.358
20	Hot when baking	Revert	3.713	0.49	0.424

D OTHER QUESTION TYPES PRESENTED IN OUR SURVEY

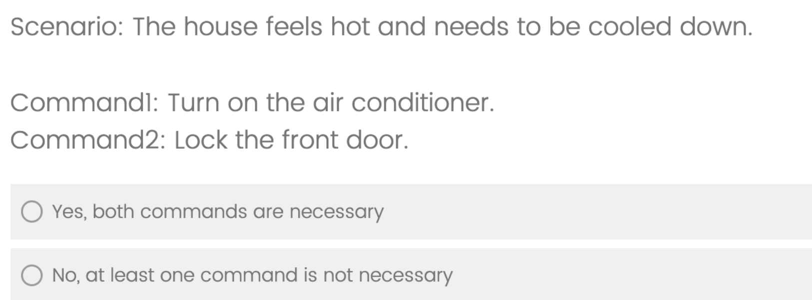


Fig. 4. A participant must determine if both smart device commands are needed to accomplish the goal outlined in the Scenario.

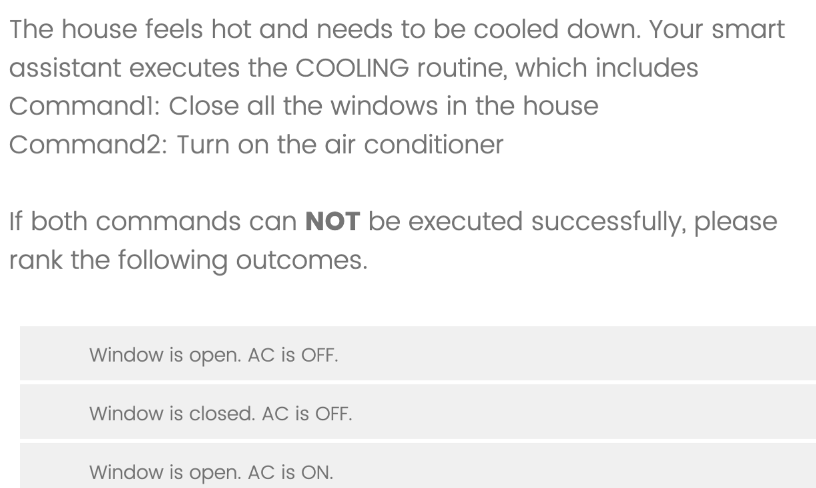


Fig. 5. A participant must rank the following 3 outcomes, described solely by the end-states of the two smart devices mentioned in the scenario.