Understanding Circumstances for Desirable Proactive Behaviour of Voice Assistants: The Proactivity Dilemma

Nima Zargham
zargham@uni-bremen.de
Digital Media Lab
University of Bremen
Germany

Leon Reicherts
l.reicherts.17@ucl.ac.uk
University College London
United Kingdom

Michael Bonfert
bonfert@uni-bremen.de
Digital Media Lab
University of Bremen
Germany

Sarah Theres Völkel
sarah.voelkel@ifi.lmu.de
LMU Munich
Germany

Johannes Schöning
johannes.schoening@unisg.ch
University of St. Gallen
Switzerland

Rainer Malaka
malaka@uni-bremen.de
Digital Media Lab
University of Bremen
Germany

Yvonne Rogers
y.rogers@ucl.ac.uk
University College London
United Kingdom

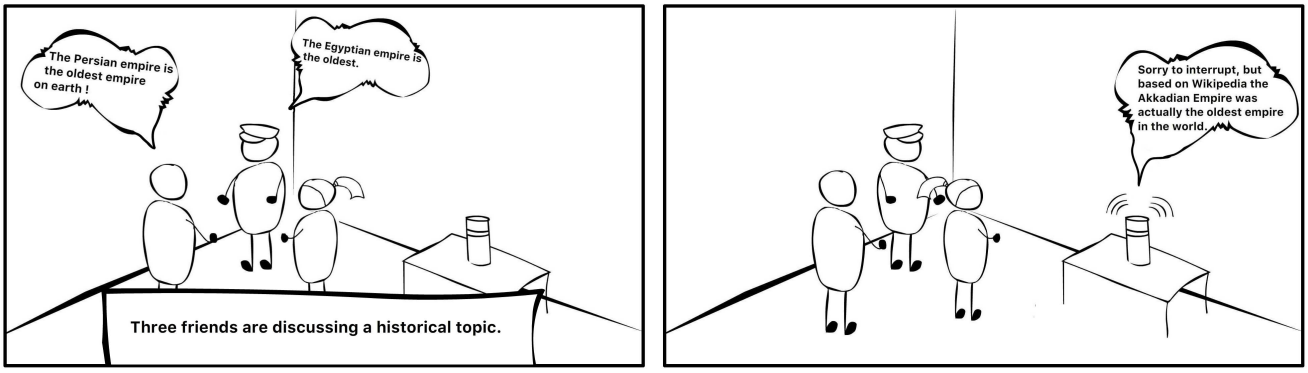


Figure 1: An example storyboard used in the study. In this scenario (S4), the agent proactively approaches users based on their conversation.

ABSTRACT

The next major evolutionary stage for voice assistants will be their capability to initiate interactions by themselves. However, to design proactive interactions, it is crucial to understand whether and when this behaviour is considered useful and how desirable it is perceived for different social contexts or ongoing activities. To investigate people's perspectives on proactivity and appropriate circumstances for it, we designed a set of storyboards depicting a variety of proactive actions in everyday situations and social settings and presented them to 15 participants in interactive interviews. Our findings suggest that, although many participants see benefits in agent proactivity, such as for urgent or critical issues, there are concerns about interference with social activities in multi-party settings, potential loss of agency, and intrusiveness. We discuss our implications for designing voice assistants with desirable proactive features.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CUI 2022, July 26–28, 2022, Glasgow, United Kingdom
© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-9739-1/22/07…$15.00
https://doi.org/10.1145/3543829.3543834

CCS CONCEPTS

• **Human-centered computing → Natural language interfaces**; *Empirical studies in HCI*; Scenario-based design.

KEYWORDS

Proactive Agents, Voice Assistants, Conversational Agents, Smart Speakers, Smart Home

ACM Reference Format:
Nima Zargham, Leon Reicherts, Michael Bonfert, Sarah Theres Völkel, Johannes Schöning, Rainer Malaka, and Yvonne Rogers. 2022. Understanding Circumstances for Desirable Proactive Behaviour of Voice Assistants: The Proactivity Dilemma. In *4th Conference on Conversational User Interfaces (CUI 2022), July 26–28, 2022, Glasgow, United Kingdom.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3543829.3543834

1 INTRODUCTION

Voice assistants (VAs) are becoming more intelligent and capable of supporting increasingly complex tasks and dialogues. People use them for controlling smart home devices, information seeking, entertainment, shopping, and activity management, among others [30]. As more and more VAs are finding their way into our homes in the form of smart speakers, they play a greater role as digital everyday helpers. However, despite the broad range of use cases and increasingly advanced language understanding and "conversational abilities" of smart speakers, these devices are still interacting mainly in a reactive manner, responding to the users' inquiry. The interaction starts with users saying the wake word followed by an inquiry, and only then the agent responds to the user.

With rapid advances in artificial intelligence, natural language understanding, and sensing techniques, VAs are becoming more capable of understanding users' behaviours, preferences, intentions, and surroundings, which opens up a broad landscape of opportunities for proactive interactions. Proactive behaviours from VAs are considered as agent-initiated interactions triggered by events related to the user(s) and their environment, as opposed to user-initiated inquiries or pre-configured actions, such as reminders, alerts, or routines set by the user [31]. In several studies, researchers have begun to examine proactive behaviour of VAs and proposed to use it for specific situations and environments [20, 27, 33]. Others have looked into the timing of proactive VA interruptions [7] and how such interruptions should be designed [11]. Moreover, certain commercial assistants already support some basic proactive services, such as reminding users of their schedule[1], automating home routines[2], and supporting home safety and security[3].

Nevertheless, the current state of research still lacks a deeper understanding of how people perceive and feel about such interactions at home. As proactive VAs require to monitor and process users' behaviour constantly, privacy concerns are likely to intensify. Tabassum et al. [38] highlighted that privacy remains a key concern that limits users' willingness to use proactive VAs. However, other factors driving users' acceptance of proactive VAs beyond privacy concerns, such as the usefulness and appropriateness of situations to be interrupted, remain unexplored.

To close the gap in understanding people's perceptions of proactive VAs in a domestic setting, we present an elicitation study to investigate the desirability of agent-initiated interventions, i.e., high usefulness, high appropriateness, and low invasiveness. Therefore, we address the following research questions:

RQ1: Under which circumstances is proactive behaviour by a voice assistant perceived as desirable?
RQ2: How should proactive interventions be initiated by the voice assistant?

To answer our research questions, we designed a set of storyboards illustrating a range of possible proactive interventions in a home environment. An example storyboard is shown in Figure 1. In an online elicitation study, we conducted interactive interviews. The participants went through a series of tasks based on these storyboards using a virtual, collaborative whiteboard to evaluate and contemplate on the concept from different perspectives. Our results show that key factors for a desirable proactive intervention are the people in the room, the type of ongoing activity, the urgency of the topic, the user's current emotional state, and the agent's initiation and phrasing of the intervention. The main contribution of this work is empirical evidence for the situational desirability of proactive VA behaviour, thereby providing a deeper understanding of factors influencing user acceptance. Our findings point toward a dilemma. As such, interactions are often perceived as useful but at the same time invasive or inappropriate. Therefore, we propose an initiation process model for minimizing the intrusiveness of useful features.

2 RELATED WORK

Although current commercial smart speakers support a limited set of proactive features such as Amazon Echo displaying a specific light pattern to visualise notifications and messages, or Google Home delivering proactive reminders for upcoming meetings based on the user's calendar, such devices remain primarily reactive with users initiating interactions. However, proactive interactions in such devices can open up new opportunities for supporting, probing, or inspiring their users [42]. In this section, we summarise related work on proactivity in VAs, opportune moments for VAs to proactively engage with the user, privacy concerns, and appropriateness of proactive interventions.

2.1 Proactivity in VAs

A survey with 1,550 participants by Schmidt and Braunger [33] revealed that proactivity is a favoured feature by users. Similarly, an elicitation study by Völkel et al. [41] on users' envisioned dialogues with a perfect voice assistant showed that many participants imagined proactive voice assistant behaviour to be desirable. In particular, the envisioned dialogues pointed to agents being able to anticipate the next possible actions and to give suggestions without being requested by the users. Andolina et al. [3] designed a proactive search agent that would listen to users conversing and present information about their conversation based on entities detected in their dialogue. They found that their agent could effectively support the conversation with facts and ideas without causing much interruption to the conversation's flow.

On the other hand, there are also potential downsides of proactive voice assistants, in particular concerning privacy [38], which we discuss below. In a study about in-car assistants, Braun et al. [5] reported that people have mixed opinions on whether the voice assistant should initiate conversations and that they only accept

[1]https://www.techrepublic.com/article/google-nest-smart-speakers-a-cheat-sheet, last accessed 2022-02-28
[2]https://www.theverge.com/2021/1/25/22249044/amazon-alexa-update-proactive-hunches-guard-plus-subscription, last accessed 2022-02-28
[3]https://www.cnet.com/home/smart-home/what-amazon-alexa-will-tell-us-in-2019, last accessed 2022-02-28

proactivity if the assistant can act like an authentic, human co-driver [5]. However, they did not investigate what factors influence some users' reluctance to engage with proactive VAs.

While there seems to be a demand for proactivity, there is little knowledge about what makes a proactive voice assistant desirable [2]. Since today's conversations with voice assistants are highly constrained, task-oriented, and impersonal [8–10, 12, 22, 29], proactive interactions in such devices could open up new opportunities and potentially empower a broad range of applications [42].

2.2 Opportune Moments for Proactive VAs

Opportune moments for interaction refer to moments where the disruption of the user's current activity is at a minimum level [39]. Although it is fairly easy and natural for people to assess another person's current activity before starting to interact with them, it is a big challenge to design such behaviour for machines [15, 32, 39]. Identifying these moments for VAs to start interacting with a user is particularly challenging, as speech interaction requires immediate attention and can easily interrupt users with their current activities or social interactions [37]. To achieve proactivity, the voice assistant needs to be context-aware and detect opportune moments to initiate interactions. Previous studies have explored opportune moments for VA interventions in homes [7, 16, 19, 43], cars [17, 18, 33–36], and other settings [20].

Conducting an experiencing sampling study with smart speakers in people's homes, Cha et al. [7] found that the key determinants for opportune moments are linked to personal contextual factors such as busyness, mood, and urgency, as well as the other contextual factors related to everyday routines at home, including social context such as presence of other people, and user mobility. Similarly, a study by Nothdurft et al. [28] suggests that the most important factors to decide if proactive behaviour is desired are the importance of the intervention for the user, users' surrounding and their mental state, and the accurate placement of the interaction.

Apart from identifying opportune moments for proactivity, researchers have also examined *how* the agent should initiate a conversation. Edwards et al. [11] looked at how people interrupt another person who is engaged in a complex task, as an approach to inform the design of proactive VAs. Their results showed that the level of urgency significantly affects how long it takes for people to start interrupting. Arias et al. [4] suggest that agents should notify users before proactively engaging with them to make sure they are willing to interact at the specific moment. Moreover, users should be in charge of deciding which information they are proactively told by an agent [4].

These studies underline that a lack of contextual knowledge is detrimental to users' acceptance of proactive VAs. While previous work has pointed out influencing factors such as the urgency of the task, little is known about specific situations in which users find a proactive VA appropriate.

2.3 Privacy Concerns in Voice Assistant Use

Preservation of privacy is a key to the users' acceptance of smart speakers. Specifically, in a home environment, stressing the importance of user privacy and security is crucial. A study by Lau et al. [21] showed that many people refrain from adopting smart speakers because they have privacy concerns or distrust the companies offering smart speakers. Malkin et al. [24] surveyed smart speaker owners and found that users are not comfortable with permanently preserving user recordings, especially those that include children and guests. Moreover, users were strongly opposed to the use of their data by third parties or for advertising.

When it comes to proactive services, such concerns intensify as the agents need to be more context-aware, have access to more personal data which are usually uploaded to and processed on companies' cloud services, and act out of the user's control. Previous research has shown that privacy represents the key challenge for proactively initiated interactions [27]. Tabassum et al. [38] conducted an online survey to explore user preferences and expectations of proactive VAs and found that, even though users perceived the services as useful, they were uncomfortable with the always-listening nature of such systems. Yet, many users were willing to share even sensitive conversations to receive more personalised and contextual services. Likewise, Cha et al. [7] found that users willingly accept to compromise their privacy in exchange for a smart speaker that offers personalised care.

Including privacy-preserving features is essential when designing proactive VAs. Previous work shows that giving users the option to examine the recordings and actions taken by the systems [25] as well as transparency on the recorded data are decisive factors for the acceptance of such proactive technologies [26]. But even with full control over what private data is shared or stored, the VA's active role and interference in the private sphere in domestic situations might be experienced as inappropriate, which needs further investigation.

2.4 Appropriate Proactive Interventions

At times, certain proactive behaviour can cause discomfort and be perceived as disruptive and invasive [4]. For successful proactive interventions, not only users' current mood but also cultural and social context need to be considered – in particular if the agent takes on a more human-like role, like a personal coach [27].

Luria et al. [23] identified three thresholds of agent proactivity including reactive to user requests, proactive by providing information, or proactive by providing recommendations for a course of action, with users differing in their comfort levels with each threshold. In a study to explore socially sophisticated agents in a domestic setting, they witnessed that most participants were open to the idea of a proactive agent in a multi-user situation, but nobody wanted the agent to enforce recommendations such as preventing them from ordering unhealthy food.

In a previous study [31], we conducted an online survey in which we asked users to rate a series of hypothetical storyboards depicting situations at home where an agent proactively addresses users, based on the criteria of *usefulness*, *pleasantness*, *appropriateness*. We found that even when participants perceived the agent interventions as useful in general, the ratings for appropriateness were much lower, suggesting that appropriateness given (social) context is a crucial factor for the overall acceptability of the interactions. While the quantitative study design could reveal interesting differences in people's ratings along those criteria, the purpose of

the present study is to gain a comprehensive understanding of the reasons behind these differences through a qualitative approach.

Overall, despite the popularity of proactive features in VAs, previous literature still lacks an understanding of user perceptions of desirable proactive behaviour in domestic settings considering situational factors. In this paper, we build upon prior work by engaging users to reflect on the contexts in which they would consider proactive interventions useful and appropriate.

3 STUDY DESIGN

In this work, we sought to investigate circumstances for a desirable proactive voice assistant in a home environment. We used an approach inspired by scenario-based design methods [6] and vignette experiments [1], which allows us to investigate (future) technologies despite current technological limitations. In our online interviews, we present participants with different hypothetical scenarios, which are illustrated by graphical storyboards to better visualize the situation and spatial configuration of the specific home environment, the user(s), and the smart speaker. We developed an interactive task-based interview procedure, designed to elicit participants' reflections on the scenarios from different perspectives. Hence, in addition to asking participants for their general thoughts on the scenarios, they were asked to complete different tasks on a virtual whiteboard. This allowed us to explore in-depth deliberations around proactive features and collect richer data. Ethical approval was received for the study from University College London.

3.1 Storyboards

The initial set of scenarios about proactive VAs in domestic settings was based on the eight scenarios from our previous study [31] where we used them to collect participant ratings across different dimensions such as perceived usefulness and appropriateness in an online study. In the present study, we reuse most of these scenarios and investigate them with a qualitative approach to shed light on *why* some proactive behaviours are seen as more or less desirable in certain contexts. However, we initially added eight additional scenarios to expand the range and the diversity of scenarios, based on further classification criteria we considered relevant for covering the large spectrum of conceivable circumstances, such as varying levels of urgency, the number of people present, or the extent of interruption (e.g., of an ongoing human-human interaction). We presented the extended set of 16 scenarios and classifications to two VA experts and asked them to add further scenarios and classifications they think are missing or might complement the existing ones. Considering their feedback, we refined the scenario selection and the classification scheme and asked three HCI researchers to independently code the scenarios using our scheme. Based on the coded scenarios, we selected nine scenarios that covered all the classifications. Seven of these nine scenarios were identical or almost identical to those of our previous research [31], including a scenario which highlighted potential challenges of proactive VAs (S8).

All scenarios were presented as cartoon sketches with two panels. The storyboards were designed in a way that should minimize any cultural and ethnic cues so that participants could put themselves in the shoes of the characters. The characters were designed without any facial expressions to avoid influencing participants' interpretation of the scenarios. As in the original storyboards, the cylinder-shaped appearance of the voice assistant was similar to a conventional smart speaker. The fictional agent was given the gender-ambiguous name "Jay" to reduce gender bias. The complete set of storyboards used for this study can be found in the Appendix and is briefly described in the following list.

- *S1 Meeting Reminder*: After the user has repeatedly "snoozed" the alarm, Jay reminds her of an upcoming meeting.
- *S2 Health Risk*: From the sound of the cough, Jay suspects an elderly user to have a respiratory infection and offers to arrange a doctor's appointment.
- *S3 Cooking Inspiration*: Two friends are contemplating about dinner when Jay offers to suggest recipes based on what is in the fridge.
- *S4 Fact Checking*: Three friends discuss a historical topic when Jay interrupts them to get a fact right.
- *S5 Disagreement Clarification*: Two people remember differently what they agreed on when Jay settles the disagreement by quoting what they said.
- *S6 Nudging*: When the user asks Jay to play a TV series, Jay suggests stopping earlier than last night.
- *S7 Technical Support*: A person asks their friend for help in setting up new headphones. As the friend is busy, Jay offers to assist.
- *S8 Fact Spoiler*: During quiz night, Jay reveals the correct solution before the players had a chance to answer.
- *S9 Emergency*: Jay detects a fire in the apartment, immediately calls the fire department and warns the sleeping residents.

3.2 Participants

15 people participated in the study, of which seven self-identified as female and eight as male. They were between 22 and 35 years of age ($M = 27.86$, $SD = 4.47$). Five participants had a bachelor's degree, nine had a master's degree, and one had a PhD. Participants were recruited using convenience sampling. The participation was voluntary and uncompensated. The recruitment continued until data saturation was reached, satisfying the recommended sample sizes of theoretical saturation from the literature [13, 40]. Two-thirds of our participants have previously used VAs (four rarely, six often). Seven participants (46.6 %) owned a smart speaker. All participants were proficient in English.

3.3 Procedure

Every study session was held remotely via video calls. The participants were asked to give informed consent and fill in the demographics questionnaire prior to the session. At the beginning of each session, participants were informed about the study procedure and the concept of a proactive VA. The study tasks were performed through the virtual whiteboard tool Miro[4]. All participants had a short familiarization phase with Miro and the virtual board. During

[4]https://miro.com, last accessed June 18, 2022

the sessions, the participants would share their screens with the interviewer to be guided through the tasks.

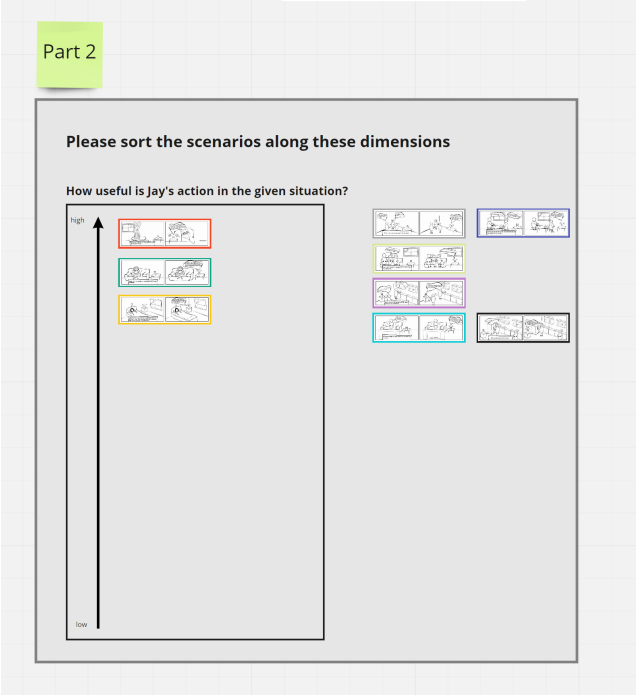


Figure 2: A partial screenshot of the card sorting task on Miro where a user is bringing the storyboards into order regarding their usefulness.

We designed a sequence of different interview parts combined with specific tasks to elicit how participants perceive the depicted (social) situation and how they think Jay's intervention affects it, as well as to understand how proactive interventions need to be designed to mitigate any negative effects on people's (social) activities. Before starting the interview, the interviewer explained that participants should assume the data is processed locally on the device. While some of Jay's features may not yet be feasible today with offline/on-device processing, we wanted to avoid participants solely worrying about data privacy, as this aspect is well researched [27, 38]. In an initial short interview, we gathered first impressions of the individual scenarios. Afterwards, participants were asked to sort the scenarios in terms of usefulness, appropriateness, and invasiveness in a card-sorting task as shown in Figure 2. After that, they speculated how each scenario may evolve and how the characters may respond to Jay's intervention. In the third task, participants were asked to choose the most invasive and the most inappropriate scenarios to then re-imagine an improved intervention. In the final task, participants were asked to decide for each scenario how they would like the VA to initiate the interaction and if it should provide a cue before starting to speak. After going through all the tasks, the session concluded with a short semi-structured interview in which participants gave their overall impression and elaborated on the potentials and challenges of proactive smart speakers. All sessions were audio-recorded for later analysis. The sessions took approximately 51.3 minutes on average (SD = 10.6).

3.4 Data Analysis

Our data analysis focused on two parts: content from the virtual whiteboards and spoken statements from the interviews. The information from the completed tasks in each participant's board was extracted and consolidated. The resulting data set was reviewed and discussed by three researchers. Some tasks were designed to produce categorical data, such as the card sorting and the cue selection tasks, which were examined using descriptive statistics. The interview segments were analysed for subsequent triangulation with the data from the boards. The transcripts of the interviews were independently coded by two researchers using inductive coding, where a single quote could be assigned to multiple codes. Codes were merged and consolidated by the two researchers. Three researchers discussed the codes, resolved disagreements, and derived themes which can be categorized into (I) perceived helpfulness, (II) privacy and mistrust, (III) consideration of social context, (IV) configuration and control, (V) and initiation and phrasing of interventions.

4 FINDINGS

Overall, participants had diverse opinions on proactive behaviours of a smart speaker. Some generally liked proactive interventions and valued the additional features, while others disliked them: "I would rather ask [for help] than getting help without asking" (P6). Some had mixed feelings: "It's like a double-edged sword: both helps and can intrude" (P5). In this section, we will share details about the conflicting appreciation and concerns of our participants. In each subsection, the results are presented first, followed directly by an interpretation in which we also discuss potential design implications.

4.1 Perceived Helpfulness

Participants sorted the scenarios in terms of usefulness, appropriateness, and invasiveness of the assistant's behaviour. The median rank of the scenarios in the order from 1st to 9th place is shown in Table 1. The scenario *Emergency* was considered most useful ($Medianrank$ = 1), most appropriate (Md = 1), and least invasive (Md = 9) by most participants. On the other hand, *Fact Spoiler* was ranked least useful (Md = 9) and least appropriate (Md = 9). Participants ranked *Disagreement Clarification* most invasive, with 86 % sorting it within the last three ranks, and highly inappropriate (Md = 8). We observed considerable similarities between the outcome of the three factors. The median ranks of how useful and appropriate the scenarios were assessed strongly correlate ($r_{Spearman} = 0.911$). The usefulness is furthermore negatively correlated with the invasiveness ($r_s = -0.830$). Similarly, this strong negative correlation occurs between appropriateness and invasiveness ($r_s = -0.902$). That is, the more useful and appropriate a situation is perceived, the less invasive it is ranked in general. However, there are several exceptions regarding invasiveness that we discuss below. All correlations are statistically significant with $p < .006$ on a Bonferroni-corrected alpha level of $\alpha = .016$.

An important factor for the proactive assistants' perceived helpfulness was the amount of time its intervention could save the user.

Scenario	Useful	Appropriate	Invasive
Emergency	1	1	9
Health Risk	2	4	5
Meeting Reminder	3	3	5
Cooking Inspiration	4	3	7
Technical Support	4	4	7
Nudging	6	6	4
Fact Checking	7	7	3
Disagreement Clarification	8	8	1
Fact Spoiler	9	9	3

Table 1: Median of how useful, appropriate, and invasive the scenarios were ranked in the card sorting task (1 being the highest and 9 the lowest rank). Scenarios are sorted based on their usefulness rankings.

The time saving aspect was mentioned in particular for the *Cooking Inspiration* (four times), the *Meeting Reminder* (five times), and – unsurprisingly – by almost all participants for the *Emergency* scenario. Also, the urgency of an agent's intervention appeared to be a key determinant for how (positively) it was perceived. One person said about the *Emergency*: "As long as someone's health is in danger, privacy would not be my priority" (P6). Similarly, regarding the *Health Risk*, 12 participants found the agent's intervention helpful as it is beneficial for the user's health: "I wouldn't mind [Jay] intruding in such cases. It's more important than me not wanting to be interrupted" (P6). For most participants, agent-initiated interactions that are time-critical but without dangerous consequences were still perceived as appropriate. Regarding the *Meeting Reminder*, one user said: "This is a good feature since [Jay] is making sure that the user won't be late for her meeting" (P3). Others concurred: "a good reason to interact" (P2). For two participants, emergency situations were the only acceptable instances for proactive interventions: "In other cases, it's just annoying" (P4). Participants also pointed out benefits for certain user groups: "This can really help with accessibility, especially for elderly and people with physical disabilities" (P10). One participant found the verbal support in the *Emergency* situation "especially helpful for children or the elderly. The system can also further instruct them" (P15).

Further, the proactive assistance for the *Technical Support* was perceived positively: "[Jay] was smart enough to understand the initial question was aimed at another person. After seeing that no solution can be found, it jumps in and helps" (P6). Reacting to indirect calls for assistance was also highlighted for the *Cooking Inspiration* scenario: "The character is mentioning that she has no clue, and she needs help" (P5) without addressing the VA. "The system was smart enough to detect a problem. It's not just answering a question, but rather trying to solve a problem it has detected" (P6). This was considered a meaningful "entry point" for the agent to proactively intervene. Speculating about the continuation of these two scenarios, all participants but two described that the proactive offer was accepted gladly by the users.

Overall, a common feeling observed during the interviews was the indecisiveness of participants to find proactivity helpful or not, when they found interactions intrusive but at the same time useful. About the *Disagreement Clarification*, one user said: "Very useful but very scary. It can destroy you but it will also cut the discussion short" (P8). In the speculated scenario continuation task, participants often thought the characters would feel violated, but still find the agent's intervention helpful, e.g., regarding *Health Risk*: "Even though she feels violated, she agrees to set an appointment" (P7). Similarly, for *Nudging*, "The user would get offended and say 'leave me alone!' But he would think about it and reflect on it later" (P6).

Interpretation. These results show that there are several situations in which users find the proactivity both useful and appropriate. However, a common pattern that we noticed was the dilemma of proactive interventions being perceived as helpful but at the same time disproportionately intrusive. We call this the proactivity dilemma. For several scenarios, participants were ambivalent about whether the intervention was overall desirable or not: a double-edged sword. This conflict of useful but invasive interventions, such as regarding health risks, is also visible in the quantitative results shown in Figure 3 (right) where one can clearly see that the relationship between both dimensions is not as uniform on the right graph (invasiveness-usefulness) as it is on the left (usefulness-appropriateness) and that the former also shows a somewhat wider spread.

Further, our results confirmed that urgency plays a big role in the appropriateness of proactive interventions, supporting the findings of previous works [7, 11, 28]. We observed throughout the study that the agent's proactive intervention was perceived as highly useful when users' health might be at risk. In such cases, people would not prioritise their privacy but were still concerned about insensitive interventions, aligning with previous research by Tabassum et al. [38]. Generally, the more serious and urgent the topic, the more useful and appropriate it was found to provide proactive assistance, e.g., when facing potential financial or professional damage. Proactively reminding users about their important upcoming activities or events was also highlighted as an appropriate and useful intervention. The familiarity of such interactions through existing digital services could be a reason for the acceptance of this form of proactive intervention.

4.2 Privacy and Mistrust

Even though we asked our participants to put aside any privacy and data protection concerns, they were the biggest worries among participants. One user mentioned: "I don't want the big companies to use all my data" (P4). A common demand amongst the interviewees was transparency and control in data processing: "If I know where my information is being processed and used, I can decide better to use such systems or not" (P12). Some participants were concerned about the misuse of personal data for hidden agendas or providing proactive advertisements: "[the agent] might give me suggestions that are influenced by political reasons or advertisements and try to control my behaviour based on that" (P10). Another concern was about an entity intruding into the private environment: "It's like another person is always at your home" (P12). They found it "really scary that everything could be monitored" (P8). These participants argued that people would constantly feel "observed"

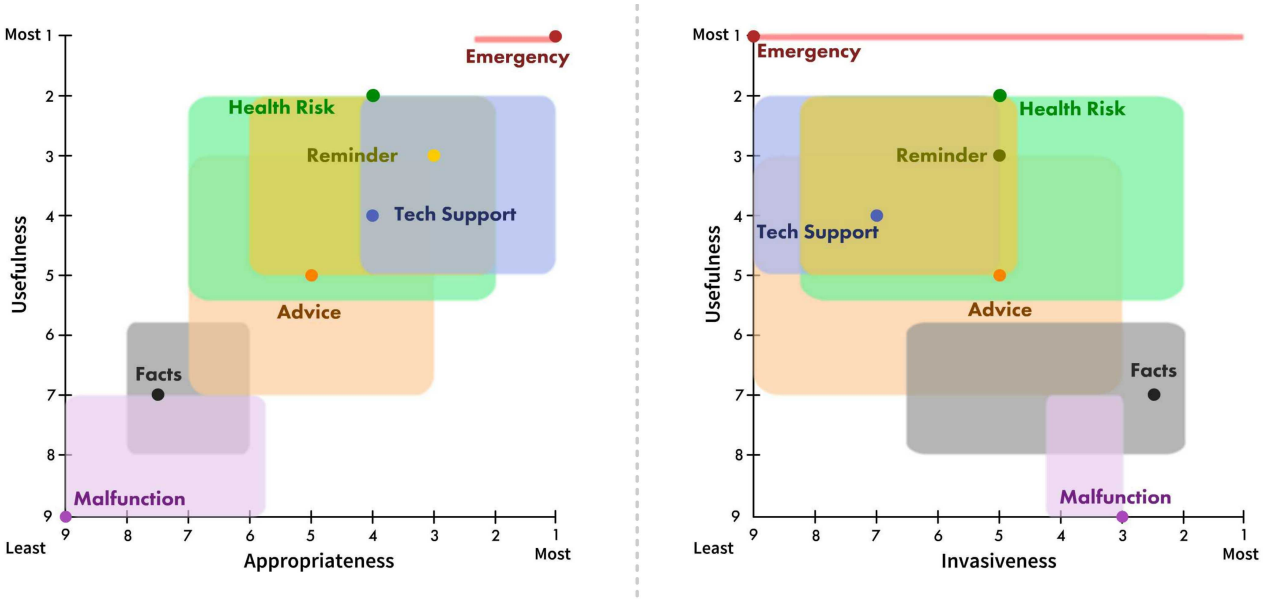


Figure 3: The medians and 75% ranges of the card sorting ranks of mapping usefulness to appropriateness (left) and to invasiveness (right) for different types of interactions: Malfunction (S8), Facts (S4, S5), Advice (S3, S6), Technical Support (S7), Reminder (S1), Health Risk (S2), Emergency (S9).

or "judged". This was especially prominent for scenarios where the agent interrupted a conversation. Mistrust was further expressed about "false alarms" and "misinterpretations" of certain situations and user states or behaviours by the agent which might create anxiety or cause stress in people: "If it's diagnosing you and it's wrong, it will create additional anxiety" (P9). Two even indicated mistrust about the reliability of the *Emergency* alarm.

Interpretation. In order for VAs to be proactive, they require more information about users' environment and behaviour, meaning more personal data needs to be processed to provide such services. During our sessions, it became evident that participants' main hurdle for adopting proactive VAs was the privacy aspect, in line with previous literature [23, 38]. Participants were worried about the misuse of their personal data by companies providing such assistants and third parties. Another concern was related to having an additional entity in the home that is not just a passive servant – like current VAs – but rather a character that aims at taking an active role in users' private space and family life. The participants associated these interferences with paternalism and a lack of control over the device, fearing negative social repercussions in multi-user settings. Therefore, to build trust and set boundaries, one approach could be that proactive VAs initially (e.g., in the first weeks of use) initiate proactive conversations in single-user contexts only.

4.3 Consideration of Social Context

Generally, participants were sceptical about the agent's social awareness. Seven participants found Jay's interventions disruptive and intrusive when they interfered with ongoing conversations:"[Jay] should not stop the thinking process and break conversations. It damages the human-human interaction" (P1). The proactive intervention was then considered "ruining the magic of the discussion" (P15). Two participants even perceived these interruptions as "creepy". The interjections were considered unwelcome because the agent was seen "as a tool rather than an equal conversation partner" (P7). With this unassailable interlocutor present, it felt to one participant "like a contract: everything is noted down. That's very stressful" (P15). The content of the conversation was described as an important factor for proactivity by seven participants: "If it is an intimate conversation, [Jay] should not really intervene" (P10). Two participants were concerned about the missed opportunity of socializing and bonding with another person due to the imposed help by the agent: "This is not received as an act of helping, but rather programmed" (P15). Further, the presence of people in the room was a common theme: "Emotional connection between me and my visitors is the key factor" (P3). In the presence of other people, 12 participants preferred the agent to be proactive only if it was an urgent matter.

Moreover, most participants found it frustrating or unpleasant when the agent corrected users: "People would feel bad about it. No one wants to be corrected" (P14). One person was torn as "this can be helpful, but it can hurt people's feelings" (P3). When the agent was contradicting one user while supporting another, participants found it even more insensitive. Regarding the *Disagreement Clarification*, verifying what was previously agreed was seen as the assistant was taking sides and seven people suspected dissatisfaction of one of the parties. They believed that such well-intended interventions "can potentially cause users to argue" (P13) and that "this could add more oil to the flame" (P1). About the *Fact Checking*, however, one participant assumed: "I think in this case, none of [the users] is correct, so the speaker was being helpful. If one of those people

was right, then the others would feel bad" (P14). Four participants speculated that the users in this scenario might feel offended, and three presumed that the proactive intervention would cause social awkwardness. In contrast, a small number of participants were in favour of these interventions, because "it's nice to be corrected" (P7) or "it's factual and cuts the discussion short" (P8). Similarly, two people appreciated the *Disagreement Clarification*: "I love this example. I think these arguments come up quite often and everyone thinks *they* are right. Personally, in this situation, I would like to have that. I always dreamed about having such a system to check for the truth" (P6).

Interpretation. In multi-user scenarios, the interventions in which the agent would help people to resolve an issue and save time were perceived positively. However, other than emergency situations, these were only perceived to be appropriate when the people had a chance to first try to resolve the matter by themselves. Participants generally thought that when the agent detected a question that was aimed at other people, responding to such questions before the intended person got a chance to respond was perceived as annoying and interfering. However, if the intended person could not properly respond to these questions or inquiries, the agent's intervention was considered useful and appropriate. For example, in the *Technical Support* scenario, the agent intervenes based on a request for help but only does so after the addressed person says they are not able to help at that point. Participants assumed that the agent was aware of the context and could appropriately detect an opportune moment to engage in the ongoing conversation. However, participants raised a concern about the agent taking away an opportunity of bonding, even if it is being helpful. They frequently mentioned that the agent's intervention in social situations is disruptive and could potentially damage human-human interaction. In accordance with previous research [27, 42], understanding the relationship between the people who are co-located, as well as the seriousness and intimacy of the conversation, were pointed out as important factors for the appropriateness of the agent's intervention in these situations.

Moreover, when the agent corrected people, some participants found it inappropriate, annoying, patronising or even insulting. The *Disagreement Clarification* scenario was rated most invasive and ranked second to last in terms of appropriateness. One reason for this was that in this scenario, the conversation was perceived as private. Additionally, the agent's intervention contradicts one of the people present and approves of the other, which resolves the disagreement but could further fuel the conflict. Nevertheless, some participants still found this highly useful and wished for such systems in their households, e.g., to cut discussions short. This example illustrates well that there seem to be major individual differences in how the proactive interventions are perceived.

4.4 Configuration and Control

A common desire amongst the participants was the ability to control and configure the system's proactive actions, in particular concerning the timing and topics. Three participants suggested the possibility to switch proactivity off temporarily. Four wanted to regulate interventions based on who is present in the room. Limiting proactive interventions at specific times of the day was suggested by three participants. One proposed to set the agent's proactivity extent using a slider in the settings. Hence, the users' agency was raised as a concern among participants. They found certain proactive interventions of Jay patronising and imperious. Participants did not like the assistant playing the role of someone who is controlling certain aspects of their lives: "I'm a person and I decide for my life. AI should not decide for me" (P4). This was particularly the case for the *Nudging* scenario. Eight participants explained that proactive features without prior approval would not be acceptable, in particular when the agent tries to nudge users towards a healthier behaviour: "If I have activated this in the settings, I would be more open to it. But if it is unasked for, I would be really annoyed" (P10). Without having asked for advice, a participant had the impression as if the agent "is judging you" (P13). Accordingly, ten participants expected users to ignore the intervention, seven said users would get frustrated, and two thought users would even disconnect the intrusive device. For one participant, the *Meeting Reminder* scenario was all about who is in control: "I feel like the system is forcing you to be productive and be a useful part of society. It takes my mind to dark places where people cannot control the system anymore. Autonomy is more important to me" (P7) Beyond customisation, participants also hoped for the system to automatically adjust over time. Whether manual or automatic, for one person "it needs to be adapted enough to the user's needs in order to understand when it's really needed – and when not" (P9).

Interpretation. Participants were concerned about their possible loss of agency. The feeling of being controlled and patronised by an agent was expressed as a worry. Similar to the findings by Luria et al. [23], several participants did not like it when the agent was suggesting healthier behaviour, i.e., avoiding extensive binge-watching. Based on our observations, the factors that would increase the chance of appropriateness for such interventions were the phrasing and the predictability of the interaction based on pre-configuration by the users. It was recommended for the agent's phrasing to be polite, calming, and suggestive rather than imposing. Correspondingly, participants wanted to have control over proactive interventions and be able to configure times and topics so that they could anticipate interactions to some extent and have more authority. To this end, such proactive VAs are ideally highly customisable and personalised based on individual user needs and preferences as suggested by previous research, such as regarding how short users want their VA's responses to be [14].

4.5 Initiating and Phrasing Interventions

How to introduce proactive interventions was a recurring theme during the interviews. For most of the interactions, participants suggested that the agent should ask for permission or give some kind of cue before speaking: "Maybe it is more acceptable if [Jay] says 'sorry to interrupt' " (P14). Some thought it would be a good compromise to first announce the subject without being too specific yet, such as: "I noticed something about your TV usage. Would you like me to share it?" In the proposed solutions, we identified three levels of initiation:

- *Non-Verbal Cues*: The agent indicates an intervention with a visual or auditory signal but then waits for the user's prompt to proceed.

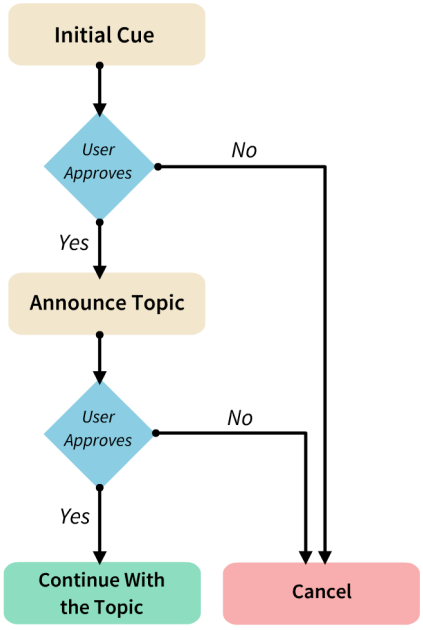


Figure 4: The figure shows the proposed initiation process model to proactively interact with people.

- *Verbal Cue*: The agent announces the subject but waits for the user's permission to proceed.
- *Direct Interventions*: The agent brings up the subject directly.

Direct interventions were mainly suggested for urgent or health-related scenarios but also for saving time. When interrupting conversations between people, non-verbal cues were preferred as they are the least distracting. In the interview task in which we asked participants to re-imagine the agent interventions to improve invasive or inappropriate scenarios, they either wanted the system to give a cue first or to be reactive. The intentionally misplaced *Fact Spoiler* was strongly criticised by the participants. Twelve people speculated that people would disconnect the device in such situations. Specifically, because the agent would not ask for the users' permission to speak, participants found it highly invasive and most inappropriate. One person was reminded of "the annoying kid in the class that screams the answer" (P10).

Further, when re-imagining scenarios, the wording was often adjusted. Concerning health-related issues, participants proposed phrasing the suggestion more cautiously: "Some people may perceive such news as shocking and get some other effects from it. It can create anxiety" (P11). Others did not want the agent to sound patronising or judgmental: "I have a tip for you regarding your health, do you want me to share it with you?" (P1). Instead of assuming a medical diagnosis and booking a doctor's appointment, two participants recommended asking clarifying questions, e.g., how they are feeling or if they have any other symptoms: "It is better if [the agent] gathers more information before making a conclusion and providing suggestions" (P13). Two participants indicated that, where possible, the agent should even help the user deal with stress, such as: "You don't have to worry, I can help you with that" (P8). Overall, the participants phrased their suggested initiations in a polite and calming manner, gently "building up" potentially distressing topics while keeping them goal-oriented and succinct.

Interpretation. From these insights, we can conclude that in most situations, participants expected the agent to ask for permission before conversing. This is in line with results by Arias et al. [4] who suggested that the agent should make sure the users are willing to interact at the specific moment. This permission request could be communicated in various forms. Verbal cues would have high conversational fidelity in relation to human conversations, such as addressing the user by name ("Excuse me, Alex?") or polite phrases ("May I interrupt?") as we suggested in our previous work [31]. A more subtle approach could be non-verbal cues of different modalities, such as abstract audio or light indicators. Depending on the ongoing activity, the preferences of our participants differed. The cue should not distract people from their activity unless it is an urgent matter requiring a striking cue. Verbal cues were described as most distracting, followed by audible cues. Visual cues were described as the least distracting. Based on our findings, we propose an initiation process model for VAs to proactively approach users in non-urgent situations, as illustrated in Figure 4. It starts with an *initial cue*, where the agent indicates that it would like to speak. After user approval, the agent moves on with introducing the *topic of intervention*. If that is also approved by the user, the agent can proceed with the action. In urgent cases, for less sensitive topics or in single-user settings, the second step could be skipped or combined with the first. Although this model could help make certain proactive behaviours more acceptable, the configuration of and control over the types of proactive behaviours as outlined in the previous section must always come first when designing such systems.

5 LIMITATIONS AND FUTURE WORK

In this research, we investigated proactive VA behaviours in a home environment as one of the most predominant use cases for VAs through a selection of storyboards depicting everyday situations. Although the broader insights of this evaluation can be applied to other settings, future work should investigate proactive VAs in other environments such as work and public environments. Moreover, the sample was skewed toward young ($M = 27.86$) and on average, more educated users, and therefore, may not be fully representative of possible VA users. This is particularly relevant when considering that in the scenarios, users with various demographics were present (e.g., the elderly person in the *Health Risk* scenario). Future studies could validate our findings by investigating a wider population and the specific user groups that certain proactive features may be designed for. In our study, we witnessed that individual personal differences can also be a decisive factor in terms of finding proactive VAs appropriate. Differences in user traits (e.g., personality) may lead to different preferences on proactive VAs, which should be incorporated and reflected in further studies.

The method applied in this investigation has its limitations and advantages. Since proactive VAs that have comparable capabilities to those illustrated in our storyboards are not yet available in the

market, we explored people's opinions on these features by presenting hypothetical scenarios. We conducted interactive interviews involving various tasks on a digital whiteboard that engaged participants to contemplate about the presented design space from different angles. As our method requires the participants to immerse and speculate, it enables the investigation of interactions with future technologies that would be intricate or expensive to build. It further enables evaluating aspects of the system that would be impossible to simulate realistically, such as emergency situations or delicate private settings. However, since participants did not experience the situations and proactive behaviours themselves, their perceptions may not reflect real-world experience. Furthermore, it is important to note that some of the services presented in the storyboards may also be supported by other technologies and not solely VAs. In this study, we sought to explore what needs to be considered when developing such features for Voice Assistants.

6 CONCLUSION

This research explores desirable circumstances for proactive interventions by VAs in domestic settings. The findings of our scenario-based study show that people see great benefit in proactivity, specifically in cases of important reminders, time-saving interventions, or emergency support. However, great concerns such as privacy implications, potential loss of agency, and interference with social activities may inhibit the adoption of such systems. Based on the interpretation of our results, we believe that the desirability of proactive interactions depends on the following factors.

Significance. The more urgent or critical the topic, the more appropriate it is for a VA to proactively intervene. The desirability is high under circumstances with a large scope or grave consequences.

Social Context and Environment. Proactive VAs should accurately identify the environmental and social context including the presence of other users or guests, the closeness of their relationships, the type and sensitivity of the ongoing activity, and the time of the day.

Agency and Control. The user needs to be able to adjust and configure proactive features including the times and topics for interventions. Users should have control over when the agent is allowed to listen and observe its environment, and when it is allowed to intervene so that they could anticipate such interactions.

Individual User Factors. As there seem to be major differences between individuals in how certain interventions are perceived, proactive VAs should be able to consider individual user factors such as physical and cognitive abilities (e.g., of young children or elderly users), the current physical and emotional state (e.g., stress level, sadness, or fatigue), and the personality and preferences of the user (e.g., privacy needs or agency).

Form of Execution. When initiating an interaction, the agent should generally first request permission using verbal or non-verbal cues, and announce the topic of intervention – unless it is time-critical as in an emergency. Furthermore, the intent should be phrased so that it is polite, not imposing, and does not create a feeling of unease, while at the same time being goal-oriented and concise. When users got used to certain interventions over time or gave permission, the VA may get right to the point.

Altogether, as long as the proactivity dilemma is carefully considered by finding a positive balance with suggestions that are perceived as more helpful than invasive, there seems to be great potential in proactive VAs.

ACKNOWLEDGMENTS

This work was partially funded by the Leverhulme Trust as part of the Doctoral Training Programme for the Ecological Study of the Brain (DS-2017-026) and the Klaus Tschira Foundation.

REFERENCES

[1] Herman Aguinis and Kyle J. Bradley. 2014. Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies. *Organizational Research Methods* 17, 4 (2014), 351–371. https://doi.org/10.1177/1094428114547952 arXiv:https://doi.org/10.1177/1094428114547952

[2] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300233

[3] Salvatore Andolina, Valeria Orso, Hendrik Schneider, Khalil Klouche, Tuukka Ruotsalo, Luciano Gamberini, and Giulio Jacucci. 2018. Investigating Proactive Search Support in Conversations. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong, China) *(DIS '18)*. Association for Computing Machinery, New York, NY, USA, 1295–1307. https://doi.org/10.1145/3196709.3196734

[4] Kika Arias, Sooyeon Jeong, Hae Won Park, and Cynthia Breazeal. 2020. Toward Designing User-centered Idle Behaviors for Social Robots in the Home.

[5] Michael Braun, Anja Mainz, Ronee Chadowitz, Bastian Pfleging, and Florian Alt. 2019. At your service: Designing voice assistant personalities to improve automotive user interfaces. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–11.

[6] John M. Carroll. 1999. Five Reasons for Scenario-Based Design. In *Proceedings of the Thirty-Second Annual Hawaii International Conference on System Sciences-Volume 3 - Volume 3 (HICSS '99)*. IEEE Computer Society, USA, 3051.

[7] Narae Cha, Auk Kim, Cheul Young Park, Soowon Kang, Mingyu Park, Jae-Gil Lee, Sangsu Lee, and Uichin Lee. 2020. Hello There! Is Now a Good Time to Talk? Opportune Moments for Proactive Interactions with Smart Speakers. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 3, Article 74 (Sept. 2020), 28 pages. https://doi.org/10.1145/3411810

[8] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, Vincent Wade, and Benjamin R. Cowan. 2019. What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300705

[9] Benjamin R. Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. "What Can I Help You with?": Infrequent Users' Experiences of Intelligent Personal Assistants. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vienna, Austria) *(MobileHCI '17)*. ACM, New York, NY, USA, Article 43, 12 pages. https://doi.org/10.1145/3098279.3098539

[10] Philip R. Doyle, Justin Edwards, Odile Dumbleton, Leigh Clark, and Benjamin R. Cowan. 2019. Mapping Perceptions of Humanness in Intelligent Personal Assistant Interaction. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services* (Taipei, Taiwan) *(MobileHCI '19)*. ACM, New York, NY, USA, Article 5, 12 pages. https://doi.org/10.1145/3338286.3340116

[11] Justin Edwards, Christian Janssen, Sandy Gould, and Benjamin R. Cowan. 2021. Eliciting Spoken Interruptions to Inform Proactive Speech Agent Design. In *CUI 2021 - 3rd Conference on Conversational User Interfaces* (Bilbao (online), Spain) *(CUI '21)*. Association for Computing Machinery, New York, NY, USA, Article 23, 12 pages. https://doi.org/10.1145/3469595.3469618

[12] Emer Gilmartin, Brendan Spillane, Maria O'Reilly, Ketong Su, Christian Saam, Benjamin R. Cowan, Nick Campbell, and Vincent Wade. 2017. Dialog Acts in Greeting and Leavetaking in Social Talk. In *Proceedings of the 1st ACM SIGCHI International Workshop on Investigating Social Interactions with Artificial Agents* (Glasgow, UK) *(ISIAA 2017)*. ACM, New York, NY, USA, 29–30. https://doi.org/10.1145/3139491.3139493

[13] Greg Guest, Arwen Bunce, and Laura Johnson. 2006. How many interviews are enough? An experiment with data saturation and variability. *Field methods* 18, 1 (2006), 59–82.

[14] Gabriel Haas, Michael Rietzler, Matt Jones, and Enrico Rukzio. 2022. Keep It Short: A Comparison of Voice Assistants' Response Behavior. In *CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 321, 12 pages. https://doi.org/10.1145/3491102.3517684

[15] Scott Hudson, James Fogarty, Christopher Atkeson, Daniel Avrahami, Jodi Forlizzi, Sara Kiesler, Johnny Lee, and Jie Yang. 2003. Predicting Human Interruptibility with Sensors: A Wizard of Oz Feasibility Study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. Association for Computing Machinery, New York, NY, USA, 257–264. https://doi.org/10.1145/642611.642657

[16] Soowon Kang, Heepyung Kim, Youngtae Noh, and Uichin Lee. 2021. *Poster: Toward Context-Aware Proactive Conversation for Smart Speakers*. Association for Computing Machinery, New York, NY, USA, 38–40. https://doi.org/10.1145/3460418.3479306

[17] Auk Kim, Woohyeok Choi, Jungmi Park, Kyeyoon Kim, and Uichin Lee. 2018. Interrupting Drivers for Interactions: Predicting Opportune Moments for In-Vehicle Proactive Auditory-Verbal Tasks. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 4, Article 175 (Dec. 2018), 28 pages. https://doi.org/10.1145/3287053

[18] Auk Kim, Jung-Mi Park, and Uichin Lee. 2020. Interruptibility for in-vehicle multitasking: influence of voice task demands and adaptive behaviors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–22.

[19] Mitsuki Komori, Yuichiro Fujimoto, Jianfeng Xu, Kazuyuki Tasaka, Hiromasa Yanagihara, and Kinya Fujita. 2019. Experimental Study on Estimation of Opportune Moments for Proactive Voice Information Service Based on Activity Transition for People Living Alone. In *Human-Computer Interaction. Perspectives on Design*, Masaaki Kurosu (Ed.). Springer International Publishing, Cham, 527–539.

[20] Matthias Kraus, Marvin Schiller, Gregor Behnke, Pascal Bercher, Michael Dorna, Michael Dambier, Birte Glimm, Susanne Biundo, and Wolfgang Minker. 2020. "Was That Successful?" On Integrating Proactive Meta-Dialogue in a DIY-Assistant Using Multimodal Cues. In *Proceedings of the 2020 International Conference on Multimodal Interaction* (Virtual Event, Netherlands) *(ICMI '20)*. Association for Computing Machinery, New York, NY, USA, 585–594. https://doi.org/10.1145/3382507.3418818

[21] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. 2018. Alexa, Are You Listening? Privacy Perceptions, Concerns and Privacy-Seeking Behaviors with Smart Speakers. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 102 (Nov. 2018), 31 pages. https://doi.org/10.1145/3274371

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

[23] Michal Luria, Rebecca Zheng, Bennett Huffman, Shuangni Huang, John Zimmerman, and Jodi Forlizzi. 2020. *Social Boundaries for Personal Agents in the Interpersonal Space of the Home*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376311

[24] Nathan Malkin, Joe Deatrick, Allen Tong, Primal Wijesekera, Serge Egelman, and David Wagner. 2019. Privacy attitudes of smart speaker users. *Proceedings on Privacy Enhancing Technologies* 2019, 4 (2019), 250–271.

[25] Donald McMillan. 2017. Implicit Interaction Through Machine Learning: Challenges in Design, Accountability, and Privacy. In *Internet Science*, Ioannis Kompatsiaris, Jonathan Cave, Anna Satsiou, Georg Carle, Antonella Passani, Efstratios Kontopoulos, Sotiris Diplaris, and Donald McMillan (Eds.). Springer International Publishing, Cham, 352–358.

[26] Donald McMillan, Antoine Loriette, and Barry Brown. 2015. Repurposing Conversation: Experiments with the Continuous Speech Stream. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3953–3962. https://doi.org/10.1145/2702123.2702532

[27] O. Miksik, I. Munasinghe, J. Asensio-Cubero, S. Reddy Bethi, S-T. Huang, S. Zylfo, X. Liu, T. Nica, A. Mitrocsak, S. Mezza, R. Beard, R. Shi, R. Ng, P. Mediano, Z. Fountas, S-H. Lee, J. Medvesek, H. Zhuang, Y. Rogers, and P. Swietojanski. 2020. Building Proactive Voice Assistants: When and How (not) to Interact. arXiv:2005.01322 [cs.HC]

[28] Florian Nothdurft, Stefan Ultes, and Wolfgang Minker. 2015. Finding appropriate interaction strategies for proactive dialogue systems—an open quest. , 73–80 pages.

[29] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal, QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3174214

[30] Aung Pyae and Tapani N. Joelsson. 2018. Investigating the Usability and User Experiences of Voice User Interface: A Case of Google Home Smart Speaker. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (Barcelona, Spain) *(MobileHCI '18)*. Association for Computing Machinery, New York, NY, USA, 127–131. https://doi.org/10.1145/3236112.3236130

[31] Leon Reicherts, Nima Zargham, Michael Bonfert, Yvonne Rogers, and Rainer Malaka. 2021. May I Interrupt? Diverging Opinions On Proactive Smart Speakers. In *CUI 2021 - 3rd Conference on Conversational User Interfaces* (Bilbao (online), Spain) *(CUI '21)*. Association for Computing Machinery, New York, NY, USA, Article 34, 10 pages. https://doi.org/10.1145/3469595.3469629

[32] A Joy Rivera. 2014. A socio-technical systems approach to studying interruptions: Understanding the interrupter's perspective. *Applied ergonomics* 45, 3 (2014), 747–756.

[33] Maria Schmidt and Patricia Braunger. 2018. A Survey on Different Means of Personalized Dialog Output for an Adaptive Personal Assistant. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization* (Singapore, Singapore) *(UMAP '18)*. Association for Computing Machinery, New York, NY, USA, 75–81. https://doi.org/10.1145/3213586.3226198

[34] Maria Schmidt, Wolfgang Minker, and Steffen Werner. 2020. User Acceptance of Proactive Voice Assistant Behavior. In *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2020*, Andreas Wendemuth, Ronald Böck, and Ingo Siegert (Eds.). TUDpress, Dresden, Dresden, Germany, 18–25.

[35] Maria Schmidt, Daniela Stier, Steffen Werner, and Wolfgang Minker. 2019. Exploration and assessment of proactive use cases for an in-car voice assistant. In *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2019*, Peter Birkholz and Simon Stone (Eds.). TUDpress, Dresden, Dresden, Germany, 148–155.

[36] Rob Semmens, Nikolas Martelaro, Pushyami Kaveti, Simon Stent, and Wendy Ju. 2019. Is Now A Good Time? An Empirical Study of Vehicle-Driver Communication Timing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300867

[37] Petra-Maria Strauss and Wolfgang Minker. 2010. *Proactive spoken dialogue interaction in multi-party environments*. Springer, Cham.

[38] Madiha Tabassum, Tomasz Kosiński, Alisa Frik, Nathan Malkin, Primal Wijesekera, Serge Egelman, and Heather Richter Lipford. 2019. Investigating Users' Preferences and Expectations for Always-Listening Voice Assistants. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4 (2019), 1–23.

[39] Liam D. Turner, Stuart M. Allen, and Roger M. Whitaker. 2015. Interruptibility Prediction for Ubiquitous Systems: Conventions and New Directions from a Growing Field. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 801–812. https://doi.org/10.1145/2750858.2807514

[40] Konstantina Vasileiou, Julie Barnett, Susan Thorpe, and Terry Young. 2018. Characterising and justifying sample size sufficiency in interview-based studies: systematic analysis of qualitative health research over a 15-year period. *BMC Medical Research Methodology* 18, 1 (Nov. 2018), 148. https://doi.org/10.1186/s12874-018-0594-7

[41] Sarah Theres Völkel, Daniel Buschek, Malin Eiband, Benjamin R. Cowan, and Heinrich Hussmann. 2021. Eliciting and Analysing Users' Envisioned Dialogues with Perfect Voice Assistants. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445536

[42] Jing Wei, Tilman Dingler, and Vassilis Kostakos. 2021. Developing the Proactive Speaker Prototype Based on Google Home. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–6.

[43] Jing Wei, Tilman Dingler, and Vassilis Kostakos. 2022. Understanding User Perceptions of Proactive Smart Speakers. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 4, Article 185 (dec 2022), 28 pages. https://doi.org/10.1145/3494965

A APPENDIX

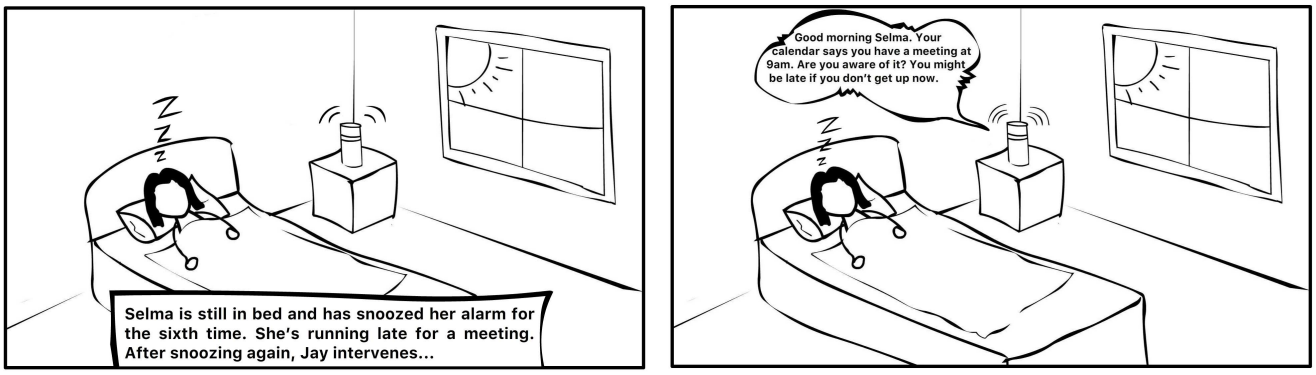


Figure 5: Scenario 1: Meeting Reminder

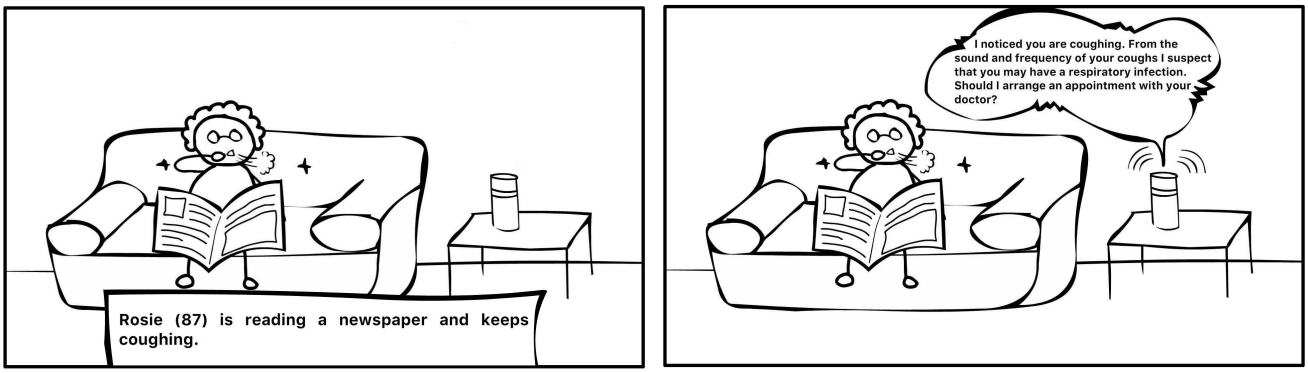


Figure 6: Scenario 2: Health Risk

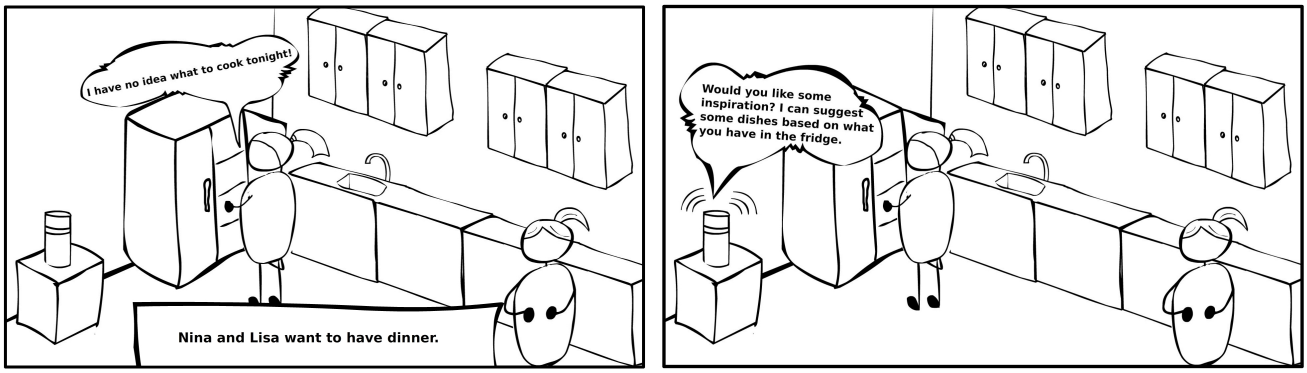


Figure 7: Scenario 3: Cooking Inspiration

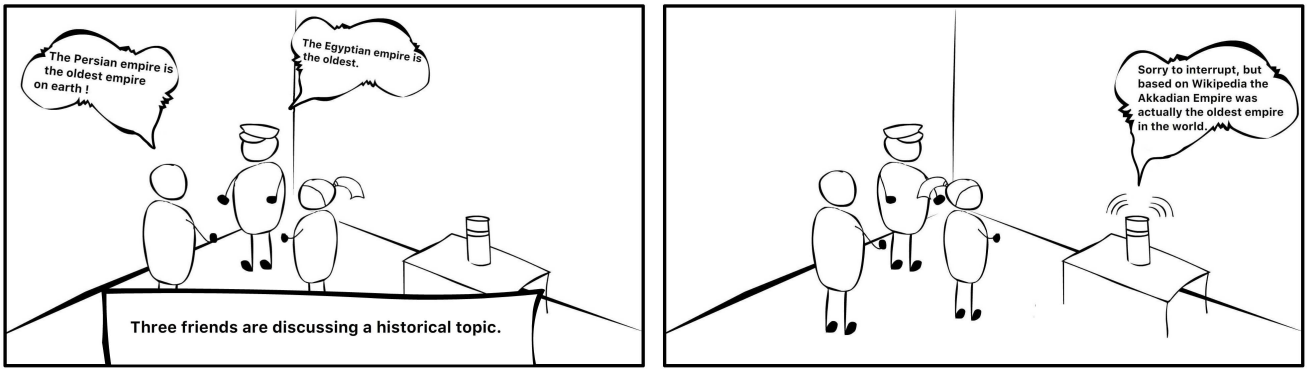


Figure 8: Scenario 4: Fact Checking

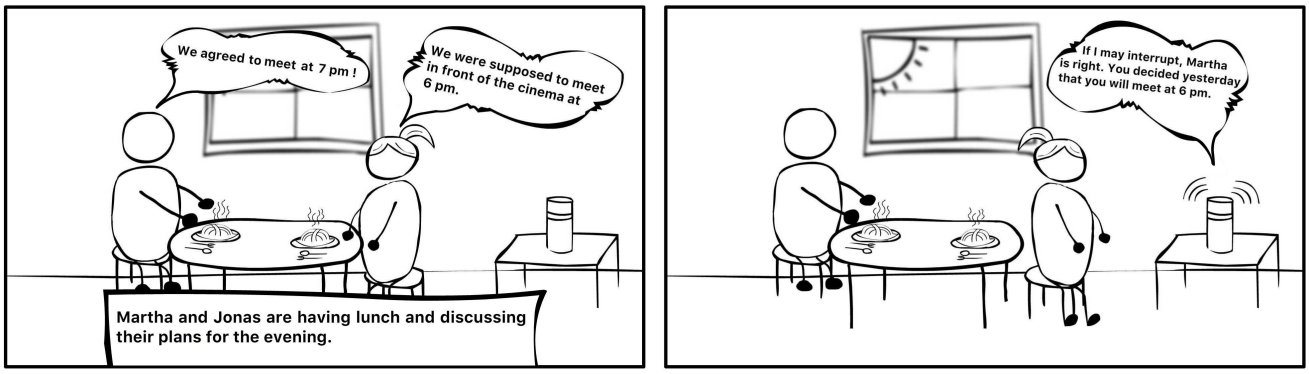


Figure 9: Scenario 5: Disagreement Clarification

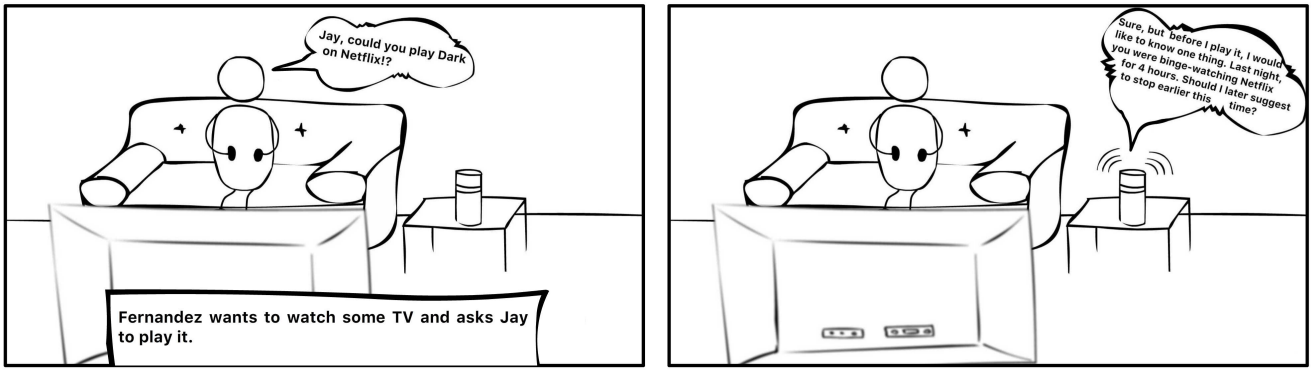


Figure 10: Scenario 6: Nudging

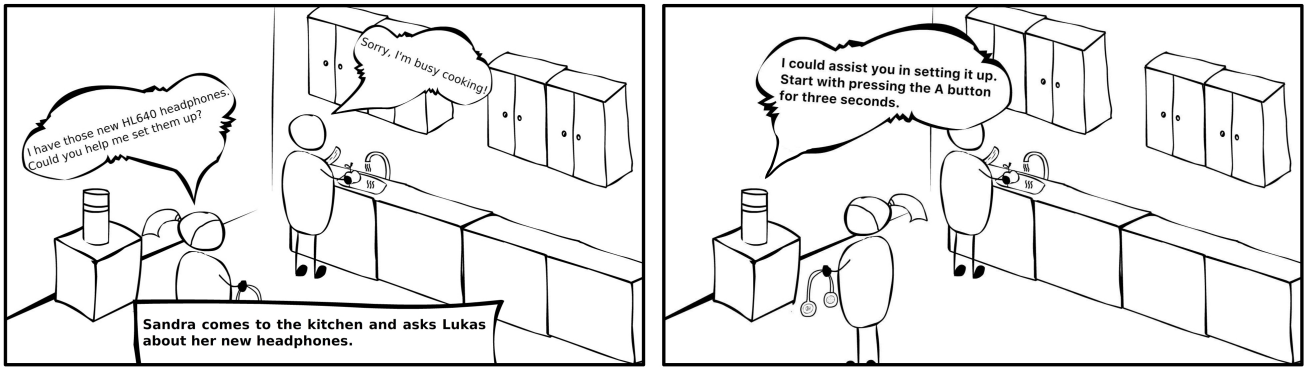


Figure 11: Scenario 7: Technical Support

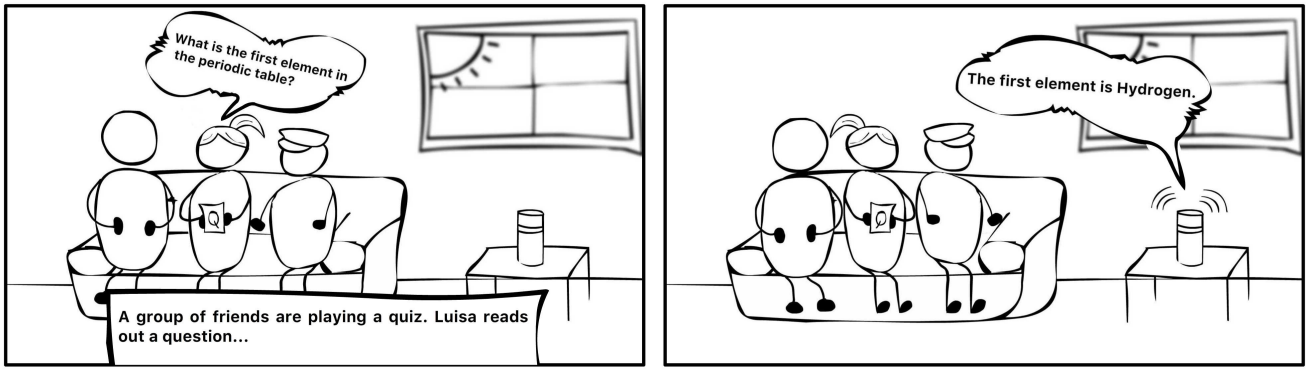


Figure 12: Scenario 8: Fact Spoiler

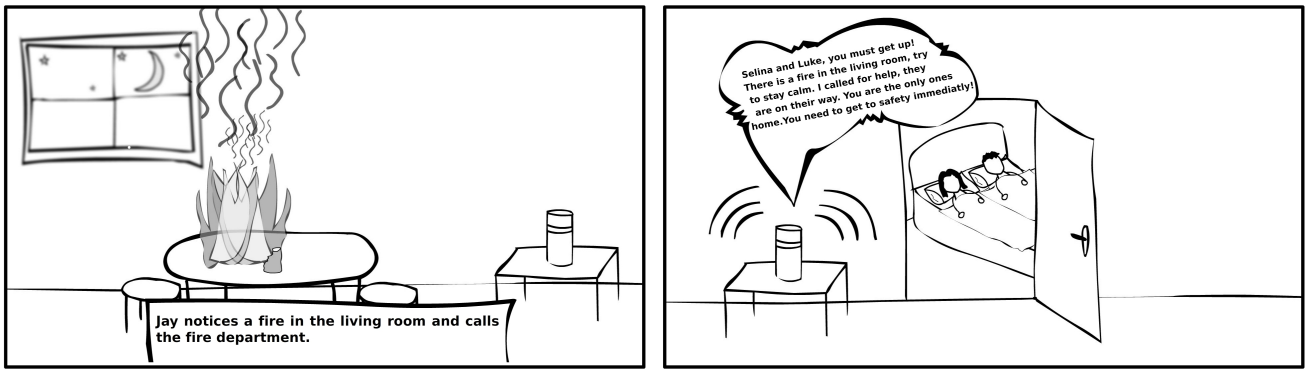


Figure 13: Scenario 9: Emergency

View publication stats