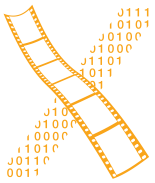


Chair for Computer Science 10 (Media Computing and Human-Computer Interaction)

RWTH AACHEN UNIVERSITY

# *Can Smart Home Action Multiplexing Solve Priority Conflicts?*

Bachelor's Thesis
submitted to the
Media Computing Group
Prof. Dr. Jan Borchers
Computer Science Department
RWTH Aachen University

*by*
*Ilinca Baicu*

Thesis advisor:
Prof. Dr. Jan Borchers

Second examiner:
Prof. Dr. Ulrik Schroeder

Registration date: 24.11.2022
Submission date: 31.03.2023

# Contents

# List of Figures

# List of Tables

# Abstract

Smart home users frequently experience conflicts as a result of rules they have created using trigger action programming. This can represent a source of irritation and give a feeling of helplessness to the user. Although there are strategies for avoiding and resolving conflicts, such as allocating priorities, they don't necessarily provide the user with a comprehensive picture of all simultaneously happening events, making it possible for the user to not even realize there was a conflict in the first place.

With the aim to prevent user confusion while yet indicating simultaneous rule triggering, this bachelor's thesis introduces another method called action multiplexing. The concept underlying multiplexing is action mixing as opposed to mutual exclusion, enabling the potential to signalize more simultaneous activities. In order to gain insights into the users' preferences and understanding capacity of conflict resolution methods a user study was conducted. The user study puts to test the participants' surprise level, confidence in their answers and cognitive challenge level while dealing with multiplexing, as well as other conflict resolution methods. Three classes of outcomes as a result of conflicts are tested and compared: state based on the last event to take place; action multiplexing; error indicators.

The conducted user study reveals interesting findings about the users' preferences and comprehension of conflict resolution methods in smart homes. On average, users tend to incline toward techniques that were more familiar and seemed less cognitively demanding, without giving a second thought to whether their understanding of the situation at hand is accurate or not. Furthermore, action multiplexing's suitability as conflict resolution method seemed to be dependent on various factors such as the number of given TAP rules, as well as the nature of the rules and the device that performs the action. These factors and others are presented and discussed in this work.

# Überblick

Smart-Home-Benutzer erleben häufig Konflikte aufgrund von Regeln, die sie mithilfe von Trigger-Action-Programmierung erstellt haben. Dies kann eine Quelle von Verwirrung darstellen und dem Benutzer ein Gefühl der Hilflosigkeit vermitteln. Obwohl es Strategien zur Vermeidung und Lösung von Konflikten gibt, wie beispielsweise die Zuweisung von Prioritäten, liefern sie dem Benutzer nicht unbedingt ein umfassendes Bild aller gleichzeitig stattfindenden Ereignisse. Es kann der Fall sein, dass der Benutzer nicht einmal merkt, dass es zu einem Konflikt gekommen ist.

Um Verwirrung beim Benutzer zu vermeiden und gleichzeitig simultanes Auslösen von Regeln anzuzeigen, stellt diese Bachelorarbeit eine weitere Methode vor, die als Aktionsmultiplexing bezeichnet wird. Das Konzept, das dem Multiplexing zugrunde liegt, ist Aktionsmischung im Gegensatz zu gegenseitigem Ausschluss, wodurch das Potenzial ermöglicht wird, mehrere gleichzeitige Aktivitäten zu signalisieren. Um Einblicke in die Präferenzen und das Verständnis der Benutzer von Konfliktlösungstechniken zu erhalten, wird eine Benutzerstudie durchgeführt. Die Benutzerstudie soll die Praktikabilität und das Verständnis des Multiplexing durch den Endbenutzer testen. Es werden drei Klassen von Ereignisse getestet und verglichen: Zustand basierend auf dem letzten stattfindenden Ereignis; Aktionsmultiplexing; Indikatoren für Fehler.

Die durchgeführte Nutzerstudie liefert interessante Erkenntnisse über die Präferenzen und das Verständnis der Nutzer von Konfliktlösungstechniken im Smart Home. Im Durchschnitt tendieren Benutzer dazu, Methoden zu bevorzugen, die ihnen vertrauter sind und weniger kognitiv anspruchsvoll erscheinen, ohne darüber nachzudenken, ob ihr Verständnis der vorliegenden Situation korrekt ist oder nicht. Darüber hinaus scheint der Nutzen des Aktionsmultiplexing von verschiedenen Faktoren wie der Anzahl der vorgegebenen TAP-Regeln sowie der Art der Regeln und des Geräts, das die Aktion ausführt, abhängig zu sein. Diese und andere Faktoren sind in dieser Arbeit vorgestellt und diskutiert.

# Acknowledgements

Foremost, I would like to express my gratitude to my supervisor Adrian Wagner for his continuous support of my thesis and research. His feedback helped improve my work significantly.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Dr. Borchers and Prof. Dr. Schroeder for their time, for the insightful comments and for examining this thesis.

The completion of this study could not have been possible without all the people who volunteered their time to participate in the user study that researched the need for multiplexing in smart homes.

Last but not least, I would like to thank my family, friends and boyfriend for always providing moral support.

# Conventions

Throughout this thesis, we use the following conventions.

*Text conventions*

Definitions of technical terms or short excursus are set off in colored boxes.

**EXCURSUS:**
Excursus are detailed discussions of a particular point in a book, usually in an appendix, or digressions in a written text.

Definition:
*Excursus*

The whole thesis is written in American English.

In order to make certain tables and graphs easier to understand and analyze, the Likert-Scale values were ascribed numbers, 1 representing the best option and 5 representing the worst option.

At some points in the thesis, "we" will be used in place of "I" or passive constructions. This does not imply that this thesis was written by someone other than me and that it is only included for readability purposes.

For reasons of politeness, unidentified third persons are described in plural form.

# Chapter 1

# Introduction

Recently smart homes have gained more popularity.

In order to improve the quality of their lives, more and more people are opting to turn their houses into smart homes, incorporating smart technology into their daily lives according to Gunge and Yalagi [2016]. From light bulbs and thermostats to speakers and curtains, this transformation can take many different forms. Regardless of the devices chosen, they are all designed to respond to human needs and take appropriate action (Miandashti et al. [2020]).

Smart Home Configuration Conflicts can be frustrating for end users and deter them from using a system.

As shown by Miandashti et al. [2020] in their work, users expect certain behavior from their devices and can become quickly irritated when an unwanted event occurs. Users who encounter problems get confused and demotivated, which discourages them from continuing a smart-home lifestyle. Residents of smart homes can establish trigger-action programming (TAP) rules to accomplish the desired behavior by using end-user development. Due to a variety of factors, many users wind up creating rules that are in conflict with one another.

A smart home configuration conflict can be reached in few steps.

The steps that lead to conflicts in smart home rules, typically include the following: First, the user creates a new rule in addition to the ones that already exist and work well together, without considering potential future conflicts. Second, at least two rules that act on the same device get triggered simultaneously, resulting in unexpected

behavior from the device. The final result could be unclear and counter to what the user had hoped for. This thesis concentrates on conflicts caused by two rules that are simultaneously triggered on the same device.

Existing solutions do not provide one that expresses multiple active states.

Existing work (Perumal et al. [2016],Coppers et al. [2022], Miandashti et al. [2020]) focuses on foreseeing and preventing conflicts before they emerge by assigning certain rules a higher priority than others or signaling to the user when something in the set rules might be in conflict. A solution in real-time, that indicates more than one active state one the same device, however, is missing.

Despite already existing technologies, multiplexing might still be necessary.

So far, conflicts can also be detected and avoided via static (Corno et al. [2019]) and dynamic analysis (Chaki and Bouguettaya [2021]). These methods have, however certain shortcomings that can possibly be eliminated using multiplexing. Through static analysis, applications can be tested and assessed without having to run the program, while in the case of dynamic analysis, the testing and assessing happen at runtime. The aim of this thesis is to study if multiplexing is a suitable smart home conflict resolution method. Instead of requiring the user to modify or delete competing rules, a novel method would be to combine them. This might be an alternative way to react to rules that get triggered simultaneously and act on the same device. In the best-case scenario, no further tweaks and adjustments to the rules would be necessary, and the user would still be able to recognize that two rules were triggered simultaneously and offer a favorable outcome.

Confidence and surprise level, as well as cognitive load, related to multiplexing should be defined through a user study.

Multiplexing is a potential solution to conflicts in smart homes. Instead of mutual exclusion, this option might be thought of as action mixing. This way the user is not forced to choose between two conflicting rules by assigning priorities or having to remove one of the rules. This method aims to eliminate as much as possible of the information loss associated with other conflict resolution techniques, as well as other limitations. This concept is intriguing, but it has to be studied if there are any drawbacks. Therefore, a user study should offer information on the user's confidence and surprise level as well as the cognitive load of this potential substitute.

This thesis is structured in five chapters.

Following this introductive chapter, the second chapter deals with related work concerning already existing smart home conflict resolution techniques and their use, as well as their limitations. The first part of the third chapter presents the motivation behind the research for this thesis as well as open questions and hypothesis regarding the newly introduced multiplexing. The study's design and the results are presented in this chapter as well. The user study is conducted using a Philips Hue smart light bulb [1]. In chapter four, the results and findings regarding multiplexing, as well as the other presented conflict signals are discussed and evaluated based on the information gained through the user study. The primary conclusions and contribution of this thesis are summarized in chapter five, along with suggestions for further research.

[1] https://www.philips-hue.com/de-de/explore-hue/how-it-works

# Chapter 2

# Related work

## 2.1 Smart homes and their relevance

More users decide on a smart home due to certain advantages.

After decades of research, homeowners may now add even more comfort to their homes than ever before because of how far modern technology has come (Harper [2006]). As mentioned in multiple publications such as the ones by Ali and Yusuf [2018],Balta-Ozkan et al. [2014], automatic, context-aware technology has swept the market by storm, and as a result, the number of people using smart homes has grown significantly over time. Furthermore, smart home technology has recently become accessible to more people since before they tended to be expensive and complex, as determined by Cook et al. [2012] and Ur et al. [2014]. Residents of a smart home can remotely operate appliances, lighting, thermostats, security systems, and other items in and around their living space using a smartphone, tablet, voice assistant, and an internet connection. Therefore, thanks to smart home technologies, homeowners can experience the convenience and financial savings as discovered by Ali and Yusuf [2018].

The Internet of Things (IoT) helps smart homes communicate with various devices.

The Internet of Things (IoT) is a topic of relevance at the moment, as stated by Rose et al. [2015] in their work 'The internet of things: An overview'. It permits connectivity between devices and the cloud and is of great help when it comes to setting up smart homes (Meng et al. [2018],Alaa

et al. [2017]). According to Atzori et al. [2010], IoT devices are a component of the broader idea of home automation. Because of the role IoT plays in the smart home configuration, this technology is associated with the idea of a smart home on the consumer market and makes up an increasing share of it, as noted by Korneeva et al. [2021].

## 2.2 TAP, a paradigm for end-user development

The end-user plays an important role in setting up the smart home using TAP rules.

By establishing rules, the user themselves contributes significantly to the setting of a smart home. This action is referred to as Trigger-action programming (TAP), a paradigm for end-user development, in the paper with the same name by Lieberman et al. [2006]. The user creates rules in order to achieve the desired behavior in the house. These rules have the form: IF [trigger] WHILE [conditions] THEN [action], as also described by Fernandes et al. [2018]. Based on the user-specified conditions, the trigger causes the rule to be activated, and when the rule fires, the specified action is carried out.

TAP is of high significance in making smart homes more accessible.

The paper 'Practical Trigger-Action Programming in the Smart Home' by Ur et al. [2014] explains why TAP can be of considerable assistance to the average user. Because TAP allows for customization and enables homeowners to control how their homes behave, smart homes are now significantly more appealing. TAP is crucial because by making the process simpler to grasp and more approachable, it helps fight the stereotype that smart homes and other technologies are only appropriate for people with a technical background, programming abilities, or a passion for technology (Zhao et al. [2021]). Furthermore, TAP can convey the majority of user-desired behavior, which makes it a significant asset, as established by Ur et al. [2014] in their user study, by Dey et al. [2006] and highlighted by Huang and Cakmak [2015] in their work.

## 2.3 Conflicts in smart home setups

While configuring smart homes, bugs might arise. They were grouped in classes and types by Brackenbury et al. [2019].

Users having the opportunity to configure their smart homes also comes with possible risks. The paper 'How Users Interpret Bugs in Trigger-Action Programming' by Brackenbury et al. [2019] discuss the possibility of conflicts emerging as a result of the increased complexity of the user's generated rules and categorizes these conflicts into 3 classes and 10 types, as seen in table 2.1. The mentioned table provides a short description of the 10 types of bugs, the temporal paradigm and also gives examples for some of them. While the control-flow bugs and inaccurate user expectation classes represent a challenge for end-users and should not be overlooked, in this thesis, of interest are the timing bugs class and priority conflicts from the inaccurate user expectation class. They help build a foundation for this topic.

Nondeterministic timing bugs appear because the system can not handle two triggers being active simultaneously.

Brackenbury et al. [2019] have determined that the system's inability to handle two triggers being active at once leads to nondeterministic timing bugs. There is a clear need for a solution for such situations because these kinds of scenarios can arise frequently in daily life, particularly when a novice user is involved, and can result in unanticipated behaviors. For example, if a user sets a rule: "IF the clock strikes 6 am AND the sun rises WITHIN 10 minutes, THEN open the window" and then also creates a second rule: "IF the clock strikes 6:05 am AND the sun rises WITHIN 5 minutes, THEN close the window". A sunrise that happens at 6:10 am triggers both rules simultaneously, creating an undesired and confusing situation for the user as seen in Figure 2.1. So far this type of bug can be detected and prevented to a certain extent by applying static or dynamic analysis, as brought up in the same paper.

Time is an important factor that often gets overlooked by users when creating TAP rules, therefore causing difficulties.

Brackenbury et al. [2019] also hint at the fact that time is a crucial component and neglecting it frequently leads to unexpected events. In case of extended action bugs, a rule will continuously trigger as long as the action has not been fully executed yet. As a result, because the activity takes a given length of time to complete, the rule will continually be activated. Suppose that the user creates a rule that calls for

| Bug Name | Paradigm | Description |
|---|---|---|
| **Control-flow bugs** | | |
| Infinite Loop | All | One rule triggers another rule, which then triggers the first, ad infinitum. |
| Contradictory Action | All | An infinite loop over an extended period. (e.g. alternating between heating and cooling) |
| Repeated Triggering | All | A rule repeatedly triggers because an event occurs many times or a state remains true. |
| **Timing bugs** | | |
| Nondeterministic Timing | All | The order in which nearly concurrent events are processed changes the system's behavior. |
| Extended Action | State–State | Rules fail to account for the extended timing of an action (e.g. brewing coffee). |
| **Inaccurate user expectation** | | |
| Missing Reversal | All | Even when no rule exists to undo an action (e.g. turning on lights), users assume one exists. |
| Secure-Default Bias | All | Users assume an action is performed (e.g. locks are locked) because doing so is more secure. |
| Time-Window Fallacy | Event–Event | The specified time window is ignored in favor of a more intuitive interpretation. |
| Priority Conflict | State–State | Users ignore the stated priorities of rules in favor of what would make the rules match intent. |
| Flipped Triggers | Event–Event, Event-State | In trigger conjunctions, assuming a reversal of the triggers could still trigger |

**Table 2.1:** Visual representation of the three bug classes and 10 bug types. Brackenbury et al. [2019]

a taxi as long as none arrived yet. There is a strong likelihood that many taxis will arrive even though only one was necessary because this rule continues to be triggered when a taxi is on the way but not yet in front of the building. A visual representation of the situation can be found in Figure 2.2. Even though calling only one taxi at a time could appear logical to the user, in their view the time factor is being overlooked and the created rule has an unexpected result. Brackenbury et al. [2019] indicate that a solution to prevent this type of bug would be static analysis.

Priority conflicts have been widely discussed.

Priority conflicts are also of great interest for this thesis. They have already been studied and debated in various works, including Sikder et al. [2020],Coppers et al. [2022], Manca et al. [2019]. A priority conflict arises when multi-

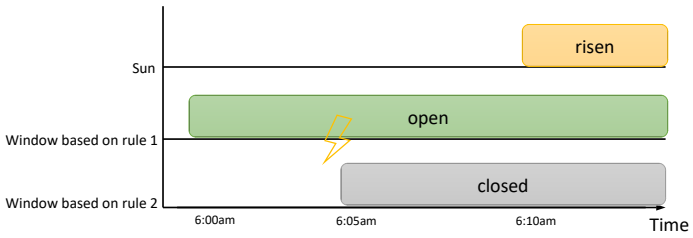



**Figure 2.1:** Example of Nondeterministic timing bug caused by two rules acting on the window.

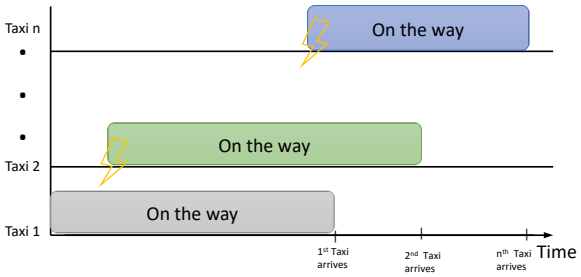



**Figure 2.2:** Example of Extended action bugs, where multiple taxis are called.

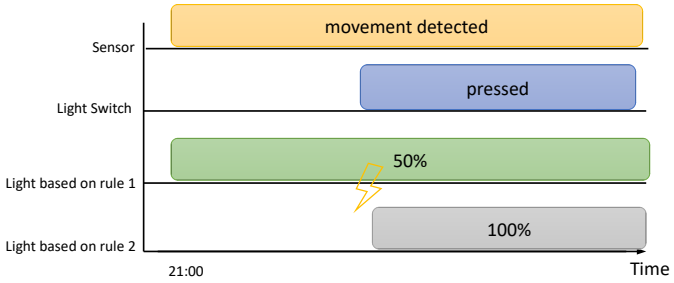



**Figure 2.3:** Example of Priority Conflict caused by the rules for a night with a full moon.

ple rules attempt to bring the same device into a particular state at the same time by acting on it. For instance, if a user decides to customize the light intensity in their apartment based on various factors and establishes the following rules: ”IF it is night, THEN light intensity is at 50 % ” and ” IF it is a full moon, THEN light intensity at 20%’, a conflict will arise on a night with a full moon. This is caused by the fact that both rules act on the same device but try to bring it into different states, as can be deduced from Figure 2.3. In ’Supporting end-user debugging of trigger-action rules for IoT applications’ by Manca et al. [2019], it was stated that assigning priorities is the most effective way to address this issue.

## 2.4 Solutions to Smart Home Conflicts

So far, it seems like most conflict resolution in smart homes is realized through anticipating and foreseeing possible clashes among rules, as deduced from papers by Coppers et al. [2022],Corno et al. [2019],Zhang et al. [2019],Fiorenza

and Mariani [2015],Chaki and Bouguettaya [2021]. This is possible using static or dynamic analysis.

Static analyses can be used for conflict resolution but it also has limitations.

Static analyses are carried out without the need to run the rules first (Ayewah et al. [2008]). The paper 'Empowering End Users in Debugging Trigger-Action Rules' by Corno et al. [2019] puts the static analysis to use and presents a system that assists end-user in debugging trigger-action rules. It helps them identify rule inconsistencies and assists them in anticipating the behavior of their rules during run-time through step-by-step simulation. This system is designed to aid the user during the creation of the desired rules. However, if a conflict is detected, the user's options are to either delete one of the conflicting rules and leave the other one in place or reformulate at least one of the rules so that it no longer causes a conflict. Although beneficial, a solution like this is not ideal because the user must make concessions.

Dynamic analysis is another option for conflict resolution but has certain shortcomings.

Conflict resolution in smart homes has also been successfully accomplished via dynamic analysis. It executes the rules, as opposed to the static analysis, and looks for conflicts. Coppers et al. [2022] presents a technique in their work 'FortClash: Predicting and Mediating Unintended Behavior in Home Automation', using this kind of analysis. The idea of one-time exceptions, which allow the user to temporarily suppress a rule's action, was developed because users occasionally struggle to come up with a general solution or reformulation of the rules in order to avoid conflicts. As might be expected, there are certain limitations as well, because the user still has to decide which one of the conflicting rules to suppress.

Conflicting rules can be avoided by using AutoTap, presented by Zhang et al. [2019].

Another interesting approach to conflict resolution is presented by Zhang et al. [2019] in their paper 'AutoTap: Synthesizing and Repairing Trigger-Action Programs Using LTL Properties'. In order to avoid conflicts occurring, the user has the option to state what requirements the system must meet, instead of giving event-driven rules. The task of the user becomes therefore easier since the TAP rules can be automatically generated to fit the desired behavior, as well as already given rules being checked and analyzed. A limitation of this method, as stated in the paper, is the fact

that it sometimes fails to synthesize a TAP program.

Introducing multiplexing as a potential substitute in order to address other method's limitations.

In this thesis, the concept of multiplexing is introduced to address these drawbacks. Multiplexing provides the user with a different method of addressing conflicts in comparison to the ones suggested in numerous publications over the years because it may be seen as action mixing rather than mutual exclusion. This could be an aspect of making smart homes and smart technologies even more appealing to consumers since the user isn't required to devote that much thought and attention to creating rules.
To test the applicability and how the end-user might perceive this method, a user study is conducted and then presented in the following chapter.

# Chapter 3

# User Study

The user study is prepared once the pilot study has been written and carried out. It aims to assess the user's confidence and surprise level as well as the cognitive load of certain smart home conflict resolution methods and propose multiplexing as a potential additional resolution method. The Philips Hue smart light bulb [1] utilized to show participants specific scenarios and settings was essential to carrying out this study. It is displayed in Figure 3.1. With this technology, the light bulb's color, brightness and on/off function can be controlled remotely using an installed application on the user's smartphone.

The user study makes use of a smart light bulb to present conflict resolution techniques.

## 3.1 Research Motivation

As established in the chapters before and mentioned by Zhang et al. [2023], TAP is now widely utilized. Consequently, many users create rules that are in conflict with one another. This may result in a variety of undesired behaviors. Conflicts in smart homes can occur at any given time and might vary in difficulty to resolve. When such conflicts arise, the user may become irritated because they are unable to determine what caused or how to resolve them. In order to better understand what people anticipate would

The user study aims to gain insight into users' understanding of conflicts by presenting everyday scenarios.

[1] https://www.philips-hue.com/de-de/explore-hue/how-it-works


**Figure 3.1:** Light bulb used to carry out the study

occur in the event of a conflict and the best course of action, this study was driven by common scenarios that users may encounter while establishing their smart homes.

A good illustration of a conflict that may arise in daily life is two rules trying to bring the same light bulb into a certain state at the same time.

Two TAP rules that act on the same light bulb can serve as a very good illustration of such a situation. If the user creates the rules in this scenario without taking into consideration both of them, they may both be triggered simultaneously and attempt to change the status of the light bulb at the same time. Therefore, if the user wanted the light bulb to become yellow when the door is open but blue when the window is open, then both of them being open at the same time would result in a conflict and leave the light bulb in an uncertain state. As previously indicated, there are already potential answers to this kind of conflict, but they are predicated on anticipating the conflict.

Another example that was taken into consideration in-

volves two rules that both operate on the same light bulb, but this time also incorporate a motion sensor: "IF the light switch is on, THEN turn on bright lights." and, "IF the sensor detects motion WHILE it is night, THEN turn on dimmed lights." These rules were given in order to minimize discomfort to the eyes while allowing residents to perform certain tasks at night, such as drinking a glass of water. If a person is identified by the motion sensor at night and also flips on the switch because of the need for brighter light, there may be a conflict. In a situation like this, a conflict between manual intervention and automatism arises.

Another example is a conflict between automatism and manual intervention.

From these types of situations emerges the need to understand what the average end-user prefers and is more familiar with when it comes to smart home conflict resolution. A possible way to do so is to let people confront themselves with daily potential conflicts and provide them with different outcomes and solutions, in order to see their comprehension of their surroundings and of the just occurring events. Different users will have different opinions and preferences when it comes to dealing with conflicts. Therefore, conducting a user survey and allowing the participants to voice their perspectives on the same occurring conflicts, possible outcomes, and potential solutions is an excellent method to keep track of these preferences and opinions.

User preferences and understanding of smart home conflict resolution play an important role and should not be underestimated.

## 3.2 Research Questions

The research questions we attempt to tackle in this thesis are:

- What brings more understanding to the user, action multiplexing or other smart home non-priority conflict resolution methods?
- What is better perceived by the user, action multiplexing or other smart home non-priority conflict resolution methods?

- Is action multiplexing a suitable solution for smart home conflicts?

## 3.3 Hypothesis

Definition: *Hypotheses of the user study*

**HYPOTHESES OF THE USER STUDY:**
The premises of this study are:

H1: Action multiplexing lets users be more confident in their comprehension of smart home conflicts than other non-priority conflict resolution methods, where conflicts occur.

H2: Action multiplexing makes the occurrence of conflicts less surprising for users than other non-priority conflict resolution methods, where conflicts occur.

H3: Action multiplexing is less cognitively challenging than other non-priority conflict resolution methods, where conflicts occur.

## 3.4 Study Design

The user study is composed of two sets of situations and a questionnaire. The order of testing the situations is randomized.

All presented conflict resolution methods are evaluated during the user study using the same sets of questions, through a questionnaire. The first set of situations requires the participants to identify the events that take place in the room based on two provided TAP rules and the color change in the light bulb. The second set of situations requires the participants to also identify the events that take place in the room based on two given TAP rules but this time using the changing brightness level of the light bulb. The order in which each participant receives the set of situations as well as the situation itself is entirely randomized to ensure a significant outcome. After each situation, the participants have to answer a series of questions regarding

it.

The original survey was composed of six sets of situations and required multiple devices, such as a thermostat, a speaker and smart curtains. Following the first conducted pilot study the survey is noticeably condensed due to time issues and in order to keep the focus and motivation of the participants as high as possible for optimal results.

Following the pilot user study, the number of sets of situations is reduced.

### 3.4.1 Setup and Surrounding

- A Philips Hue smart light bulb attached to a lamp ( as seen in Figure 3.1)
- The questionnaire containing the questions the participants have to answer in writing ( can be found in appendix A "USER STUDY" )
- A silent room where the study can be conducted without any interruptions so that the participants can fully focus and take part in the survey without any interference or influence from the outside

## 3.5 Study Procedure

First of all, all of the equipment needed for the study, such as the survey and the lamp with the smart light bulb and the study set up as a whole, is ready and set out on the table. The survey is included in appendix A "USER STUDY".

The components of the study are prepared.

Once the participant is comfortably sitting at the table, they are asked to read and sign the consent form, which can be found in appendix A "USER STUDY" in figure A.1 and give their consent to be recorded. This is a crucial phase because it contributes to the accuracy of the study's evaluation. The participant is then given a description of the study and the tasks they must do after this stage. Throughout the study, the participant learns more about each task and is always free to ask questions. Since the study is heavily based on

Users are informed of the study's methodology and what is expected of them.

**Figure 3.2:** Colors of the light bulb used throughout the study.

**Figure 3.3:** Intensities of the light bulb used throughout the study.

color and brightness level changes, to improve the likelihood that the participant would be able to distinguish between and recognize them, part of the colors and brightness levels used in the study are shown to the participant beforehand. In order to execute the study correctly some colors and brightness levels are not presented to the participant. However, it is mentioned that other colors and brightness levels than the ones shown might appear during the study.

For the user study, five colors and five intensities are chosen.

The exact colors and brightness levels used for this study can be seen in Figures 3.2 and 3.3. The default setting for the study's color part is cool white, whereas the TAP rules specify the colors yellow for the door and blue for the window. Red and green are included in the exhibited colors as conflict signs due to mixing and error state.
For the study's intensity part, a default setting of 10% is used, while dimmed lights are set at 50% and bright lights at 100%. In addition, 30% represents an error state through random intensity, and 75% represents the mixing of intensities.

**CLASSES:**
In order to better understand the end user's preferences, four categories of conflict resolution methods were presented in this user study:

- Prioritizing: priorities were assigned to the given TAP rules
- Multiplexing: consisting of mixing and alternating, indicates that two states are active simultaneously
- Error state: light bulb turning off, red for the color part of the study and random intensity level (30%) for the intensity part of the study, indicating error caused by two rules getting triggered simultaneously
- Last event to take place: light bulb maintains the state in which it was brought by the last event to take place

Definition:
*Classes*

The study begins when the participant is given the first set of TAP rules and presented with the first situation. For each set, composed of 16 situations, the participant has to answer a question regarding the understanding of the TAP rules, 4 questions for each situation they saw and 5 general questions at the end of each set of situations. During this time, the participant's reactions is observed and noted, as well as any made comments regarding the situation at hand. In the end, the participant is asked to answer the general questions, which concern a comparison of all the non-priority presented conflict resolution methods. The participant has to observe all outcomes caused by a conflict in the given rules and in the end, has to rank them based on personal preference. In order to fully analyze the results, this comparison is crucial.

Participants have to answer a questionnaire while their actions are observed.

## 3.6 Subjects

The participants' age range is 20 to 27.

This study includes twenty-four participants, aged 20–27. Seven females and seventeen males took part in it. The mean of the participant's age is 22.6, and the standard deviation is 2. One of the male participants had trouble telling the difference between red and green because he is color-blind. Ten of the participants are smart home users, five have plans to get smart home devices, and nine have no interest in getting any in the near future. Of the smart home users, three stated that they almost never configure their smart home devices; two did so yearly, other two configure their devices several times a year; and only one did so monthly, several times a month, and several times a week. When configuring their smart home gadgets, the participants mentioned coming in contact with Amazon Alexa, Google Nest, Philips Hue, HomeKit, as well as other applications.

Not all participants are smart home users in their daily lives.

## 3.7 Results

The study is reviewed with the aid of the recordings, the questionnaire, and the study notes.

The study's findings are assessed using the participants' comments, the audio recordings, and the responses to the questionnaire. In the event that I wasn't able to take down every single comment, the participant stated throughout the study, the recordings of the sessions are helpful. Statistical tests are used to analyze the Likert Scales and the points that participants used to assess the various presented situations that signalize conflicts.

Participants are asked to rate how simple to understand they considered the provided TAP rules to be.

The first step is for the participants to rate how easy it is to understand the presented TAP rules using a Likert scale. In order to compare how understandable the two sets of rules are for the participants, numbers are attributed to the answers, 1 denoting "very easy" and 5 denoting "very hard" (to understand). The overall score of the TAP rules understanding for the light color part, as well as the light intensity part of the study, can be seen in figure 3.4.

After seeing both sets of situations, the participants in this

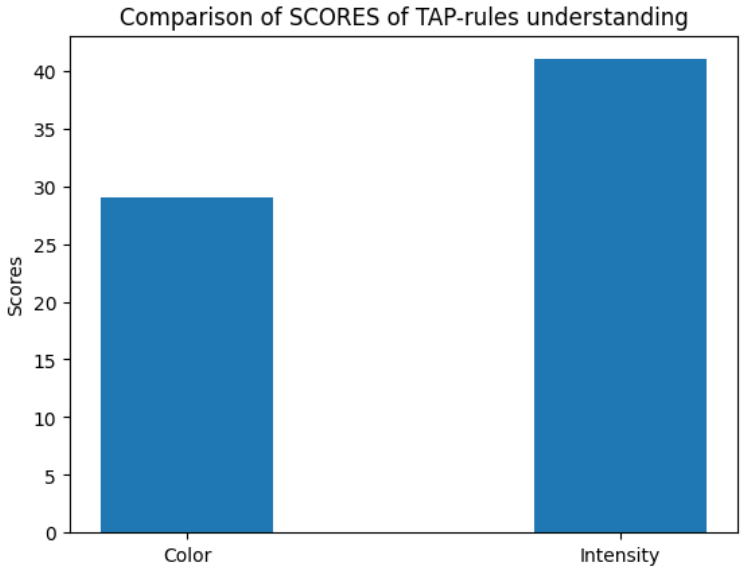



**Figure 3.4:** Summed scores of understanding of provided TAP rules, according to participants.

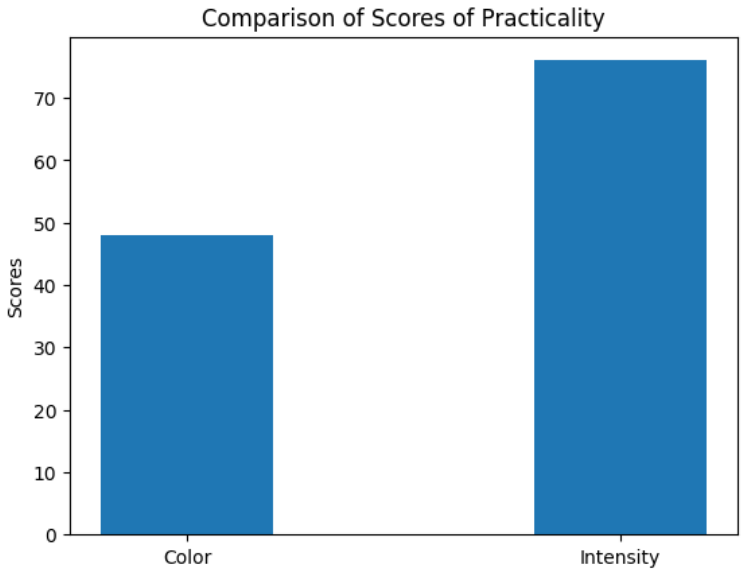



**Figure 3.5:** Summed scores of light color and light intensity practicality, according to participants.

Participants rate the practicality of color and intensity as event indicators using a Likert scale.

study are asked to rate the practicality of light color and light intensity as indicators of events, using a Likert scale. Five answers on the Likert scale, ranging from ″very practical″ to ″very impractical″ are available. In order to compare the practicality of the light color and light intensity, numbers are attributed to the Likert Scale answers, 1 representing ″very practical″ and 5 representing ″very impractical″. Depicted in figure 3.5 is the summed score of the participants' answers for both cases as a comparison.

Definition: *Presented conflict resolution methods/indicators*

**PRESENTED CONFLICT RESOLUTION METHODS/INDICATORS:**
In the next chapters, the following short forms will be used in order to refer to situations from the user study:

*(D)*: Just the door opens.
*(W)*: Just the window opens.
*(DF)*: Door opens first.
*(WF)*: Window opens first.

Non-priority situations:
*OFF*: The light bulb turns off.
*RED*: The light bulb turns red.
*LE*: Light keeps color/intensity based on the last event to take place.
*ALT*: The light bulb alternates between colors/intensities.
*MIX/75%*: Colors/Intensities get mixed together.
*M/100%*: (manual) The light switch gets pressed by the user, triggering 100% light intensity.
*A/50%*: (auto) Motion sensor detects movement, triggering 50% light intensity.
*(AF)*: Automatism intervenes first.
*(MF)*: Manual intervention first

### 3.7.1 Statistical tests

H0 assumes the following:

**H0 HYPOTHESES FOR STATISTICAL TEST:**
H0: There is no significant difference between the answers for the question of the user study for the seven situations where priority was not assigned.

Definition: *H0 HYPOTHESES FOR STATISTICAL TEST*

We will be running the Friedman Test on all seven situations from each part of the user study, where priorities were not assigned: *D, W, ALT, MIX, OFF, RED/30%, LE* for each of the three questions from the user study. The predetermined level of significance is $\alpha$ = 0.05. The situations which are composed out of *(WF)* and *(DF)* or *(AF)* and *(MF)* were merged into only one situation, using the mean of the two, in order to make the analyzing of data easier to read. An overview of the results of the Friedman test can be seen in table 3.1

The Friedman test is performed in order to find out if there are significant differences among the situations without assigned priorities.

The Friedman test reveals a statistically significant difference in participants' answers across the seven situations regarding the second question of the user study, for the color part. In fact, the p-value is very small, with p = $1.09548e^{-9}$ and statistic = 53.14796, therefore indicating strong evidence against the null hypothesis.
A statistically significant difference in participants' answers across the seven situations for the second question of the user study, for the intensity part, is revealed by the Friedman test as well. The p-value is in this case also very small (p = 3.79766e$^{-5}$; statistic = 30.079395), which means that the H0 hypothesis must be rejected.

H0 is rejected with strong evidence against it.

In the case of the third user study question for the color part, there is also a statistically significant difference in participants' answers. The p-value is very small (p = 7.02032e$^{-17}$, statistic=88.24237) and the H0 hypothesis has to be rejected. There is also a statistically significant difference in the responses given by participants for the third user study question for the intensity part. Because the p-value is so low (p = 1.47175e$^{-8}$, statistic = 47.52254), the H0

H0 is rejected.

hypothesis must be rejected.

H0 is rejected.

Regarding the fourth question of the user study, there is a statistically significant difference in participants' answers for both the color and intensity part of the user study. The color part scored a very low value (p = 1.18135e$^{-13}$, statistic = 72.62644) and therefore indicates strong evidence against the null hypothesis. The intensity part of the study shows a low p-value as well ( p = 0.00041, statistic=24.54197), also providing evidence against the null hypothesis.

| | statistic value | p-value |
|---|---|---|
| Second question (COLOR) | 53.14796 | $1.09548e^{-9}$ |
| Second question (INTENSITY) | 30.07939 | 3.79766e$^{-5}$ |
| Third question (COLOR) | 88.24237 | 7.02032e$^{-17}$ |
| Third question (INTENSITY) | 47.52254 | 1.47175e$^{-8}$ |
| Fourth question (COLOR) | 72.62644 | 1.18135e$^{-13}$ |
| Fourth question (INTENSITY) | 24.54197 | 0.00041 |

**Table 3.1:** Overview of the statistic and p values of the Friedman test (rounded to 5 decimals).

H1 is rejected.

A post-hoc analysis using Wilcoxon signed rank tests is used to evaluate the confidence level in participants' answers for the other presented situations, where priorities were not assigned, compared to multiplexing, for the color part of the user study. A table showing an overview of the p-values of the Wilcoxon signed rank test can be seen in B "Results of the user study" in B.13 and B.14. Applying the Wilcoxon signed ranked test, no significant difference between mixing and other non-priority conflict resolution methods is found, as well as between alternating and other non-priority conflict resolution methods
The same is being done for the intensity part of the study. By applying the Wilcoxon test, it shows that there is not a significant difference between multiplexing and other presented situations where priorities were not assigned. Therefore, multiplexing does not necessarily let users be more confident in their comprehension of smart home conflicts than other conflict resolution methods. On these grounds, H1 must be rejected because H1 claims that multiplexing lets users be more confident in their comprehension of smart home conflicts than other non-priority conflict res-

olution methods.

H2 is uncertain.

To evaluate the surprise level multiplexing brings to the user vs. the surprise level the other non-priority conflict resolution methods bring, the Wilcoxon signed ranked test is used again: for the color part of the study, there is a significant difference between mixing compared to last event to take place (z = 2 ; p = 0.000153), as well as between alternating and the light turning off (58 = 0 ; p = 0.042332), the light turning red (z = 21.50 ; p = 0.004627) and the last event to take place (z = 5 ; p = 0.001633). As for the intensity part of the study, regarding mixing, the outcomes are similar: there is a significant difference between mixing compared to last event to take place (z = 20 ; p = 0.011227). Significant differences were also identified through the Wilcoxon test between alternating and last event to take place (z = 9.50 ; p = 0.003869). Therefore, only alternating in the color part of the study made the occurrence of conflicts less surprising for the user and consequently, H2 can't be proved nor rejected because it can't be considered clarified. A table showing an overview of the p-values of the Wilcoxon signed rank test can be seen in B "Results of the user study" in B.15 and B.16.

H3 is retained.

Finally, we use the Wilcoxon Signed Ranked test to asses the cognitive load of multiplexing vs. other techniques. A table presenting an overview of the p-values of the Wilcoxon signed-rank test can be seen in B "Results of the user study" in B.17 and B.18. For the color part of the study, there is a significant difference between multiplexing compared to last event to take place (z = 8 ; p = 0.004500 for mixing) (z = 8 ; p = 0.007618 for alternating).
For the intensity part of the study, there is a significant difference identified between multiplexing and last event to take place (z = 22 ; p = 0.026291 for mixing) (z = 9.50 ; p = 0.016238 for alternating). Therefore, multiplexing does not necessarily mean less cognitive load than other conflict resolution methods. On these grounds, H3 must be retained.

### 3.7.2 Participants' confidence level

The participants' confidence level is analyzed by looking at their answers to the second question of the study.

In order to gain insight into the participants' level of confidence while confronting themselves with the presented situations and the occurring events during the study, we have to look at and analyze their answers to the second question of the questionnaire.The participants answered this question using a Likert scale. The responses to the second survey question for the study's color part, "How confident are you in the answer given above?" for all scenarios, are summarized in tables B.1 and B.2 and for the intensity part in tables B.3 and B.4 that can be found in appendix B "Results of the user study".

Participants feel most confident in scenarios where priorities are allocated and least confident when the light turns red or off.

For every scenario where the TAP rules have a predetermined priority, the majority of participants seem to be quite confident in their answers. The door opening, the window opening, and *RED(DF)* all exhibit the same level of assurance. In situations like *OFF(WF), LE(WF), ALT(DF)* and *MIX(WF)*, participants show less confidence. The worst levels of confidence is displayed in situations like *OFF(DF), LE(DF), MIX(DF), ALT(WF)* and *RED(WF)*. Similar to the mode, the median supports the pattern that scenarios such as the ones with assigned priorities or the simple ones, where only the door or the window opens score higher in confidence, while scenarios such as *OFF(WF), LE(WF), ALT(DF), MIX(WF), LE(DF), MIX(DF), ALT(WF)* and *RED(WF)* show less confidence with medians of 2 and 2.5 respectively. The remaining situations *OFF(DF)* and *RED(DF)* have a median of 3. Participants state that the situations where the light turns red or the situation where the light turns off are very confusing.

Most confidence is generally shown when the TAP rules are given priority, and least confidence is displayed when the light is turned off or presented at a random intensity of 30%.

In this case, participants seem to be very confident in their answers when the TAP rules have assigned priorities too. Priorities, according to one participant, are really useful. Participants maintain their confidence in other scenarios including *M, A, OFF(MF), ALT(AF), LE(AF)* and *ALT(MF)*. Less confidence is shown in the following situations: *75%(MF), LE(MF), 30%(MF)* and *75%(AF)*. Situations like *OFF(AF)* and *30%(AF)* show that participants are generally the least confident when it comes to them and that

they are more difficult to explain. While *Prio(MF)*, *Prio(AF)* and *M* score a median of 1 and *Prio(A)* and *Prio(M)* one of 1.5, most presented situations have a median of 2 *A*, *75%(MF)*, *ALT(AF)*, *30%(MF)*, *LE(AF)*, *ALT(MF)*,*75%(AF)*, showing confidence among the participants' answers and respecting the trend set by the modus. *OFF(MF)* and *30%(AF)* have a median of 2.5 and 3 respectively, indicating that people are only somewhat confident.

### 3.7.3 Participants' surprise level

The participants' answers to the third question offers insights into how surprised they are by the seen situations.

We must examine and evaluate the participants' responses to the third questionnaire question in order to learn more about how surprised they felt when confronted with the situations that were presented to them, as well as the changes in the light bulb throughout the study. In order to express their level of surprise, participants used a Likert scale. In appendix B "Results of the user study", in tables B.5 and B.6 the summarised answers to the third question: "How surprised were you by what you saw happen as a result of the TAP-rules being triggered?" for the color part of the study, can be found. The results for the same question but for the intensity part of the survey can be seen in tables B.7 and B.8.

The situations in which the light turns off, red, or the colors mix are overall the most surprising.

Participants evaluate their reaction to the following scenarios as "very unsurprised" in most cases: *Prio(D)*, *Prio(W)*, *Prio(DF)*, *D* and *F*, whereas situations such as *Prio(WF)*, *LE(DF)*, *LE(WF)* and *ALT(WF)* scored mostly "unsurprised". After seeing *OFF(DF)*, *MIX(DF)*, *OFF(WF)*, *RED(WF)* and *MIX(WF)*, participants report feeling "somewhat surprised" and "surprised" respectively. One of the participants mentions they would "have no clue what is happening and assume failure" in these scenarios. For all situations except *RED(WF)* and *ALT(WF)*, the median follows the pattern of the mode.

In the intensity part of the study, scenarios in which the light turns off

Participants in this study report being "very unsurprised" by *Prio(A)*, *Prio(AF)*, *Prio(M)*, *M* and *A* and "surprised" by *Prio(MF)*, *75%(MF)*, *LE(MF)*, *ALT(AF)*, *30%(MF)*, *LE(AF)* and *75%(AF)*. *OFF(MF)*, *OFF(AF)* and *ALT(MF)* are re-

and alternating intensities are the most unexpected.

garded as somewhat surprising, however, *30%AF* is rated as surprising since it is perceived as random by the participants. The median confirms the trend that *Prio(A), Prio(AF), Prio(M), M, A* are very unsurprising scenarios, while *Prio(MF), 75%(MF), LE(MF), ALT(AF), 30%(MF), LE(AF), 75%(AF)* are perceived as unsurprising and *OFF(MF), OFF(AF)* and *ALT(MF)* are somewhat surprising.

### 3.7.4 Cognitive Load

Participants express their opinion on the cognitive load level of the situations in the user study.

Participants express their opinion using a Likert Scale regarding how cognitively challenging the presented situations are by answering the fourth question of the user study: "How cognitively challenging is the situation that just occurred?". The answers to this question help us understand how difficult it was for the participants to make sense out of what they see occurring in the room. The results to this question can also be found in appendix B "Results of the user study", in tables B.9, B.10, B.11 and B.12

Multiplexing and the light color turning red are regarded as the two instances that were most cognitively demanding out of the presented situations.

Situations such as *Prio(D), Prio(W), Prio(DF), D, W* and *LE(DF)* score mostly "very cognitively unchallenging", while more participants saw *PRIO(WF), OFF(DE), ALT(WF), OFF(WF), RED(DF), LE(WF)* as "cognitively unchallenging" than other situations. As "somewhat cognitively challenging" are mostly seen the following situations: *MIX(DF), ALT(DF), RED(WF)* and *MIX(WF)*. One participant comments that they have not anticipated the mixing of colors and that red and green should not appear in accordance with TAP rules. Regarding the answers to the fourth question, the median supports the pattern of which situations are generally seen as "very cognitively unchallenging" and which situations tend to be more "cognitively challenging" as well.

Participants consider in the intensity part of the study the light turning off more cognitively challenging

The majority of answers for *Prio(A), Prio(MF), Prio(AF), M, A* and *75%(AF)* are "very cognitively unchallenging". Situations such as *Prio(M), 75%(MF), LE(MF), LE(AF)* assess as "cognitively unchallenging". The mode of *OFF(MF), ALT(AF),30%(MF), ALT(MF), 30%(AF)* is 3, indicating that most participants perceive these situations as "somewhat

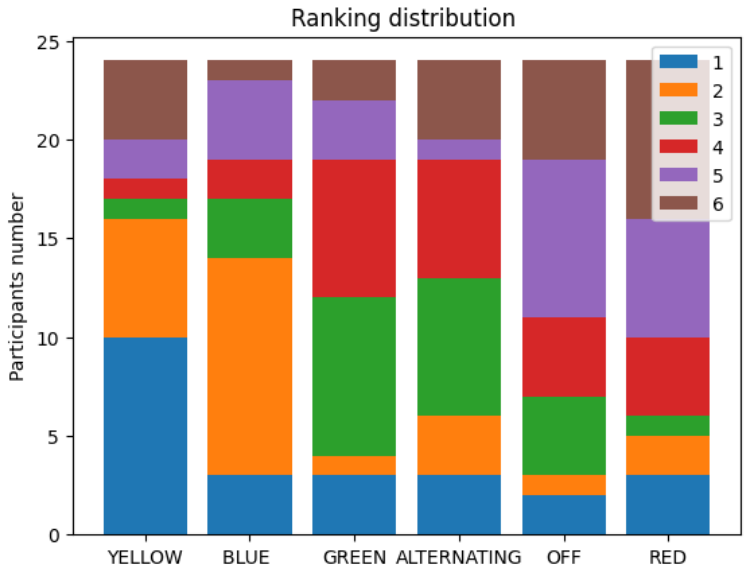



**Figure 3.6:** Distribution of answers expressed using colors for the color part of the study

cognitively challenging", while *OFF(AF)* is considered to be "cognitively challenging", with one of the participants stating that there is no explanation for the light turning off. Similar to the mode, the median confirms the trend that scenarios such as *Prio(A), Prio(MF), Prio(M), Prio(AF), M, A, 75%(MF), LE(AF), LE(MF)* score higher in how cognitively unchallenging they seem to be, while scenarios with medians of 3.5 and 3, such as *OFF(AF), OFF(MF), ALT(AF), ALT(MF), 30%(AF)* seem to be cognitively challenging. The median for *75%(AF)* is 3.5 although multiple participants consider the situation to be "very cognitively unchallenging".

than other presented situations.

### 3.7.5 Color ranking

The participants are asked to rank the presented non-priority situations using distinct numbers from 1 to 6 based on how understandable they appear to be as conflict resolution methods. In this case, 1 represents the most understandable and 6 the least understandable. This ranking

Participants express their understanding of the non-priority situations, as well as opinions through a ranking.

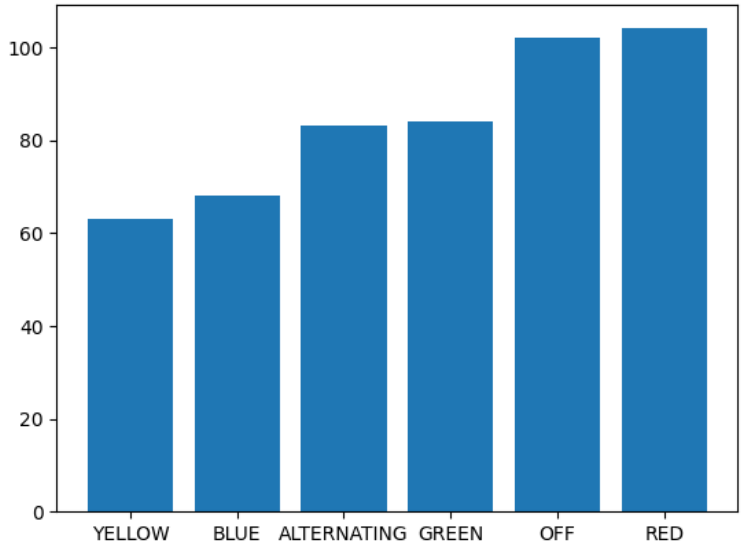



**Figure 3.7:** Ranking's sum of the points for each situation in the color part of the study

offers a good overview of the participants' understanding and perception of the presented situations, as well as offers them an opportunity to voice their opinion, by explaining their choices regarding the ranking.

The ranking and answer distribution for the color part of the study can be seen in figures 3.6 and 3.7.

The color-coded bars displayed in figure 3.6 show the distribution of answers among participants, regarding the non-priority situations. Looking at the y-axis, we can read how many participants chose the response associated with a certain color. The legend regarding the relation between colors and ranking numbers can be seen in the upper right corner of the figure.
In figure 3.7, the height of the bars represent the sum of the given points by the participants, therefore of the overall ranking and the values can be read from the y-axis of the graph.

Yellow and Blue score best in the ranking, while *OFF* and *RED* are positioned last.

Figure 3.7 can be interpreted as follows, since 1 represents the most understandable and 6 the least understandable: Participants argue that the user will anticipate the colors yellow and blue since they are mentioned in the TAP rules, making them the first two most understanding methods of

communicating a conflict. They are followed by alternating between colors and mixing the colors, between which there is not a big difference in points. While some of the participants think that mixing and alternating between colors clearly show that there is a conflict, others find mixing to be perplexing because it causes the appearance of a new color that was not specified in the TAP rules. When it comes to alternating, while some people stated it is a useful approach to represent events happening simultaneously, others judged it as "slightly hard to comprehend and unpleasant". The light turning off and the light turning red are on the last two spots of the ranking. "I didn't really comprehend 100% that both were open at the same time when the lights went red/off", a participant admits.

Participants prefer the color yellow.

Looking at the distribution of the answers, we can observe that ten participants rank yellow as the most intelligible indicator for conflicts (41.67%), six choose the second position (25%) only one person places it in the third place, and one in the fourth place (4.17%). Two participants opt for fifth place (8.33%) and four for sixth one (16.67%).

Participants' answer distribution choice for blue.

Looking at figure 3.6, only three participants (12.5%) choose blue as their first or third option, but eleven (45.83%) choose it as their second. Blue is chosen to fill the fourth spot by two (8.33%), the fifth by four (16.67%), and the sixth spot by just one participant (4.17%).

Participants' preference for alternating.

Only three participants (12.5%) select the alternating-colors scenario as their first or second option. Alternating, according to participants, is a choice that is obvious and most intuitive but may irritate the user. Only one person chooses to place it in third place (4.17%), seven (29.17%) in fourth, six (25%) chose to place it in fifth, and four (16.67%) in sixth.

Participants'answer distribution for color mixing.

The mixing of colors situation scores close to the alternating situation. Three participants (12.5%) also select this option as their first choice; one of them mentions that utilizing a combination of yellow and blue is preferable since it makes the issue easier to understand. It is chosen as the second choice by one person (4.17%) and the third choice by eight people (33.33%). Seven participants (29.17%) decide to select color mixing as their fourth option, with one of them

stating that it is more difficult to distinguish because it resembles yellow and blue. The mixing of colors is ranked fifth by three participants (12.5%) and sixth by the remaining two (8.33%).

Participants did not rank the light turning off very high.

Two participants (8.33%) rank the situation where the light goes out first, with one of them claiming it is the ”most obvious to the sight”. Only one person (4.17%) chooses the light turning off as their second ranking, while four (16.67%) placed it in the third and fourth spot, saying it ”shows when something is wrong”. It receives lower ranks from seven participants (29.17%) who rank it fifth and five (20.83%) who rank it sixth because ”it suggests trouble but the reason is unclear” and one participant seems confused and said they would just assume failure.

Red as conflict resolution method is chosen by the least participants as a top position in the ranking.

The worst score overall is received by the light turning red. Three participants (12.5%) choose to rank it as their most understandable option, only two (8.33%) place it as second and one (4.17%) ranks it third. Four people (16.67%) opt for the light turning red to be their fourth option and six (25%) their fifth. Eight (33.33%) out of the total participants choose to rank it as the least understood circumstance, classifying it as the most unexpected and startling, a little strange, or just not understanding what the color is trying to represent at all.

Participants prefer expected outcomes.

All things considered, it seems like participants tend towards more familiar methods, that involve already in the TAP rules mentioned colors and intensity levels, but also consider multiplexing an understandable conflict resolution method and are able to identify the occurrence of a conflict through it.

### 3.7.6 Intensity ranking

Insight into the participants' understanding and opinions is gained through the ranking.

For the intensity part of the study, participants were asked as well to rank the non-priority seen situations using distinct numbers from 1 to 6 based on how understanding they are as resolution methods. 1 represent the most understanding, while 6 represents the least understanding. As in

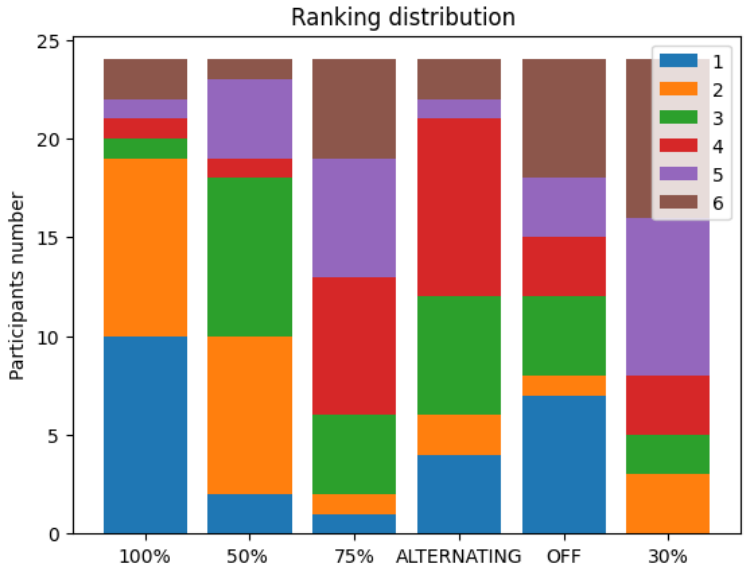



**Figure 3.8:** Distribution of answers expressed using colors for the intensity part of the study

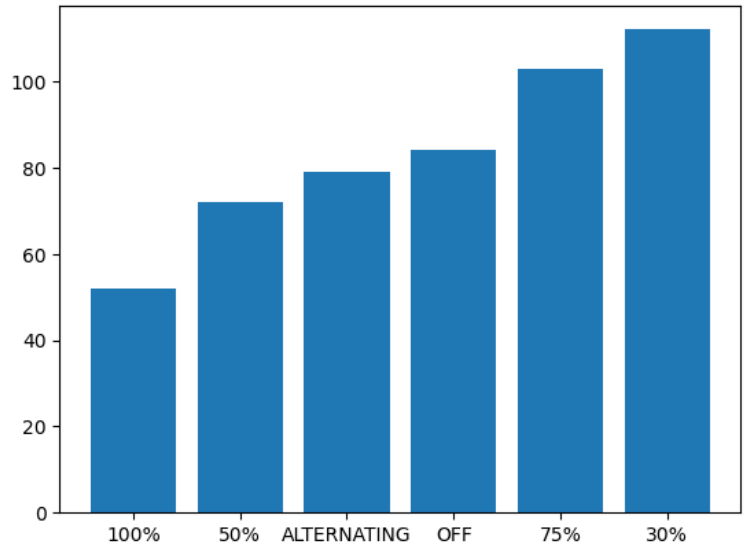



**Figure 3.9:** Ranking's sum of the points for each situation in the intensity part of the study

the case of the ranking for the color part of the study, this ranking, as well as the distribution of the answers offers insights into the participants' understanding, perspective and opinions of the non-priority situations of the study.

Participants are asked to rank the situations from the intensity part of the study based on understanding.

Figure 3.9 visually represents the overall points awarded by the participants in the user study to each presented situation, excluding the ones where priorities were assigned to the TAP rules, in the intensity part. The y-axis of this graph can be used to read the total number of points that participants gave to certain situations. Figure 3.8 shows the distribution of the answers, which displays the number of participants who chose the response corresponding to each color. The ranking numbers' allocated colors are displayed in the upper right corner.

The intensities mentioned in the TAP rules are rank leaders, while *75%* and *30%* are barely noticed by participants and placed on the last positions.

From figure 3.9 we can easily conclude that *100%* is leading the ranking, while *50%* is in the second place, followed by *ALT* as a close third. A participant notes that changing intensity to the last event to take place is quite clear, while another one found alternating to be a clear signal but uncomfortable for the eyes. *OFF* was ranked fourth among the participants, while *75%* and *30%* land the fifth and sixth positions, respectively. Participants stated that, when it comes to the light turning off, it seems like there is no conflict but at least differs from other values which makes you think, as well as that they didn't once perceive the *30%* brightness and that when it comes to *75%* and *30%*, they could not identify the intensities and the differences.

Participants believe that because *100%* is anticipated, it is simple to understand.

Looking at the color distribution for the scenario that received the highest total score, we can observe that 10 participants (41.67%) rank 100% as first, while 9 (37.5%) rank it as second. Because the result matches the user's initial expectation, some participants claim that 100% is an anticipated result and *100%* is the most understandable. It is chosen as the third position by one participant (4.17%) and as the fourth position by another (4.17%). Moreover, only one person chose the fifth position and the sixth position (8.33%), respectively. One participant says that 100% intensity could be confused with other situations and another says that it's impossible to know for sure whether one of the situations, or both, are present at any given time.

*50%* intensity also has a high ranking due to it being mentioned in the TAP rules

Only two participants (8.33%) give the situation when the light intensity changes to *50%* the top ranking, but eight (33.33%) gave it the second and third places because they believe *50%* makes sense given the TAP rules. Only one participant (4.17%) thinks the fourth or sixth position was appropriate, while four (16.67%) choose the fifth.

*ALT* is placed third in the overall ranking.

Alternating intensities comes third in the overall ranking and is chosen as the first option by three participants (12.5%), disclosing that alternating makes the occurrence of a problem clear and that it signalized the problem effectively. Other two participants (8.33%) choose to place alternating intensities on the second place, while three put it on the third (12.5%). More participants think the fourth position fits, with nine (37.5%) selecting it and supporting their decision with the statements that it is simple to get used to and a good way to indicate conflicts. Due to their perception of alternation as stressful and worries that it may induce a seizure, one participant (4.17%) and two others (8.33%) rank it in the fifth and sixth positions, respectively, among the remaining participants.

Turning the light off comes in the fourth position as some participants see it as potentially dangerous.

Another presented situation to the participants is the light turning off. Seven (29.17%) of them consider the first place fitted for the situation, affirming that it is obvious evidence that both incidents occurred at the same time and is simple to comprehend. Four (16.67%) choose the third spot, while only one (4.17%) thinks the second place is appropriate. Four (16.67%) participants rank it as their fourth option, three (12.5%) as their fifth, and six (25%) decide it is the least understandable situation and rank it in sixth place. One of the participants says they had to guess from the context that a conflict was happening, while another states that turning the light off might be dangerous.

*75%* is problematic for some participants to identify.

The situation that is placed in the fourth position in the overall ranking represents the mixing between the two intensities that are already mentioned in the given TAP rules. Only one (4.17%) participant ranks it as their first option and another one as their second, mentioning you can tell it is the average between 50% and 100%. Four (16.67%) participants out of the twenty-four see fit to place the mixing of intensities on the third position and seven on the fourth,

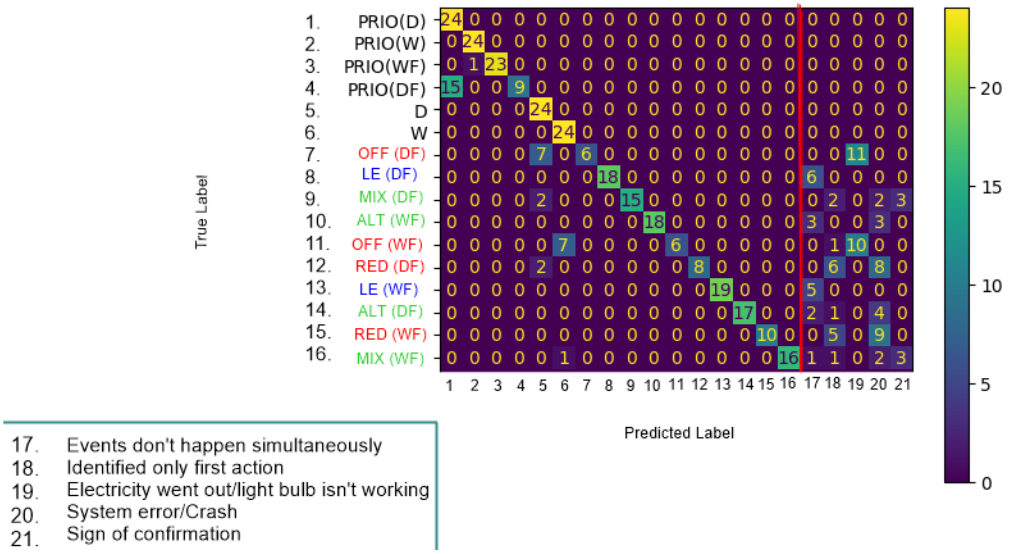



**Figure 3.10:** Confusion Matrix for the color part of the study.

while the last ranking places are chosen by six (25%) and respectively five (20.83%) participants. The ones opting for the last positions say they were not entirely sure what was going on and had a hard time distinguishing between the intensities mentioned in the given TAP rules and *75%*.

Due to participants' inability to distinguish the *30%* intensity from the others, it obtained the worst total score.

The worst overall score is received by *30%*. No participant considers this situation to be the most understandable one and only three (12.5%) place it in second place, one of them mentioning that "it makes you think". It is placed third by two participants (8.33%) and fourth by three (12.5%). The majority of participants chose either fifth place with five (20.83%) or sixth place with eight (33.33%). In fact, they even refer to it as not noticeable and said that *30%* was a random intensity that made it difficult to even detect it had happened.

### 3.7.7 Confusion matrix

Confusion matrices show how participants interpreted the presented scenarios.

In order to evaluate the performance of the classification problem, we created two confusion matrices, one for the color part of the study and one for the intensity part of the study. The two matrices can be seen in figure 3.10 and fig-

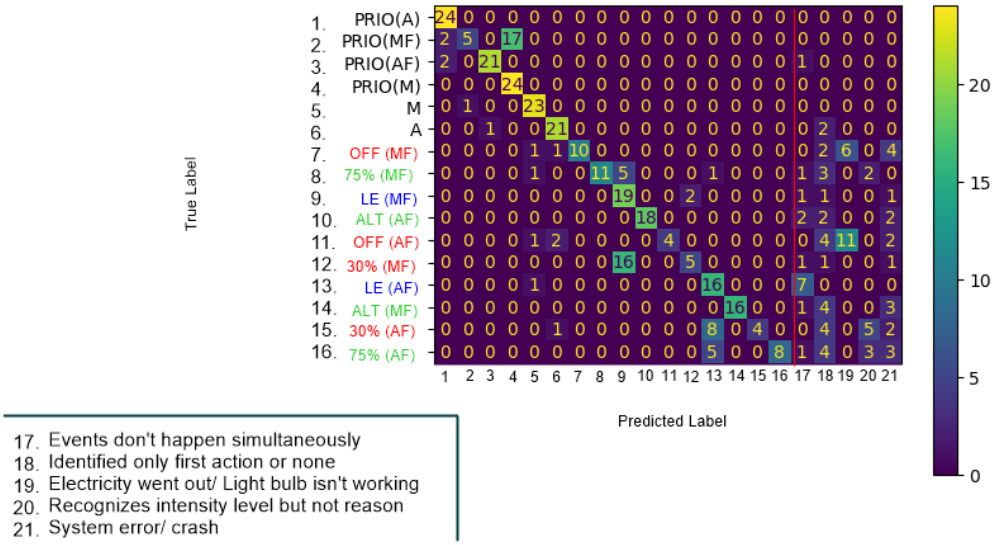



**Figure 3.11:** Confusion Matrix for the intensity part of the study.

ure 3.11. They provide an overview of how participants interpreted the presented situations throughout the study. They show how many participants actually managed to identify the presented situations correctly and provide an explanation for what they had observed during the study and how many interpreted the situations as being other scenarios.

Five columns are added to the confusion matrix to express often-made interpretations regarding the situations.

Five more columns are added to represent the interpretations that participants frequently make of the events that are presented during the user study in order to cover all possible interpretations. The added columns are delimited, in the figures, by a red line from the other, presented during the study, situations. As seen in figure 3.10, picturing the confusion matrix for the color part of the study, some of the participants have a hard time figuring out that events happen simultaneously, as they rather think presented events happen one after the other. Another problem seems to be that participants, in certain situations, only identify the first occurring event, mentioning they have no idea what just happened. A significant number of participants think that the electricity went out or the light bulb malfunctioned when confronted with the *OFF* situation. The thought of a system error or crash is also popular among the participants. Interesting to observe is that three of the partici-

pants interpreted the color green as a sign of confirmation, that everything works fine, and not as the mixing between yellow and blue.

Participants confront themselves with similar difficulties in the intensity part of the study.

For the intensity part of the study, the added columns are somewhat similar to the color part. Participants have trouble identifying whether situations occur simultaneously or not, have a hard time identifying more than one situation, sometimes even none and also seem to think the electricity went out while confronted with the situation *OFF*. As mentioned before, participants stated they struggle to identify different intensities and tell them apart. Therefore, especially for *75%* and *30%* some of the participants were able to recognize the intensity level but not the reason behind it. Lastly, participants thought that in certain situations a system crash occurs

The matrix's cells each stand for a unique combination of actual and expected events. They fall into the following categories:

- True positives (TP) represent the participants that identified the presented situation correctly.
- False positives (FP) show the participants that mistook other presented situations for the one at hand.
- False negatives (FN) express the participants that mistook the presented situation for another.
- True negatives (TN) stand for the participants that correctly identified the presented situation not to be another situation.

Performance metrics are employed to determine how well the user study participants performed.

We can quantify the performance of the participants during the course of the study by using performance metrics including accuracy, precision, recall, and F1-Score. Table 3.2 for the study's color part and table 3.3 for the study's intensity part, both provide the measurement results. The accuracy can be calculated by applying the formula: (TP + TN) / (TP + TN + FP + FN). It represents the proportion

of situations that were identified correctly out of all identified situations by the participants. Calculated as follows: TP / (TP + FP), precision represents the proportion of true positive identifications among all positive ones. In order to obtain the proportion of true positive identifications out of all positive cases, the recall has to be calculated as TP / (TP + FN). The F1-Score represents the harmonic mean of precision and recall and therefore can be obtained as follows: 2 * (precision * recall) / (precision + recall).

Despite a high accuracy rate, some situations score a fairly low precision value, indicating many false positives.

Overall, the accuracy rate is high for all of the situations presented, ranging from 92.7% to 99.7%. Generally speaking, a high accuracy rate suggests strong overall performance.
While analyzing the results for *PRIO(D)*, it is easily noticeable that despite a high accuracy rate, the precision score, being the lowest, points towards numerous participants who mistake other presented situations for the one at hand. The confusion matrix suggests that 15 participants may have mistaken the situation *PRIO(DF)* for *PRIO(D)*, failing to notice that the window was also open, due to the light bulb only displaying the color yellow.
*D* is comparable in this regard. It doesn't have a very high precision value, which signalizes a rather increased number of false positives, although the accuracy rate is high. Two participants see the mixing of colors after the door opened first as just the door opening and the other two confuse the light turning red after the door opens first, *RED(DF)*, with *D*. Seven participants think the light bulb turning off after the door opens first, was just the door opening. Just the window opening also produces a fairly low precision value of 75%. In this instance, eight participants confuse it with another situation that is provided. Seven of them believed= that *OFF(WF)* was just the window being opened, while the eighth mistake *W* for *MIX(WF)*.

Participants have a difficult time identifying the situation in which the light turns off, accordingly.

*OFF* has the lowest recall value, which is only 25%. This means that the situations in which the light turned off as an indicator of conflict is missing participants that identified it accordingly. Just six participants are able to accurately identify the situation. Additional instances with low recall values include *PRIO(DF)*, which nine participants iden-

Multiplexing and Last event to take place score good recall values.

tify correctly, *RED(DF)* and *RED(WF)*, which only eight and 10 participants identify correctly, respectively. A better score is observed in situations such as *MIX(DF)*, *MIX(WF)*, *ALT(DF)*, *ALT(WF)*, *LE(WF)* and *LE(DF)*, with more than fifteen participants being able to identify the situations correctly. Overall, the best recall scores are seen in the basic situations where only one of the events presented in the predefined TAP rules occurs.

The light turning off also has a low F1 score due to too few people identifying the situation.

The F1 score provides a balance between recall and precision, therefore a high F1 score means both, a good recall score and a good precision score. The best F1-scores can be found in *PRIO(W)* and *PRIO(WF)*. In these scenarios participants usually identify them well and also did not mistake other situations for them. Other situations with good F1-scores are *W*, *LE*, *ALT* and *MIX(WF)*. The worst F1 score, however, is attained by *OFF*. This result is due to the very few participants who managed to identify the situation.

| | ACCURACY | PRECISION | RECALL | F1 score |
|---|---|---|---|---|
| PRIO(D) | 0.961 | 0.615 | 1 | 0.762 |
| PRIO(W) | 0.997 | 0.960 | 1 | 0.98 |
| PRIO(WF) | 0.997 | 1 | 0.958 | 0.979 |
| PRIO(DF) | 0.961 | 1 | 0.375 | 0.545 |
| D | 0.971 | 0.686 | 1 | 0.814 |
| W | 0.979 | 0.750 | 1 | 0.857 |
| OFF(DF) | 0.953 | 1 | 0.25 | 0.4 |
| LE(DF) | 0.984 | 1 | 0.75 | 0.857 |
| MIX(DF) | 0.977 | 1 | 0.625 | 0.769 |
| ALT(WF) | 0.984 | 1 | 0.75 | 0.857 |
| OFF(WF) | 0.953 | 1 | 0.25 | 0.4 |
| RED(DF) | 0.958 | 1 | 0.333 | 0.5 |
| LE(WF) | 0.987 | 1 | 0.792 | 0.884 |
| ALT(DF) | 0.982 | 1 | 0.708 | 0.829 |
| RED(WF) | 0.964 | 1 | 0.417 | 0.588 |
| MIX(WF) | 0.979 | 1 | 0.667 | 0.8 |

**Table 3.2:** Results from the confusion matrix for the color part of the study, containing: accuracy, precision, recall and F1 score (rounded to 3 decimals).

For the intensity part of the study, the accuracy rate is high as well. It varies between 93% and 99%. Although situations where the accuracy rate is high but the precision rate is low can also be observed. *PRIO(M), LE* are suitable examples of the statement. For these situations, numerous participants confused other presented situations with them. The confusion matrix indicates that no less than seventeen participants could not tell *PRIO(M)* and *PRIO(MF)* apart, while twenty-one struggle with *LE(MF)*, as 5 of them mistake *75%(MF)* for it and the other sixteen *30%(MF)*. Eight participants confuse *LE(AF)* with *75%(AF)* and eight with *30%(AF)*.

Despite having a good overall accuracy rate in the intensity part of the study, some situations such as *PRIO(M)* snd *LE* score poorly when it comes to the precision rate.

The lowest recall values can be observed in *OFF(AF), 30%* and *PRIO(MF)*, being 16.7% and 20,8%. Participants find it challenging to correctly identify the aforementioned events. The situations *OFF(AF)* and *30%(AF)* are recognized only by four participants, while *30%(MF)* and *PRIO(MF)* by five. *75%* und *OFF(MF)* also show somewhat low recall values. Regarding *75% (AF)*, eight participants managed to identify the situation, ten recognized *OFF(MF)* and eleven *75% (MF)*. *PRTIO(M)* and *PRIO(A)* score best. All participants manage to identify both situations. Very good recall values have *M, A* and *Prio(AF)* as well. *M* managed to be recognized by twenty-three and the former by twenty-one participants.

Best recall values are obtained by *M, A* and *Prio(AF)*.

A good precision score and a good recall score are shown by *PRIO(A), PRIO(AF), M, A* and *ALT* through good F1 score values, varying between 80% and 92,3%. This means that participants both identify the situations well and did not mistake other situations for them. Not the same can be said about *OFF(AF)* and *30%* since they had the worst F1-scores with 28.6% and 32.3%. In these cases, participants could not identify the situations well.

The F1 score balances precision and recall. The best F1 score was achieved by *PRIO(A), PRIO(AF), M, A* and *ALT*.

| | ACCURACY | PRECISION | RECALL | F1 score |
|---|---|---|---|---|
| PRIO(A) | 0.990 | 0.857 | 1 | 0.923 |
| PRIO(MF) | 0.948 | 0.833 | 0.208 | 0.333 |
| PRIO(AF) | 0.990 | 0.955 | 0.875 | 0.913 |
| PRIO(M) | 0.956 | 0.585 | 1 | 0.738 |
| M | 0.987 | 0.852 | 0.958 | 0.902 |
| A | 0.982 | 0.840 | 0.875 | 0.857 |
| OFF(MF) | 0.964 | 1 | 0.417 | 0.588 |
| 75%(MF) | 0.966 | 1 | 0.458 | 0.629 |
| LE(MF) | 0.932 | 0.475 | 0.792 | 0.594 |
| ALT(AF) | 0.984 | 1 | 0.75 | 0.857 |
| OFF(AF) | 0.948 | 1 | 0.167 | 0.286 |
| 30%(MF) | 0.945 | 0.714 | 0.208 | 0.323 |
| LE(AF) | 0.943 | 0.533 | 0.667 | 0.593 |
| ALT(MF) | 0.979 | 1 | 0.667 | 0.8 |
| 30%(AF) | 0.948 | 1 | 0.167 | 0.286 |
| 75%(AF) | 0.958 | 1 | 0.333 | 0.5 |
| Events don't happen simultaneously | 0.961 | 0 | - | - |
| Identified only first action or none | 0.930 | 0 | - | - |
| Electricity went out/Light bulb isn't working | 0.956 | 0 | - | - |
| Recognizes intensity level but not reason | 0.974 | 0 | - | - |
| System Error/Crash | 0.953 | 0 | - | - |

**Table 3.3:** Results from the confusion matrix for the intensity part of the study, containing: accuracy, precision, recall and F1 score (rounded to 3 decimals).

# Chapter 4

# Discussion

In this chapter, the results and layout of the conducted user study are discussed. Furthermore, the quantitative and qualitative results are being analyzed in order to answer the research questions from the previous chapter.

## 4.1 False assurance caused by the familiarity of prioritizing

Participants appear to be more confident in their answers when priorities are assigned to the TAP rules.

In general, the participants in the user study appear to be more confident in their given answers in situations where priorities are assigned. This can be observed based on the modes and medians of the situations with assigned priorities, compared to the situations without. As seen in tables B.1 and B.2, that can be found in appendix B “Results of the user study”, for the color part of the user study, the most picked answer to the question ”How confident are you in the answer given above?”, for *PRIO(D), PRIO(W), PRIO(WF)* and *PRIO(DF)* is ”very confident”. The median in these situations is 1 and for *PRIO(DF)* 1,5. This also sustains the statement and shows the distribution of the answers over the Likert scale. The same can be observed in the intensity part of the study. The modes and medians of the situations where priorities are assigned to the given TAP rules overall score higher than in the other cases. The

results can be found in tables B.3 and B.4 and show that the most picked answer is ”very confident” as well and the median varies between 1 and 1.5.

Participants seem to be less surprised when priorities are involved.

Moreover, participants did not seem to be surprised by the occurring changes in the light bulb when it comes to the situations with assigned priorities in either part of the study. The most chosen answer to the question ”How surprised were you by what you saw happen as a result of the TAP rules being triggered?” is ”very unsurprised” and ”unsurprised” and the median has the values either 5 or 4.5. All this suggests that participants were, to a certain extent, expecting the colors and intensities.

Priority situations do not seem too cognitively challenging to the participants.

In addition,the cognitive load of the priority is considered to be low by participants. Some say that priorities make it easier to identify events and are very understandable. The most chosen answer to describe the cognitive load of the priority situation is ”very cognitively unchallenging”. Participants seem to not have to make a big cognitive effort to identify occurring events based on events that happen in the room when priorities are involved.

Users show a strong need for assurance and control over situations.

According to the user study’s findings, it seems that participants feel more assured when priorities are given to the presented TAP rules, stating that priorities simplify things and make clear how conflicts and the system work, as well as the possible options. Regarding the color part of the study, one participant mentions, that priorities limit the color output to only two, which is easier to understand. For the intensity part of the study, another participant claimed that they are beneficial because manual intervention always prevails in conflicts, keeping the situation under control. These statements made by participants show the user’s need and desire for situational control and also for knowing what will happen next.

The confusion matrices in figure 3.10 and figure 3.11 show that the feeling of assurance and comfort one might have as a result of knowing the rules have assigned priorities and thinking the occurrence of a situation is clear might be indeed just a mere feeling. In fact, the study subjects are still unable to name every occurrence that took place in

the room during the study. Some of them are quick to realize that priorities might have some shortcomings and can still make the user miss out on some events if they happen simultaneously. Participants observe that while priorities can be valuable in some situations, they can also cause one to miss some occurrences. One even voices concern about how priorities often fail to accurately depict conflicts. To further explain this statement, the following examples are mentioned based on the provided TAP rules and presented situations during the study for both, the color and intensity part: A participant says that due to the rule regarding the door having priority, it is impossible to tell if the door and window are open at the same time in case the door opens first because the light color remains yellow, while another participant finds the combination between dimming lights and priorities to be confusing. Another issue brought up by a participant is the difficulty in recalling the priority that was set.

The confusion matrices show that participants have trouble distinguishing between certain situations where priorities are assigned.

Besides, the confusion matrix reveals that only a small percentage of participants can genuinely recognize that another event might be occurring simultaneously if the event with the greater priority occurred first. The nine participants that consider the possibility of the window opening while the door is still open and the five participants that consider the motion sensor getting triggered after the switch has been pressed, still can't tell for sure if only one of the events is happening or both events are occurring at the same time: "Door opens. The window might be open as well but we can't tell for sure." This shows that, despite the participants' high confidence, low level of surprise, as well as the cognitive load considered to be low, the majority still can't identify the conflict. Furthermore, it can be deduced that the participants seem to get a feeling of familiarity and assurance from prioritizing since it's a widely used method, making them lean towards it and giving it a better evaluation.

Only a few participants recognize that it's hard to tell how many events occurred simultaneously if the one with the higher priority happens first.

Considering this information, it becomes clear that while assigning priority in smart homes to prevent conflicts is a very helpful and familiar alternative for the user, it can also be deceptive and make it difficult to recognize occurring events in some circumstances. As a result, there may be a

Priorities might not be enough to inform the user of all occurring events.

requirement for an alternative that can inform the user that several rules are being triggered at once.

## 4.2 User's understanding

Less cognitively challenging situations are easier for participants to identify.

Looking at the results reveals that predefined events are overall better perceived than unexpected ones, as well as simpler sequences of events, are easier to understand than more complex ones. What participants in the user study considered to be the most understandable when it comes to the situations where priorities are not assigned is depicted in figure 3.7 and figure 3.9. They almost always rank the expected outcomes such as the light bulb turning yellow or blue and the light intensity being at 50% or 100% as their first preferences, mentioning that yellow and blue are the most understandable since these were the colors set in the initial TAP rules. 50% and 100% are easy to understand and expected by the user as well. The confusion matrices, in figures 3.10 and 3.11 also sustain this statement. All participants are able to identify the situations accordingly where only one event occurred in the color part of the study, regardless of whether priorities were assigned or not, and none of the participants had trouble identifying single events when priorities were assigned in the intensity part of the study. The majority of them also recognized correctly the situations where only one event occurred and priorities were not assigned. On the other hand, some participants argued that 50% or 100% make it seem like there is no conflict and nothing is out of the ordinary and that displaying the same colors as the one mentioned in the given TAP rules is disorienting.

Some participants accurately identified certain situations because exit rules are not taken into account,

It is important to note that while participants rank yellow, blue, 50%, and 100% as the most intuitive ways to identify a conflict, according to the rankings, very few of them clearly comprehended the concept underlying these light bulb states as conflict indicators. Only two participants mention the fact that the light bulb keeps the state in which it is brought by the last event to take place. Due to the fact that exit rules are not taken into consideration in this study, the other participants accidentally properly identify

the events that take place in the room as a result of the TAP rules being triggered. Therefore participants stating that the door is opened and then the window is opened, or that the motion sensor is triggered and then the light switch is pressed would identify the situation correctly, although it is not directly mentioned by the participants that these events happen simultaneously. Overall the Last Event to take place class has better recall values than most situations.

albeit it is unclear whether they comprehend that events were occurring simultaneously.

On the other hand, a class participants appear to experience clear difficulties with is the error class. Overall, an error state such as the light bulb turning off or red or displaying a random brightness level seem to confuse the participants and make them rather assume a system failure, the electricity being cut off, or just leave them confused without an idea what had just happened, rather than make them realize that two rules were triggered at the same time. This becomes clear while looking at the confusion matrices (3.10, 3.11) as just a small percentage identifies the situations accordingly.

The error class as conflict indicator is barely identified accordingly and creates the most difficulties for the participants.

Based on the confusion matrices as well(3.10, 3.11), it can be deduced, that there is a high percentage of participants (45.83%) that assume the electricity went out if the light bulb turning off is used as an indicator of conflicts. It seems that an average smart home user would just go with the first instinct in situations where an unexpected event occurs and not question whether that is the correct reason for the event or if there could be another. A participant in the user study states that they can not find any explanation for the light turning off, while another simply says the light should never be turned off. There also were some participants that thought the light turning off is a good conflict indicator, saying that it makes sense based on the given TAP rules since the light turning off was not mentioned in them. The recall values for *OFF* are very low compared to the other values in both cases and sustain the lack of participants that identify *OFF* accordingly. For the color part of the study, the value is 0.25, while for the intensity part is a bit higher, varying between 0.167 and 0.417.

If lights turning off is used as conflict resolution, there is a high chance, the user will assume the electricity was cut off.

Using the color red to indicate conflicts makes users feel alarmed and confused.

Red is also part of the error class and it is able to make only 37.5% of the participants recognize that two rules are triggered at the same time, while further 37.5% assume an error or a system crash occurred. The color red as an indicator of conflicts has a low recall value, as well. Participants state that they did not fully get the fact that both, the window and the door, are open when red is used as an indicator and that red is the most confusing and unexpected outcome of the presented ones. Some participants find *RED* to be useful and mentioned, red is a commonly used color for problems so it's the best indicator for a conflict. From the participant's reactions to the light bulb turning red, it's obvious that an alarming setting is created, especially since the color red is not expected or mentioned in any scenario. Since the color red might bring the user into an alarming state, the outcome might be the user assuming a worse situation such as system failure or crash.

30% is hard to identify and often misleading, as it can be easily confused with other intensities.

However, the worst understanding seems to be achieved by 30%. It is meant as a conflict indicator through random, unspecified intensity level and it is part of the error class as well. Not many participants rate it very well since it is also the only situation that is not ranked as the most understandable by any of the twenty-four participants in the ranking (3.9). These results probably have a strong correlation with the fact that most participants did not perceive the 30% intensity, stating they often mistook it for 50%. Only 20,8% identify the situation accordingly. This can be seen in the confusion matrix as well as the table containing the results from the confusion matrix. The recall value for this situation is the lowest out of all of them, therefore many positive cases are missing. It can be concluded, that a random, unexpected intensity probably doesn't help users identify a conflict and is rather misleading. This can also be a result of the intensities not being a practical indicator of events overall, as seen in the figure 3.5.

How good users understand multiplexing might depend on the given TAP rules as well.

Finally, taking a look at multiplexing, during the user study opinions were divided. The final scores are generally good, except for the mixing situation in the intensity part of the study, where the recall value is considerably lower than for the other multiplexing situations. This is due to the fact that mixing did not seem to be very well perceived by the

participants in the intensity part of the study since they can not tell 50%, 75% and 100% apart and mistake one for another. Alternating scores better than mixing, having better recall and F1 scores. To point out is that usually more than 50% of the participants are able to identify a conflict when expressed through multiplexing, as deduced from the confusion matrices (3.10,3.11) . When it comes to color mixing, more participants are capable to figure out that green represents the mixture between blue and yellow. One of the participants states that alternating and mixing are both good methods to make the user aware of a conflict occurrence generally, but 75% was not the best since intensities are in general confusing.

Familiarity and less cognitive load are preferred and thus, the first research question is answered.

All things considered, the first research question can be answered as follows: It is difficult to say what brings the user more understanding since multiple factors have to be considered. It seems like generally speaking, people feel more secure when dealing with expected outcomes or outcomes that are directly related to the already defined rules. The best method, according to one participant, relies on the user's knowledge and understanding since alternating would be an excellent choice if the user knows it is an indicator of conflicts, while another believes it would get easier to spot occurrences and conflicts over time as the user grows accustomed to the system. Furthermore, based on the results, we can say that generally, the multiplexing class brings more understanding to the user than the error class, as the error class put a significant number of participants in difficulty.

## 4.3 Perception of conflict resolution techniques

Participants had a false perception of the last event to take place class.

As mentioned before, the Last Event to Take place class scores very well in the overall ranking, being chosen as the first option by many participants. But, as explained before, this choice might have been biased, as participants did not fully get the idea behind this class and just thought yellow, blue, 50% and 100% are more understandable because they

are defined in the TAP rules. Participants mention these are expected outcomes and most understandable since they are the colors and intensities initially set. A participant mentions they think a third color is a better signal for a conflict than using yellow or blue. A second participant points out that in the case of 50% and 100%, they could be confused with other situations.

Participants view the error class with the worst perception.

Nevertheless,participants seem to have a less favorable perception on the error class. They state that the red outcome is the most unexpected and ambiguous one and that the light turning off implies a problem but the cause is unclear. People express their confusion and say they have no idea what just happened in the room. In the intensity part of the study, 30% is considered to be "random" and hard to identify. Most participants even admit they are not aware they have experienced this level of intensity and can not name it or explain why it is presented. Based on the gathered data and participants' opinions, it looks like out of all the presented non-priority situations, the error class was the one that created the participants the most difficulties and was least understandable.

Multiplexing is seen as a good conflict indicator by several participants.

Multiplexing seems to intrigue a few participants, as several state they would opt for this conflict resolution method in their homes and a participant says they never thought of this option when being confronted with a conflict but thinks it can be effective in the long run. Additionally, alternating, according to participants, clarifies a problem and makes it simple to pinpoint the issue. By another participant, on the other hand, alternating is regarded as an effective but very unpleasant signal, especially for the eyes. When it comes to mixing, one participant believes that mixing colors, which represent two active states, is the most suggestive option.

The multiplexing class is better perceived than the error class. It is difficult to determine whether it is better perceived than the last event to take place class.

In order to answer the second question of the research: based on the collected data, it seems that participants perceive the multiplexing class better than the error class and it is easier for them to conclude that a conflict was occurring while being presented with mixing or alternating as indicators. On the other hand, it is hard to make a statement when it comes to the last event to take place class due

to the misinterpretation by the participants as well as the uncertainty of the correctness of their answers.

### 4.3.1 Multiplexing as suitable conflict resolution

Opinions regarding the multiplexing's effectiveness and utility are divided.

Several participants believe that multiplexing is a reliable and intelligible sign of conflicts. In terms of overall rankings for the color part of the study, alternating is ranked third, mixing is ranked fourth, and for the intensity part, alternating is ranked third and mixing is ranked fifth (3.7; 3.9). While, as revealed, some participants see multiplexing as a good conflict indicator, some others voice doubts regarding the utility of multiplexing. They point out that the effectiveness of mixing depends on the chosen colors because certain colors mix together better than others and that the utility of alternating may decrease as more rules are added. On the other hand, another participant says that they find assigning priorities helpful but would choose a third color to display a conflict. People struggle to decide which multiplexing style they preferred. This can be deduced based on participants' statements such as: "The alternation is a bit stressful, so I think a middle ground such as 75% would be a better solution."

About three-quarters of the participants are able to identify conflicts based on multiplexing.

As previously stated, despite certain concerns participants have, multiplexing has a satisfactory overall score in the rankings of the non-priority conflict resolution methods. Participants state that mixing and alternating make the occurrence of a conflict clear. The confusion matrix for the color part of the study, depicted in figure 3.10 shows that between 62.5% and 75% of participants are able to identify that a conflict happens while using multiplexing as an indicator. The confusion matrix for the intensity part, shown in figure 3.11 shows a bigger difference between mixing and alternating. Only between 33.33% and 45.83% of participants identify a conflict when the used method to indicate it is mixing and between 66.67% and 75% when the used method is alternating.

Even though more than half of the participants identify the occurring conflict accordingly through multiplexing,

Although they appear to be confident in their responses, participants did find multiplexing to be somewhat cognitively demanding.

it still seems like it was mentally demanding to a certain extent. After seeing the situations where multiplexing is present, participants were asked how confident they are in their provided answer regarding what happens in the room. "Confident" and "somewhat confident" are chosen more frequently than the other options on the Likert scale. The results are slightly different for the intensity part of the study since the most popular responses are "extremely confident" and "confident". Moreover, multiplexing is viewed as "cognitively challenging" and "slightly cognitively challenging" in general. This shows that although participants are relatively confident in their capability of identifying what events occur in the room, they still find it somewhat difficult to orient themselves based on the light bulb's colors and intensities when multiplexing is used.

H2 has to be ruled as uncertain because the made statement is about multiplexing and not alternating alone.

Another point that needs to be addressed is the uncertainty of H2, as it is of significance for answering the third research question. As mentioned in 3.7 "Results", after conducting the Wilcoxon signed ranked test, H2, stating that multiplexing makes the occurrence of conflicts less surprising for users than other non-priority conflict resolution methods, where conflicts occur, is ruled as uncertain. This is due to the fact that, for the second question that participants have to answer, significant differences are only found between mixing and *LE*, as well as between alternating and the *OFF*; alternating and *RED* and alternating and *LE*, in the color part of the study. As for the intensity part, significant differences are identified between mixing and *LE* and alternating and *LE*. The results show that only alternating in the color part of the study makes the occurrence of conflicts overall less surprising to users. Since multiplexing is composed of mixing and alternating in this thesis, H2 can not be rejected but also can not be proved since the initial statement is regarding both parts of multiplexing and not just alternating.

The user's prior knowledge, expectations, and interactions with multiplexing can all play a role in the user's perspective of it.

At the same time, multiplexing can also be very dependent on the user's prior knowledge and expectations. According to the confusion matrix for the color part of the study, three participants interpret mixing between yellow and blue, which results in green, as a sign of confirmation that the event showed prior was executed successfully. Be-

cause of this, in the case in which the window opens first, some participants believed that the following green signal means that the door had opened successfully and that there was no conflict, rather than that the window has also opened and the two colors had mixed: "Door got opened and as confirmation that it worked, the light turned green." It is mentioned by participants that "the human brain associates the color green with confirmation, a sign that everything is working fine". A problem participants have as well is that in the case of alternating, some of them did not realize that events are happening simultaneously. Therefore, participants come up with explanations for why the light color and intensity are alternating, such as the door and window being opened and closed multiple times or for the intensity part, someone playing with the light switch while still getting detected by the sensor. This shows that the user's understanding and perspective of multiplexing are influenced by previous cognitive associations and it might require getting used to and forming new associations.

Multiplexing awakened some of the participants' interest.

A participant makes a remark that further sustains the idea that the user's prior knowledge and interactions with multiplexing are significant: "I would have loved it (referring to a new color as conflict indicator) in my home if an unexpected color did not appear". Another participant shows interest in multiplexing and says they would try mixing intensities as an outcome for rules being triggered simultaneously but a problem is that they did not recognize the 75% light intensity.

Colors are a better event indicator than light intensities.

Furthermore, the user's prior knowledge, expectations, and interactions with multiplexing don't seem to be the only factors that play a role in the user's perspective of it. It seems that multiplexing is also dependent on the setting in which it occurs and the given TAP rules. Figure 3.5 in 3.7 "Results" shows an overall score of how practical participants find light color and light intensity as indicators of events. The participants use a Likert scale with values from "very useful" to "very unuseful", to express their opinions. Participants consider the change of color in the light bulb to be a better indicator of events than the change in intensity. This comes as a result of participants' difficulty in differentiating between intensities. It is mentioned multiple

times, that it is hard to differentiate between brightness levels and that in cases like the presented one, priorities might be more helpful. One of the participants provided a more extensive answer by explaining, that it is difficult to differentiate between intensities and more so to understand the shown percentages on the first run as one is not used to them. The participant assumed it might become easier the more you see such a system. This can impact how suitable multiplexing is. The mixing between colors seems to be better perceived than the mixing between intensities since the participants could not recognize *75%*. This can also be seen based on the rankings (3.7, 3.9), as multiplexing is placed on third and fourth place in the color part of the study, but on third and fifth place in the intensity part.

If multiplexing is a suitable solution for smart home conflicts depends on the user's preferences.

So whether multiplexing is a suitable solution for conflicts or not depends on each user's preferences, understanding and perception of it. The data shows a considerable amount of participants are able to recognize an occurring conflict when indicated by multiplexing but some participants misinterpret mixing and alternating. Some participants question how effective multiplexing would be as an indicator of conflicts, while others pointed out that priorities can not always represent the occurrence of a conflict accordingly. All things considered, it seems that multiplexing would be a possible suitable indicator of conflicts if users would get used to it and its meaning as they would start expecting certain events to happen. The answer to this research question is also highly dependent on the provided rules. As deduced from the answers of participants in the user study, some conflicts are more appropriate to express through multiplexing than others, based on how clear the mixing is displayed and how irritating alternating is.

# Chapter 5

# Summary and future work

This thesis and the value of its findings are summarized in the first section of the last chapter. Ideas for further work are covered in the second section of the chapter. Ideas to assist resolve conflicts in smart homes are provided in order to simplify people's daily lives and make the usage of smart home technology more accessible and clear.

## 5.1 Summary and contributions

This thesis aims to look into the effectiveness and user's understanding of multiplexing with the help of a user study.

The aim of this thesis is to propose action multiplexing as a potential conflict resolution method for smart homes, to give an overview of how it may be incorporated into users' daily lives, and to look into people's understanding and perception of it. In order to achieve this, a user study was conducted, in which six possible outcomes, as a result of a conflict, were presented to the participants through action mixing, alternating, state based on the last event to take place, assigning priorities and off switch, red and random action as indicators of error. Furthermore, the study was composed of two parts, one using light color as an event indicator and the other one using light intensity as an event indicator. In the study, the participants have to

identify events happening in the room based on the outcome of TAP rules getting triggered. Through this work we set out to answer three questions: 1. What brings more understanding to the user, action multiplexing or other smart home non-priority conflict resolution methods?; 2. What is better perceived by the user, action multiplexing or other smart home non-priority conflict resolution methods?; 3. Is action multiplexing a suitable solution for smart home conflicts?

Although users tend towards less cognitively challenging scenarios, it does not mean that they are always able to identify conflicts in such situations.

A ranking shows that participants prefer already-known and somewhat expected outcomes, using actions that were mentioned in the TAP rules before because of predicted reasons: They do not demand much cognitive workload, and with them being defined in the TAP rules, the user gets a sense of control over the situation. Although, as mentioned in chapter 4, despite the participants placing the situations where the state of the last event to take place is kept high in the ranking, it does not mean that they were in fact able to identify the occurrence of the conflict as well. So, the first question can not be fully answered since the participants got a false sense of understanding and control from the color and intensities that were already mentioned through the TAP rules. But it is clear that users prefer familiar and cognitively unchallenging events and outcomes and therefore, it is best if the user expects and has already seen a certain behavior from the device.

Users might prefer the already familiar prioritizing. Multiplexing's usefulness is dependent on certain factors.

Participants stated during the study that although action multiplexing can signalize the occurrence of a conflict well, it is very dependent on the rules at the base. They hinted that the number of rules conflicting, as well as the actions themselves play a role in the multiplexing's efficiency and usefulness. Prioritizing is and will probably continue to be a user's preferred method to combat conflicts due to usually good results and widespread. Overall, multiplexing was better perceived than the error class according to the gathered data. To answer the second question, multiplexing seems to be better perceived than the error class, but users are already familiar with techniques such as assigning priorities, therefore might be inclined toward those, as well as the last event to take place class due to simplicity.

The study data shows that a significant number of participants were able to identify the occurrence of a conflict accordingly when the used method to display it was action mixing or alternation. Although the number was slightly higher in the color part of the study than in the intensity part. This is due to the light colors being perceived as a better event indicator than the light intensities, which were harder for participants to distinguish from one another. Furthermore, user preferences and opinions also have to be taken into account, since the feeling of comfort and ease of the user plays an important role in smart homes. The answer to the third question based on the gathered data is that action multiplexing could be a suitable solution for smart home conflicts but it is highly dependent on the user's liking and perception of it.

Multiplexing could be a suitable solution for smart home conflicts depending on the user's preferences.

In summary, the contribution of this work is analyzing and discussing users' perception and understanding of action multiplexing as an indicator and solution to smart home conflicts.

## 5.2 Limitations

This thesis proposes an approach for conflict signalizing and resolution in smart homes. Although the findings presented in chapter 3.7 have shown that action multiplexing has the potential to be effective as a conflict resolution method for smart homes, there are still a number of issues that need to be considered.

First, although smart homes are widely adopted, not all participants in the user study were actual smart home users in their everyday life.
Second, one of the participants was red-green color blind and therefore had a harder time determining the outcome of some of the presented situations in the color part of the study.
Furthermore, the study took place in a designated quiet room, where the participants had to observe a series of events. They were actively waiting for an event to occur and had all their attention shifted toward the light bulb

Multiplexing has certain limitations that have to be considered.

used in the study. In everyday life, rules can get triggered at any given moment and therefore, the user does not pay full attention to their surroundings, waiting for an event to take place. Consequently, there is always a chance that the user might miss events. To mention is that during the user study if a participant needed to see a situation again, it was possible, while in everyday use of smart home devices, usually, that is not the case.
For the user study, in both parts of it, only two TAP rules were provided. As participants have mentioned, in case multiple rules get triggered simultaneously, multiplexing might not be as convenient for the user anymore. For instance, regarding action mixing, since in the user study only two colors were used, they mix well together. The results might be slightly different if other colors such as red and orange would have been chosen. As for alternating, if multiple rules are involved, it can become tiring and very cognitively challenging for the user to understand which rules are triggered.
Another point to be made is the fact that the user study only had two parts, one providing rules regarding the light color and another one providing rules regarding the light intensity. Consequently, action multiplexing was tested only for this specific set of TAP rules, acting on a light bulb, and conflicts. In the everyday life of the smart home user, a lot of rules, acting on various devices, can come into conflict and depending on the device and the performed action, multiplexing might be more or less useful.
Finally, participants only had the chance to come in contact with multiplexing a very limited number of times and they even stated that it might be more useful if they could get used to it or expect it to happen. To take into account is also the fact that in the user study the TAP rules were already given, while usually smart home users set the TAP rules themselves, which increases the level of understanding. As a whole, multiplexing has some limitations, that should be considered and researched in future work.

## 5.3 Future work

Ideas for further research on the applicability and usefulness of action multiplexing emerged during the process. A possible way to move forward would be to further research the applicability and usefulness of multiplexing in different settings and using different devices as targets, as well as various numbers of rules that might end up in conflict. Letting the user get acquainted with the idea of multiplexing in a more familiar environment while also having the option to set rules they find useful for themselves might also make a difference in the results.

Further research is needed to explore the applicability and usefulness of action multiplexing in different settings and with different devices.

# Appendix A

# USER STUDY

Important forms for carrying out the user study are shown on the pages that follow. These forms were filled out by the participants during the study.

## Informed Consent Form

**Can Smart Home Action Multiplexing Solve Priority Conflicts?**

**Principal investigator:** Ilinca Baicu
Email: ilinca.baicu@rwth-aachen.de

**Purpose:** The goal of this survey is to evaluate solutions suitable for conflict resolution in smart homes through a series of questions.

**Procedure:** Participation in this survey involves observing a series of events and situations created with the help of smart gadgets or through simulation and answering a number of questions based on these situations and simulations. The investigator will present certain events to the participant and a short interview will follow.

Questions asked and information received throughout the interview process will be logged. All information will be confidential. (See 'Confidentiality' below for details.)

**Risks/Discomfort:** The survey is expected to last no longer than 60-80 minutes. If you become fatigued during the course of your participation in the survey feel free to take as many breaks as necessary during the allotted timeframe. Some of the presented situations contain slow flashing light that might represent a risk to certain participants. Should the completion of the task become distressing to you, it will be terminated immediately.

**Confidentiality:** All information collected during the study will be kept strictly confidential. Results will be aggregated, and more specific information will be pseudonymized. After evaluation we will delete all recordings of the study session. If you agree to participate in this survey, please sign your name below.

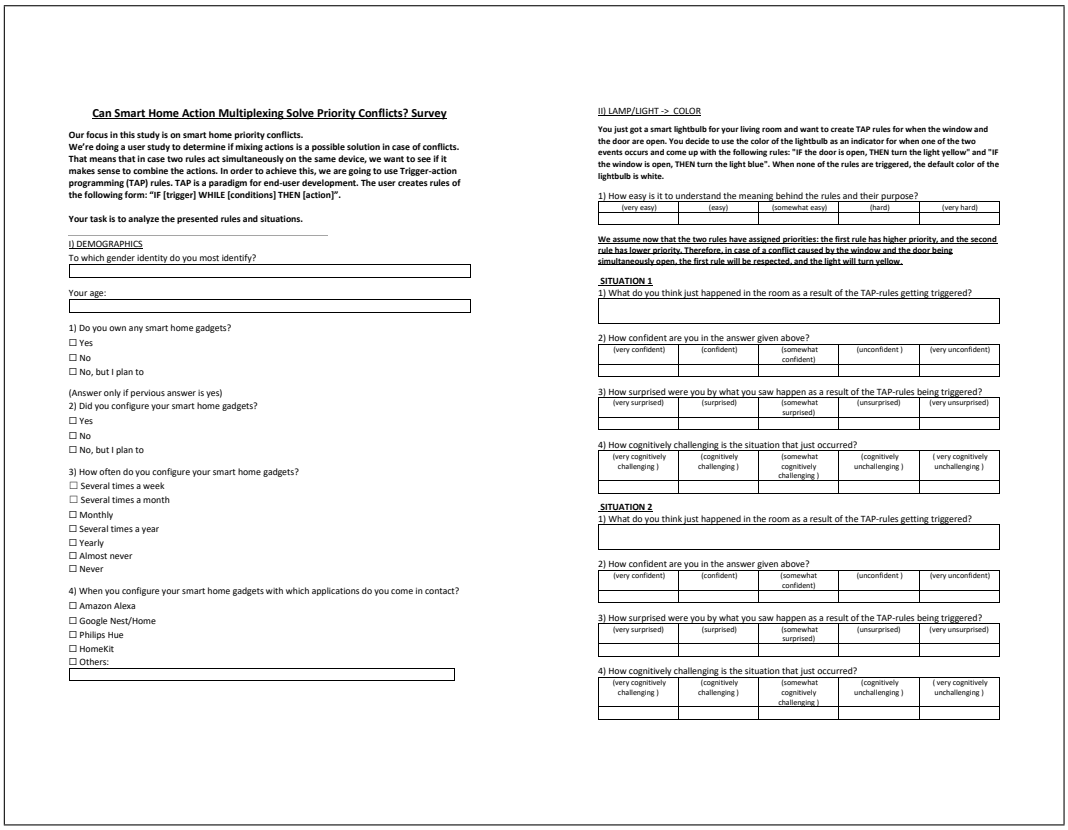
**Addendums:** Participation in this study is voluntary. You are free to withdraw or discontinue the participation. Participation in this study will involve no cost to you and you will be given compensation in form of sweets and drinks.

☐ I have read and understood the information on this form.
☐ I have had the information on this form explained to me.
☐ I grant permission to the researcher to audio tape me as part of this research.

| Participant's Name | Participant's Signature | Date |
|---|---|---|

| Principal Investigator | Date |
|---|---|

**Figure A.1:** Informed consent form of the user study



**Can Smart Home Action Multiplexing Solve Priority Conflicts? Survey**

**Our focus in this study is on smart home priority conflicts.**
**We're doing a user study to determine if mixing actions is a possible solution in case of conflicts. That means that in case two rules act simultaneously on the same device, we want to see if it makes sense to combine the actions. In order to achieve this, we are going to use Trigger-action programming (TAP) rules. TAP is a paradigm for end-user development. The user creates rules of the following form: "IF [trigger] WHILE [conditions] THEN [action]".**

**Your task is to analyze the presented rules and situations.**

I) DEMOGRAPHICS

To which gender identity do you most identify?

Your age:

1) Do you own any smart home gadgets?
☐ Yes
☐ No
☐ No, but I plan to

(Answer only if pervious answer is yes)
2) Did you configure your smart home gadgets?
☐ Yes
☐ No
☐ No, but I plan to

3) How often do you configure your smart home gadgets?
☐ Several times a week
☐ Several times a month
☐ Monthly
☐ Several times a year
☐ Yearly
☐ Almost never
☐ Never

4) When you configure your smart home gadgets with which applications do you come in contact?
☐ Amazon Alexa
☐ Google Nest/Home
☐ Philips Hue
☐ HomeKit
☐ Others:

II) LAMP/LIGHT -> COLOR

**You just got a smart lightbulb for your living room and want to create TAP rules for when the window and the door are open. You decide to use the color of the lightbulb as an indicator for when one of the two events occurs and come up with the following rules: "IF the door is open, THEN turn the light yellow" and "IF the window is open, THEN turn the light blue". When none of the rules are triggered, the default color of the lightbulb is white.**

1) How easy is it to understand the meaning behind the rules and their purpose?

| (very easy) | (easy) | (somewhat easy) | (hard) | (very hard) |
|---|---|---|---|---|
| | | | | |

**We assume now that the two rules have assigned priorities: the first rule has higher priority, and the second rule has lower priority. Therefore, in case of a conflict caused by the window and the door being simultaneously open, the first rule will be respected, and the light will turn yellow.**

**SITUATION 1**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**SITUATION 2**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

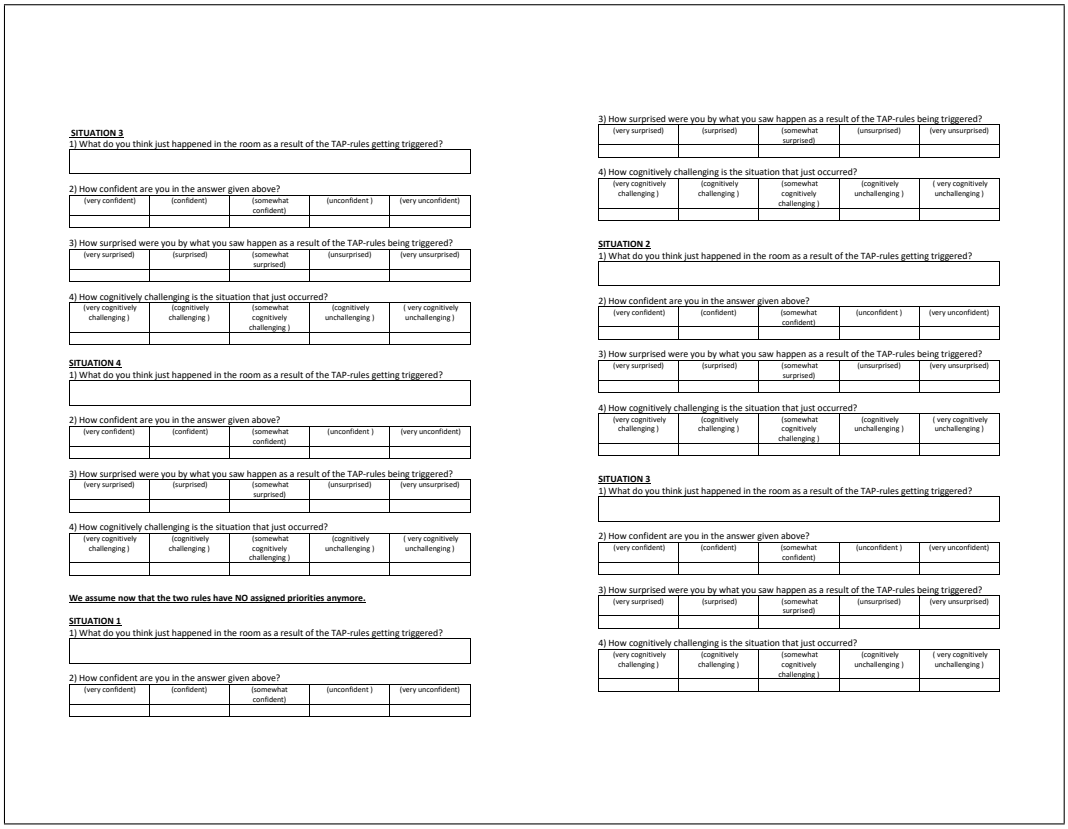
3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**Figure A.2:** Survey of the user study (1)



**SITUATION 3**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**SITUATION 4**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**We assume now that the two rules have NO assigned priorities anymore.**

**SITUATION 1**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**SITUATION 2**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**SITUATION 3**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

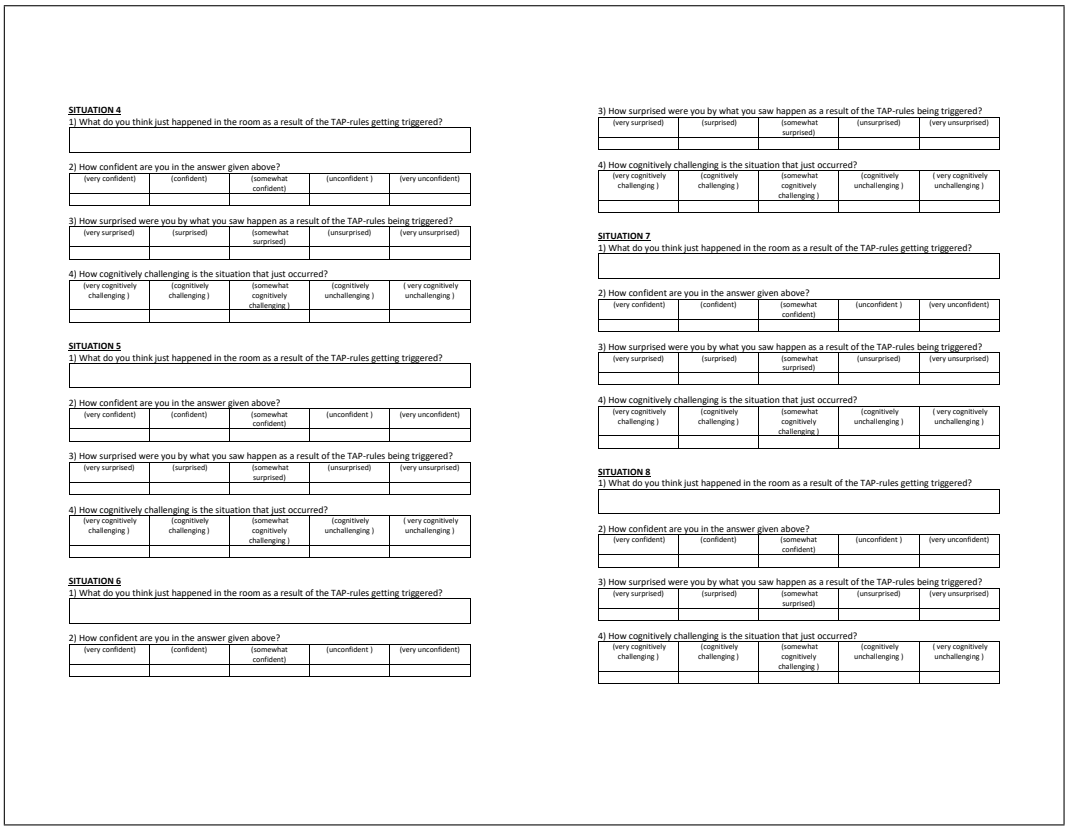
3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**Figure A.3:** Survey of the user study (2)



**SITUATION 4**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**SITUATION 5**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**SITUATION 6**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**SITUATION 7**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**SITUATION 8**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

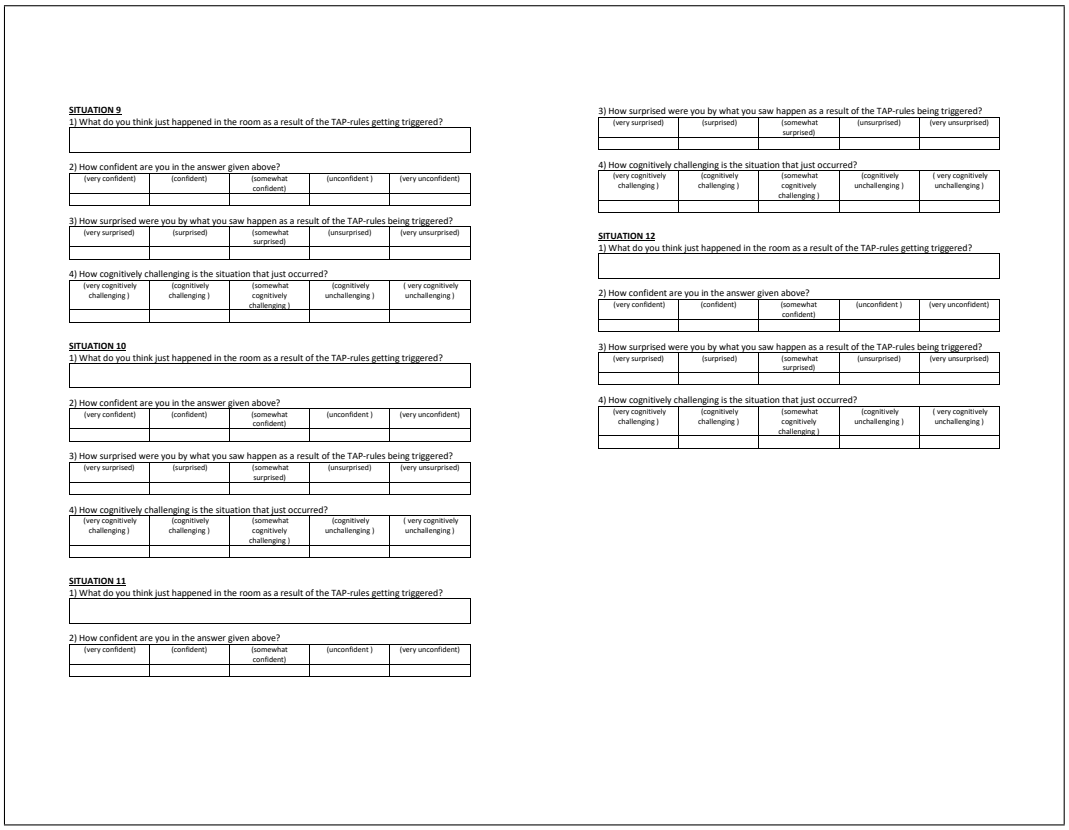
3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**Figure A.4:** Survey of the user study (3)



**SITUATION 9**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

| |
|---|
| |

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**SITUATION 10**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

| |
|---|
| |

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**SITUATION 11**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

| |
|---|
| |

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**SITUATION 12**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

| |
|---|
| |

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

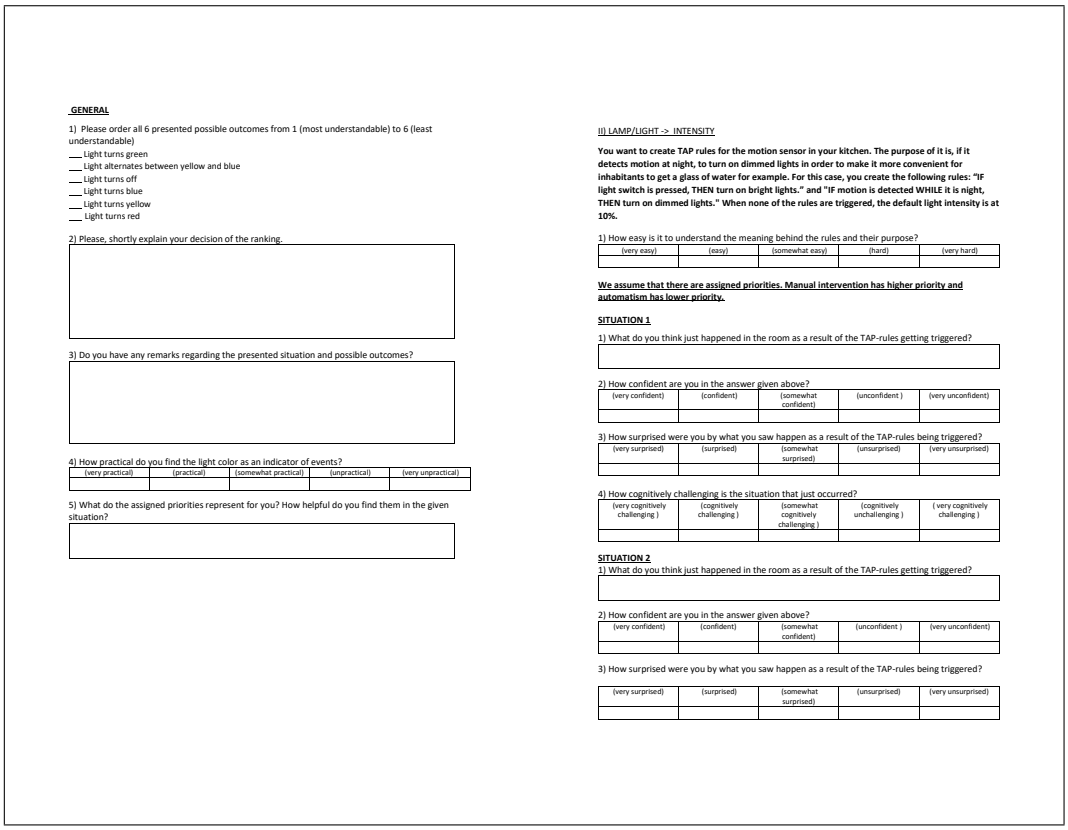
3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**Figure A.5:** Survey of the user study (4)



**GENERAL**

1) Please order all 6 presented possible outcomes from 1 (most understandable) to 6 (least understandable)

___ Light turns green
___ Light alternates between yellow and blue
___ Light turns off
___ Light turns blue
___ Light turns yellow
___ Light turns red

2) Please, shortly explain your decision of the ranking.

3) Do you have any remarks regarding the presented situation and possible outcomes?

4) How practical do you find the light color as an indicator of events?

| (very practical) | (practical) | (somewhat practical) | (unpractical) | (very unpractical) |
|---|---|---|---|---|
| | | | | |

5) What do the assigned priorities represent for you? How helpful do you find them in the given situation?

II) LAMP/LIGHT -> INTENSITY

**You want to create TAP rules for the motion sensor in your kitchen. The purpose of it is, if it detects motion at night, to turn on dimmed lights in order to make it more convenient for inhabitants to get a glass of water for example. For this case, you create the following rules: "IF light switch is pressed, THEN turn on bright lights." and "IF motion is detected WHILE it is night, THEN turn on dimmed lights." When none of the rules are triggered, the default light intensity is at 10%.**

1) How easy is it to understand the meaning behind the rules and their purpose?

| (very easy) | (easy) | (somewhat easy) | (hard) | (very hard) |
|---|---|---|---|---|
| | | | | |

**We assume that there are assigned priorities. Manual intervention has higher priority and automatism has lower priority.**

**SITUATION 1**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively challenging ) |
|---|---|---|---|---|
| | | | | |

**SITUATION 2**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

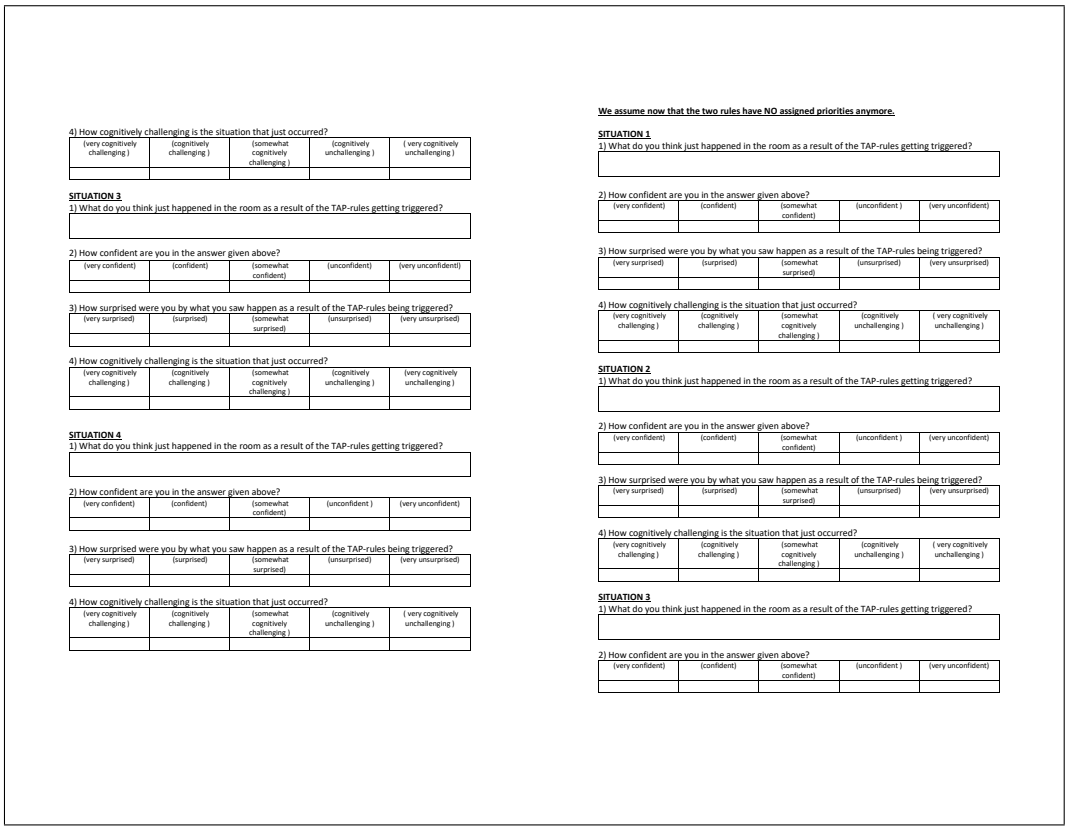
2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

**Figure A.6:** Survey of the user study (5)



4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**SITUATION 3**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident) | (very unconfident!) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | (very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**SITUATION 4**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**We assume now that the two rules have NO assigned priorities anymore.**

**SITUATION 1**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**SITUATION 2**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

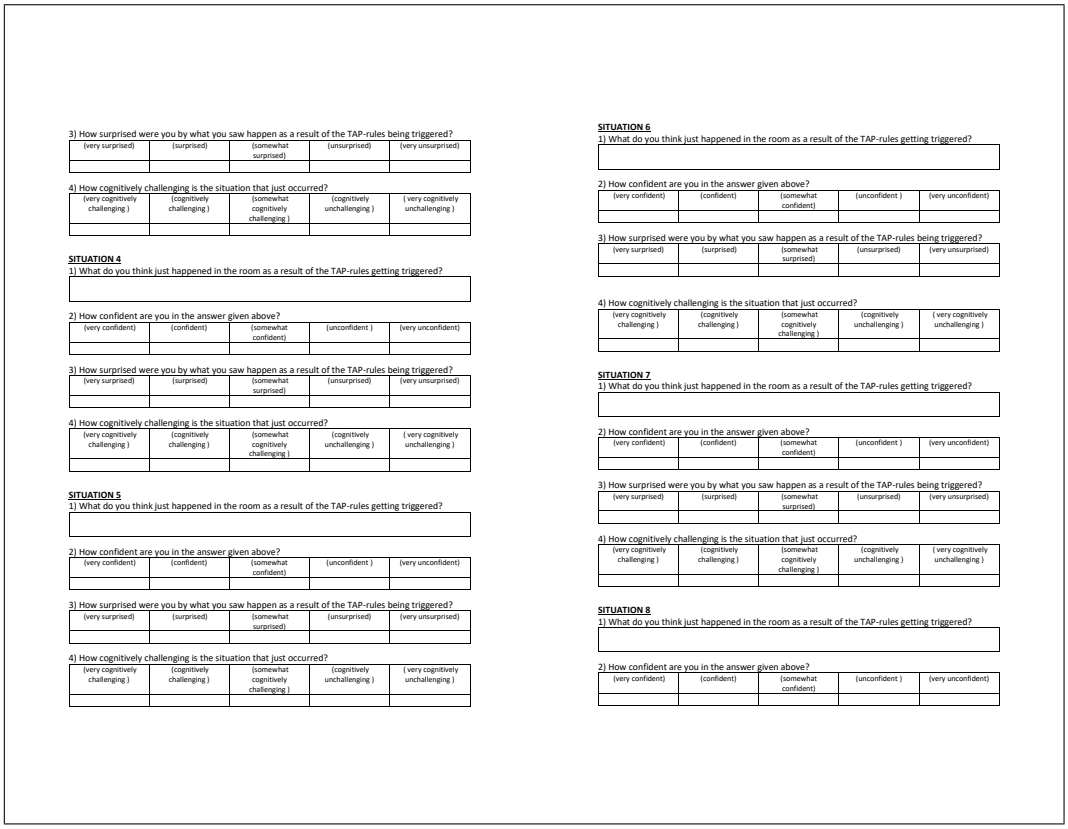
**SITUATION 3**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

**Figure A.7:** Survey of the user study (6)



3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**SITUATION 4**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**SITUATION 5**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**SITUATION 6**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**SITUATION 7**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

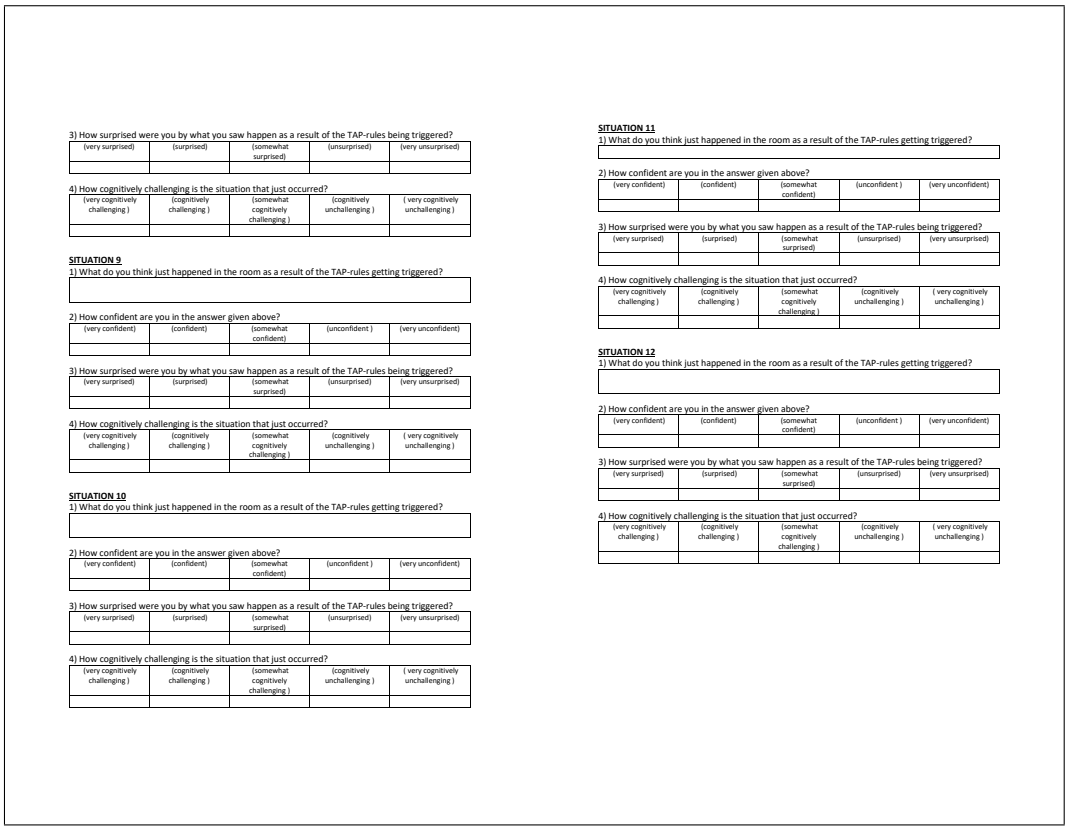
**SITUATION 8**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

**Figure A.8:** Survey of the user study (7)



3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**SITUATION 9**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

| |
|---|
| |

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**SITUATION 10**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

| |
|---|
| |

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**SITUATION 11**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

| |
|---|
| |

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**SITUATION 12**

1) What do you think just happened in the room as a result of the TAP-rules getting triggered?

| |
|---|
| |

2) How confident are you in the answer given above?

| (very confident) | (confident) | (somewhat confident) | (unconfident ) | (very unconfident) |
|---|---|---|---|---|
| | | | | |

3) How surprised were you by what you saw happen as a result of the TAP-rules being triggered?

| (very surprised) | (surprised) | (somewhat surprised) | (unsurprised) | (very unsurprised) |
|---|---|---|---|---|
| | | | | |

4) How cognitively challenging is the situation that just occurred?

| (very cognitively challenging ) | (cognitively challenging ) | (somewhat cognitively challenging ) | (cognitively unchallenging ) | ( very cognitively unchallenging ) |
|---|---|---|---|---|
| | | | | |

**Figure A.9:** Survey of the user study (8)

# Appendix B

# Results of the user study

| | Prio(D) | Prio(W) | Prio(WF) | Prio(DF) | D | W | OFF(DF) | LE(DF) | MIX(DF) |
|---|---|---|---|---|---|---|---|---|---|
| Mode | 1 | 1 | 1 | 1 | 1 | 1 | 3 | 3 | 3 |
| Median | 1 | 1 | 1.5 | 1 | 1 | 1 | 3 | 2 | 2 |
| Range | 4 | 2 | 2 | 3 | 3 | 2 | 4 | 2 | 3 |
| Minimum | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Maximum | 5 | 3 | 3 | 4 | 4 | 3 | 5 | 4 | 4 |

**Table B.1:** Data metrics for participants' answers to the second question of the survey's Color part (Part 1). For better understanding, the Likert-Scale values were ascribed numbers, 1 signifying 'Very confident' and 5 'Very unconfident.

| | ALT(WF) | OFF(WF) | RED(DF) | LE(WF) | ALT(DF) | RED(WF) | MIX(WF) |
|---|---|---|---|---|---|---|---|
| Mode | 3 | 2 | 1 | 2 | 2 | 3 | 2 |
| Median | 2 | 2 | 3 | 2 | 2 | 2.5 | 2 |
| Range | 3 | 4 | 4 | 2 | 3 | 4 | 3 |
| Minimum | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Maximum | 4 | 5 | 5 | 3 | 4 | 5 | 4 |

**Table B.2:** Data metrics for participants' answers to the second question of the survey's Color part (Part 2). For better understanding, the Likert-Scale values were ascribed numbers, 1 signifying 'Very confident' and 5 'Very unconfident.

| | Prio(A) | Prio(MF) | Prio(AF) | Prio(M) | M | A | OFF(MF) | 75%(MF) | LE(MF) |
|---|---|---|---|---|---|---|---|---|---|
| Mode | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 |
| Median | 1.5 | 1 | 1 | 1.5 | 1 | 2 | 2.5 | 2 | 2 |
| Range | 2 | 3 | 3 | 2 | 3 | 3 | 4 | 3 | 4 |
| Minimum | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Maximum | 3 | 4 | 4 | 3 | 4 | 4 | 5 | 4 | 5 |

**Table B.3:** Data metrics for participants' answers to the second question of the survey's Intensity part (Part 1). For better understanding, the Likert-Scale values were ascribed numbers, 1 signifying 'Very confident' and 5 'Very unconfident.

| | ALT(AF) | OFF(AF) | 30% (MF) | LE(AF) | ALT(MF) | 30%(AF) | 75%(AF) |
|---|---|---|---|---|---|---|---|
| Mode | 1 | 3 | 2 | 1 | 1 | 3 | 2 |
| Median | 2 | 3 | 2 | 2 | 2 | 3 | 2 |
| Range | 4 | 4 | 3 | 3 | 4 | 4 | 4 |
| Minimum | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Maximum | 5 | 5 | 4 | 4 | 5 | 5 | 5 |

**Table B.4:** Data metrics for participants' answers to the second question of the survey's Intensity part (Part 2). For better understanding, the Likert-Scale values were ascribed numbers, 1 signifying 'Very confident' and 5 'Very unconfident.

| | Prio(D) | Prio(W) | Prio(WF) | Prio(DF) | D | W | OFF(DF) | MIX(DF) | LE(DF) |
|---|---|---|---|---|---|---|---|---|---|
| Mode | 5 | 5 | 4 | 5 | 5 | 5 | 3 | 3 | 4 |
| Median | 5 | 5 | 4 | 5 | 5 | 5 | 3 | 3 | 4 |
| Range | 2 | 3 | 2 | 3 | 2 | 2 | 4 | 4 | 3 |
| Minimum | 3 | 2 | 3 | 2 | 3 | 3 | 1 | 1 | 2 |
| Maximum | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |

**Table B.5:** Data metrics for participants' answers to the third question of the survey's Color part (Part 1). For better understanding, the Likert-Scale values were ascribed numbers, 1 signifying 'Very surprised' and 5 'Very unsurprised'.

| | ALT(WF) | OFF(WF) | RED(DF) | LE(WF) | ALT(DF) | RED(WF) | MIX(WF) |
|---|---|---|---|---|---|---|---|
| Mode | 4 | 3 | 3 | 4 | 3 | 2 | 3 |
| Median | 3 | 3 | 3 | 4 | 3 | 3 | 3 |
| Range | 4 | 4 | 4 | 3 | 4 | 3 | 4 |
| Minimum | 1 | 1 | 1 | 2 | 1 | 1 | 1 |
| Maximum | 5 | 5 | 5 | 5 | 5 | 4 | 5 |

**Table B.6:** Data metrics for participants' answers to the third question of the survey's Color part (Part 2). For better understanding, the Likert-Scale values were ascribed numbers, 1 signifying 'Very surprised' and 5 'Very unsurprised'.

| | Prio(A) | Prio(MF) | Prio(AF) | Prio(M) | M | A | OFF(MF) | 75%(MF) |
|---|---|---|---|---|---|---|---|---|
| Mode | 5 | 4 | 5 | 5 | 5 | 5 | 3 | 4 |
| Median | 4.5 | 4 | 4.5 | 4.5 | 4.5 | 4.5 | 3 | 3.5 |
| Range | 2 | 2 | 2 | 2 | 3 | 3 | 4 | 3 |
| Minimum | 3 | 3 | 3 | 3 | 2 | 2 | 1 | 2 |
| Maximum | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |

**Table B.7:** Data metrics for participants' answers to the third question of the survey's Intensity part (Part 1). For better understanding, the Likert-Scale values were ascribed numbers, 1 signifying 'Very surprised' and 5 'Very unsurprised'.

| | LE(MF) | ALT(AF) | OFF(AF) | 30%(MF) | LE(AF) | ALT(MF) | 30%(AF) | 75%(AF) |
|---|---|---|---|---|---|---|---|---|
| Mode | 4 | 4 | 3 | 4 | 4 | 3 | 2 | 4 |
| Median | 4 | 3 | 3 | 4 | 4 | 3 | 3 | 3 |
| Range | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| Minimum | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Maximum | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |

**Table B.8:** Data metrics for participants' answers to the third question of the survey's Intensity part (Part 2). For better understanding, the Likert-Scale values were ascribed numbers, 1 signifying 'Very surprised' and 5 'Very unsurprised'.

| | Prio(D) | Prio(W) | Prio(WF) | Prio(DF) | D | W | OFF(DF) | MIX(DF) | LE(DF) |
|---|---|---|---|---|---|---|---|---|---|
| Mode | 5 | 5 | 4 | 5 | 5 | 5 | 4 | 3 | 5 |
| Median | 5 | 5 | 4 | 4.5 | 5 | 5 | 4 | 4 | 4 |
| Range | 2 | 2 | 3 | 2 | 2 | 2 | 3 | 3 | 3 |
| Minimum | 3 | 3 | 2 | 3 | 3 | 3 | 2 | 2 | 2 |
| Maximum | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |

**Table B.9:** Data metrics for participants' answers to the fourth question of the survey's Color part (Part 1). For better understanding, the Likert-Scale values were ascribed numbers, 1 signifying 'Very cognitively challenging' and 5 'Very cognitively unchallenging'.

| | ALT(WF) | OFF(WF) | RED(DF) | LE(WF) | ALT(DF) | RED(WF) | MIX(WF) |
|---|---|---|---|---|---|---|---|
| Mode | 4 | 4 | 4 | 4 | 3 | 3 | 3 |
| Median | 4 | 4 | 3.5 | 4 | 3.5 | 3 | 3 |
| Range | 4 | 3 | 4 | 4 | 3 | 4 | 3 |
| Minimum | 1 | 2 | 1 | 1 | 2 | 1 | 2 |
| Maximum | 5 | 5 | 5 | 5 | 5 | 5 | 5 |

**Table B.10:** Data metrics for participants' answers to the fourth question of the survey's Color part (Part 3). For better understanding, the Likert-Scale values were ascribed numbers, 1 signifying 'Very cognitively challenging' and 5 'Very cognitively unchallenging'.

| | Prio(A) | Prio(MF) | Prio(AF) | Prio(M) | M | A | OFF(MF) | 75%(MF) |
|---|---|---|---|---|---|---|---|---|
| Mode | 5 | 5 | 5 | 4 | 5 | 5 | 3 | 4 |
| Median | 4 | 4 | 5 | 4 | 4.5 | 5 | 3.5 | 4 |
| Range | 2 | 2 | 2 | 2 | 3 | 3 | 4 | 4 |
| Minimum | 3 | 3 | 3 | 3 | 2 | 2 | 1 | 1 |
| Maximum | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |

**Table B.11:** Data metrics for participants' answers to the fourth question of the survey's Intensity part (Part 1). For better understanding, the Likert-Scale values were ascribed numbers, 1 signifying 'Very cognitively challenging' and 5 'Very cognitively unchallenging'.

| | LE(MF) | ALT(AF) | OFF(AF) | 30%(MF) | LE(AF) | ALT(MF) | 30%(AF) | 75%(AF) |
|---|---|---|---|---|---|---|---|---|
| Mode | 4 | 3 | 2 | 3 | 4 | 3 | 3 | 5 |
| Median | 4 | 3 | 3 | 4 | 4 | 3 | 3 | 3.5 |
| Range | 4 | 3 | 4 | 3 | 3 | 4 | 4 | 4 |
| Minimum | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 1 |
| Maximum | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |

**Table B.12:** Data metrics for participants' answers to the fourth question of the survey's Intensity part (Part 2). For better understanding, the Likert-Scale values were ascribed numbers, 1 signifying 'Very cognitively challenging' and 5 'Very cognitively unchallenging'.

| | D | W | MIX | ALT | OFF | RED | LE |
|---|---|---|---|---|---|---|---|
| D | - | 0.317311 | 0.000931 | 0.000953 | 0.000307 | 0.000789 | 0.009669 |
| W | 0.317311 | - | 0.000269 | 0.000243 | 0.000235 | 0.000265 | 0.003071 |
| MIX | 0.000931 | 0.000269 | - | 0.895245 | 0.075507 | 0.084742 | 0.096648 |
| ALT | 0.000953 | 0.000243 | 0.895245 | - | 0.095831 | 0.096745 | 0.056291 |
| OFF | 0.000307 | 0.000235 | 0.075507 | 0.095831 | - | 0.865753 | 0.006049 |
| RED | 0.000789 | 0.000265 | 0.084742 | 0.096745 | 0.865753 | - | 0.005116 |
| LE | 0.009669 | 0.003071 | 0.096648 | 0.056291 | 0.006049 | 0.005116 | - |

**Table B.13:** Overview of the 'p' values of the Wilcoxon signed-rank test for the second question (COLOR). The significantly different situations are highlighted in blue.

| | M | A | MIX | ALT | OFF | 30% | LE |
|---|---|---|---|---|---|---|---|
| M | - | 0.458404 | 0.006678 | 0.051468 | 0.002308 | 0.014926 | 0.045721 |
| A | 0.458404 | - | 0.017154 | 0.153965 | 0.002041 | 0.030205 | 0.097822 |
| MIX | 0.006678 | 0.017154 | - | 0.311817 | 0.985981 | 0.677400 | 0.064820 |
| ALT | 0.051468 | 0.153965 | 0.311817 | - | 0.105919 | 0.269046 | 0.500265 |
| OFF | 0.002308 | 0.002041 | 0.985981 | 0.105919 | - | 0.733970 | 0.023024 |
| 30% | 0.014926 | 0.030205 | 0.677400 | 0.269046 | 0.733970 | - | 0.071152 |
| LE | 0.045721 | 0.097822 | 0.064820 | 0.500265 | 0.023024 | 0.071152 | - |

**Table B.14:** Overview of the 'p' values of the Wilcoxon signed-rank test for the second question (INTENSITY). The significantly different situations are highlighted in blue.

| | D | W | MIX | ALT | OFF | RED | LE |
|---|---|---|---|---|---|---|---|
| D | - | 0.317311 | 0.000077 | 0.000174 | 0.000081 | 0.000038 | 0.002896 |
| W | 0.317311 | - | 0.000076 | 0.000117 | 0.000079 | 0.000025 | 0.001847 |
| MIX | 0.000077 | 0.000076 | - | 0.390994 | 0.228326 | 0.139172 | 0.000153 |
| ALT | 0.000174 | 0.000117 | 0.390994 | - | 0.042332 | 0.004627 | 0.001633 |
| OFF | 0.000081 | 0.000079 | 0.228326 | 0.042332 | - | 0.694639 | 0.000151 |
| RED | 0.000038 | 0.000025 | 0.139172 | 0.004627 | 0.694639 | - | 0.000084 |
| LE | 0.002896 | 0.001847 | 0.000153 | 0.001633 | 0.000151 | 0.000084 | - |

**Table B.15:** Overview of the 'p' values of the Wilcoxon signed-rank test for the third question (COLOR). The significantly different situations are highlighted in blue.

| | M | A | MIX | ALT | OFF | 30% | LE |
|---|---|---|---|---|---|---|---|
| M | - | 0.527089 | 0.001744 | 0.003864 | 0.000121 | 0.000930 | 0.003615 |
| A | 0.527089 | - | 0.008469 | 0.007198 | 0.000310 | 0.002404 | 0.077966 |
| MIX | 0.001744 | 0.008469 | - | 0.470551 | 0.125262 | 0.360828 | 0.011227 |
| ALT | 0.003864 | 0.007198 | 0.470551 | - | 0.111159 | 0.979124 | 0.003869 |
| OFF | 0.000121 | 0.000310 | 0.125262 | 0.111159 | - | 0.256862 | 0.000349 |
| 30% | 0.000930 | 0.002404 | 0.360828 | 0.979124 | 0.256862 | - | 0.003714 |
| LE | 0.003615 | 0.077966 | 0.011227 | 0.003869 | 0.000349 | 0.003714 | - |

**Table B.16:** Overview of the 'p' values of the Wilcoxon signed-rank test for the third question (INTENSITY). The significantly different situations are highlighted in blue.

| | M | A | MIX | ALT | OFF | 30% | LE |
|---|---|---|---|---|---|---|---|
| M | - | 0.317311 | 0.000241 | 0.000117 | 0.000117 | 0.000116 | 0.001982 |
| A | 0.317311 | - | 0.000255 | 0.000117 | 0.000119 | 0.000117 | 0.002007 |
| MIX | 0.000241 | 0.000255 | - | 0.546809 | 0.882217 | 0.169382 | 0.004500 |
| ALT | 0.000117 | 0.000117 | 0.546809 | - | 0.414014 | 0.189712 | 0.007618 |
| OFF | 0.000117 | 0.000119 | 0.882217 | 0.414014 | - | 0.054520 | 0.030625 |
| 30% | 0.000116 | 0.000117 | 0.169382 | 0.189712 | 0.054520 | - | 0.002391 |
| LE | 0.001982 | 0.002007 | 0.004500 | 0.007618 | 0.030625 | 0.002391 | - |

**Table B.17:** Overview of the 'p' values of the Wilcoxon signed-rank test for the fourth question (COLOR). The significantly different situations are highlighted in blue.

| | M | A | MIX | ALT | OFF | 30% | LE |
|---|---|---|---|---|---|---|---|
| M | - | 0.772830 | 0.019527 | 0.014755 | 0.004526 | 0.024150 | 0.061945 |
| A | 0.772830 | - | 0.032407 | 0.028852 | 0.005237 | 0.017034 | 0.111852 |
| MIX | 0.019527 | 0.032407 | - | 0.622899 | 0.915642 | 0.753886 | 0.026291 |
| ALT | 0.014755 | 0.028852 | 0.622899 | - | 0.961476 | 0.721580 | 0.016238 |
| OFF | 0.004526 | 0.005237 | 0.915642 | 0.961476 | - | 0.806827 | 0.008650 |
| 30% | 0.024150 | 0.017034 | 0.753886 | 0.721580 | 0.806827 | - | 0.017169 |
| LE | 0.061945 | 0.111852 | 0.026291 | 0.016238 | 0.008650 | 0.017169 | - |

**Table B.18:** Overview of the 'p' values of the Wilcoxon signed-rank test for the fourth question (INTENSITY). The significantly different situations are highlighted in blue.

# Bibliography

Mussab Alaa, Aws Alaa Zaidan, Bilal Bahaa Zaidan, Mohammed Talal, and Miss Laiha Mat Kiah. A review of smart home applications based on internet of things. *Journal of Network and Computer Applications*, 97:48–65, 2017.

Sonny Ali and Zia Yusuf. Mapping the smart-home market, 2018.

Luigi Atzori, Antonio Iera, and Giacomo Morabito. The internet of things: A survey. *Computer networks*, 54(15): 2787–2805, 2010.

Nathaniel Ayewah, William Pugh, David Hovemeyer, J David Morgenthaler, and John Penix. Using static analysis to find bugs. *IEEE software*, 25(5):22–29, 2008.

Nazmiye Balta-Ozkan, Benjamin Boteler, and Oscar Amerighi. European smart home market development: Public views on technical and economic aspects across the united kingdom, germany and italy. *Energy Research & Social Science*, 3:65–77, 2014.

Will Brackenbury, Abhimanyu Deora, Jillian Ritchey, Jason Vallee, Weijia He, Guan Wang, Michael L Littman, and Blase Ur. How users interpret bugs in trigger-action programming. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–12, 2019.

Dipankar Chaki and Athman Bouguettaya. Dynamic conflict resolution of iot services in smart homes. In *Service-Oriented Computing: 19th International Conference, ICSOC 2021, Virtual Event, November 22–25, 2021, Proceedings 19*, pages 368–384. Springer, 2021.

Diane J Cook, Aaron S Crandall, Brian L Thomas, and Narayanan C Krishnan. Casas: A smart home in a box. *Computer*, 46(7):62–69, 2012.

Sven Coppers, Davy Vanacken, and Kris Luyten. Fortclash: Predicting and mediating unintended behavior in home automation. *Proceedings of the ACM on Human-Computer Interaction*, 6(EICS):1–20, 2022.

Fulvio Corno, Luigi De Russis, and Alberto Monge Roffarello. Empowering end users in debugging trigger-action rules. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2019.

Anind K Dey, Timothy Sohn, Sara Streng, and Justin Kodama. icap: Interactive prototyping of context-aware applications. In *Pervasive Computing: 4th International Conference, PERVASIVE 2006, Dublin, Ireland, May 7-10, 2006. Proceedings 4*, pages 254–271. Springer, 2006.

Earlence Fernandes, Amir Rahmati, Jaeyeon Jung, and Atul Prakash. Decentralized action integrity for trigger-action iot platforms. In *Proceedings 2018 Network and Distributed System Security Symposium*, 2018.

Jacopo Fiorenza and Andrea Mariani. Improving trigger action programming in smart buildings through suggestions based on behavioral graphs analysis. 2015.

Vaishnavi S Gunge and Pratibha S Yalagi. Smart home automation: a literature review. *International Journal of Computer Applications*, 975(8887-8891), 2016.

Richard Harper. *Inside the smart home*. Springer Science & Business Media, 2006.

Justin Huang and Maya Cakmak. Supporting mental model accuracy in trigger-action programming. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 215–225, 2015.

Elena Korneeva, Nina Olinder, and Wadim Strielkowski. Consumer attitudes to the smart home technologies and the internet of things (iot). *Energies*, 14(23):7913, 2021.

Henry Lieberman, Fabio Paternò, Markus Klann, and Volker Wulf. End-user development: An emerging

paradigm. In *End user development*, pages 1–8. Springer, 2006.

Marco Manca, Fabio Paternò, Carmen Santoro, and Luca Corcella. Supporting end-user debugging of trigger-action rules for iot applications. *International Journal of Human-Computer Studies*, 123:56–69, 2019.

Yan Meng, Wei Zhang, Haojin Zhu, and Xuemin Sherman Shen. Securing consumer iot in the smart home: Architecture, challenges, and countermeasures. *IEEE Wireless Communications*, 25(6):53–59, 2018.

Fereshteh Jadidi Miandashti, Mohammad Izadi, Ali Asghar Nazari Shirehjini, and Shervin Shirmohammadi. An empirical approach to modeling user-system interaction conflicts in smart homes. *IEEE Transactions on Human-Machine Systems*, 50(6):573–583, 2020.

Thinagaran Perumal, Md Nasir Sulaiman, Soumya Kanti Datta, Thinaharan Ramachandran, and Chui Yew Leong. Rule-based conflict resolution framework for internet of things device management in smart home environment. In *2016 IEEE 5th Global Conference on Consumer Electronics*, pages 1–2. IEEE, 2016.

Karen Rose, Scott Eldridge, and Lyman Chapin. The internet of things: An overview. *The internet society (ISOC)*, 80: 1–50, 2015.

Amit Kumar Sikder, Leonardo Babun, Z Berkay Celik, Abbas Acar, Hidayet Aksu, Patrick McDaniel, Engin Kirda, and A Selcuk Uluagac. Kratos: Multi-user multi-device-aware access control system for the smart home. In *Proceedings of the 13th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, pages 1–12, 2020.

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

Lefan Zhang, Weijia He, Jesse Martinez, Noah Brackenbury, Shan Lu, and Blase Ur. Autotap: Synthesizing and repairing trigger-action programs using ltl properties. In

*2019 IEEE/ACM 41st international conference on software engineering (ICSE)*, pages 281–291. IEEE, 2019.

Lefan Zhang, Cyrus Zhou, Michael L Littman, Blase Ur, and Shan Lu. Helping users debug trigger-action programs. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(4):1–32, 2023.

Valerie Zhao, Lefan Zhang, Bo Wang, Michael L Littman, Shan Lu, and Blase Ur. Understanding trigger-action programs through novel visualizations of program differences. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2021.

# Index

Typeset April 19, 2023