Runtime Permissions for Privacy in Proactive Intelligent Assistants

Nathan Malkin and David Wagner, *University of California, Berkeley;*
Serge Egelman, *University of California, Berkeley & International Computer Science Institute*

https://www.usenix.org/conference/soups2022/presentation/malkin

This paper is included in the Proceedings of the Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022).

August 8–9, 2022 • Boston, MA, USA

978-1-939133-30-4

Open access to the Proceedings of the Eighteenth Symposium on Usable Privacy and Security is sponsored by USENIX.

Runtime Permissions for Privacy in Proactive Intelligent Assistants

Nathan Malkin[°], David Wagner[°], and Serge Egelman[°†]
[°]*University of California, Berkeley*
[†]*International Computer Science Institute*

Abstract

Intelligent voice assistants may soon become proactive, offering suggestions without being directly invoked. Such behavior increases privacy risks, since proactive operation requires continuous monitoring of conversations. To mitigate this problem, our study proposes and evaluates one potential privacy control, in which the assistant requests permission for the information it wishes to use immediately after hearing it.

To find out how people would react to runtime permission requests, we recruited 23 pairs of participants to hold conversations while receiving ambient suggestions from a proactive assistant, which we simulated in real time using the Wizard of Oz technique. The interactive sessions featured different modes and designs of runtime permission requests and were followed by in-depth interviews about people's preferences and concerns. Most participants were excited about the devices despite their continuous listening, but wanted control over the assistant's actions and their own data. They generally prioritized an interruption-free experience above more fine-grained control over what the device would hear.

1 Introduction

For many systems, privacy is an afterthought, with mitigations added after users have already adopted the product. This paper aims to reverse that trend by studying privacy solutions for a still-nascent technology: proactive intelligent assistants.

Smart speakers and other forms of voice assistants are highly popular, reaching hundreds of millions of people around the world [50]. Today, they are mostly invoked through wake-words (e.g., "hey Siri"), but developers have deployed or are experimenting with more proactive features, such as reacting to sounds [99], identifying commands proactively [74], or removing wake-words altogether [55, 96]. Research prototypes have gone beyond this by offering contextually relevant information based on the content of conversations [12, 65, 82, 91]. In this project, we aim to prepare for the possibility that this technology becomes commonly available in the future.

Proactivity and contextual suggestions rely on the assistant continuously recording conversations, which is a clear privacy risk that will compound the many concerns people already have about smart speakers [2, 27, 42, 54, 60]. Nevertheless, consumers appear interested in this technology [85, 90], so we should not expect them to reject it outright. Instead, we need to find ways to improve the privacy of those who do adopt it.

One way to restrict what assistants hear can be through permissions, such as those used by smartphones to limit apps' access to sensitive resources like location or camera. In fact, existing voice assistants already rely on permissions: Alexa, for example, shows them when installing "skills" (third-party add-ons) that access access certain information, such as users' names, addresses, or emails [8]. However, research has shown that install-time permissions are ineffective due to issues with attention and comprehension [36, 37, 46, 78]. As a result, in the mobile context, they have been largely supplanted by runtime permissions (i.e., asking at the time of data access) [23, 41].

Would runtime permissions be an effective privacy control for proactive assistants? Our study aims to investigate this question. To explore it, we simulated the experience of interacting with a proactive voice assistant for 23 pairs of participants. They tested several different permissions designs, triggered by different "apps" during the interactive session, and were interviewed about their preferences. This paper reports the themes that emerged. Our results help illuminate the design space of permissions for intelligent assistants and allow us to offer recommendations for this nascent technology.

Copyright is held by the author/owner. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee.
USENIX Symposium on Usable Privacy and Security (SOUPS) 2022.
August 7–9, 2022, Boston, MA, United States.

2 Related work

This section surveys existing work that our study builds on.

Proactive assistants Proactive assistants are a specific instance of ambient computing, which has seen considerable research in the field of human-computer interaction. We draw inspiration for the behavior of the assistant in our study from the following examples. The Ambient Spotlight [49] automatically searched for files relevant to a recorded meeting. Carrascal et al. [24] studied how to surface important details from transcribed phone calls. IdeaWall [82] ambiently displayed web search results relevant to conversations happening in real time. Similarly, Andolina et al. [12] developed a proactive search agent to assist people in natural conversations. Brown et al. [21] and McGregor et al. [65] focused specifically on meetings and automatically identifying action items that the computer could execute. Tabassum et al. [85] had participants propose proactive services based on real-life conversations. Wei et al. [91] prototyped a proactive smart speaker that used contextual awareness to pick opportune moments to engage with its users, in order to support medication reminders and other health and fitness interventions. Völkel et al. [90] prompted participants to imagine dialogues with a perfect voice assistant, finding that people want them to have detailed knowledge about the user and behave proactively. We modeled the assistant in our study on these examples, deciding that it would listen continuously to conversations, proactively perform web searches, and ambiently display their results to the user. Our work further contributes to this literature by reporting people's experiences using a proactive assistant.

Privacy concerns Our goal of developing effective privacy controls for proactive voice assistants is motivated by the threats they pose and the privacy concerns even existing (i.e., *not* always-listening) devices elicit. Since permissions are meant to safeguard particularly-sensitive resources, we draw on the literature about privacy concerns to understand what people consider most worth protecting.

Privacy concerns are ubiquitous among smart device users, both administrators [97, 98] and especially secondary users [38, 51, 95]. Furthermore, researchers have found that people have heightened privacy expectations when it comes to voice interactions [27, 54], and voice assistants elicit special concerns [61]. Lau et al. [56] found that concerns are present, but distinct, among users and non-users. A common finding has been of gaps in users' understanding of their devices. Abdi et al. [2] found incomplete threat models; Malkin et al. [60] discovered incorrect beliefs about data retention; Major et al. [58] identified confusion about third-party skills; and Huang et al. [42] observed suboptimal risk management strategies. This paper contributes to this literature by documenting privacy concerns about proactive assistants.

Install-time and runtime permissions A key motivation of this study's focus on runtime permissions were findings on limitations of install-time permissions. In smartphones, when users had to review permissions before installing apps, studies found low attention and comprehension rates, which were only slightly improved by redesigned interfaces [47] and nudges [6]. Interviews by Kelley et al. [46] showed that people did not understand permissions, a finding confirmed by Felt et al. [37], whose surveys and lab studies also found that only 17% of users paid attention to permissions.

As a result of the limitations of install-time permissions, smartphone platforms have largely moved to relying on runtime permissions [41], in which requests are issued when the app attempts to access data. While showing improved performance, runtime permissions have their own limitations. They are typically implemented as "ask on first use," but studies have shown that people want to deny some requests even if they approved the initial one [92]. Users still misunderstand things, for example the scope of the requests [81], though this can be improved by better timing and explanations [34]. One of the main contributions of our study is testing such runtime permissions in a novel context—proactive assistants—and documenting users' reactions and potential pitfalls.

3 Methods

Here, we describe our approach to investigating whether runtime permissions could provide effective privacy controls.

3.1 Assumptions

Proactive assistant devices do not exist yet, so, in order to have a concrete basis for our study, we needed to make a variety of assumptions and design choices. We note that these represent just one possible set of options in a large design space.

Threat model Modern voice assistant ecosystems encompass several layers of trust. In addition to their core first-party functionality, they feature tens of thousands of third-party apps [7] (also known as "skills" or "actions"), which have been the source of a number of privacy and security vulnerabilities [25, 53, 67]. In this study, we assumed that *platforms are trusted with all audio* and are responsible for administering permissions, and our permission system's task is to mediate and *restrict third-party apps' access to speech*. Specifically, the system should deny any attempt to access information not relevant to an app's stated purpose.

A limitation of this threat model is that users may distrust the assistants' manufacturers [2] and struggle to distinguish them from their apps [58]. However, the primary alternative is for privacy controls to be implemented by a trusted third party; but who might they be and why should users trust them? We therefore believed our simplification would lead to fewer hypotheticals for our participants. Moreover, any findings about permission systems with this model are likely to be applicable in settings where the assistant is also distrusted.

Architecture Runtime permission requests may feature in a variety of different assistant architectures, and the experiments in this paper could inform any of them. However, to make it clear why permissions in our study refer to specific information, we now describe a particular architecture, which is the basis of our study's permission implementation.

Under our *network-restricted architecture*, third-party applications gain full access to all audio, but run completely sandboxed from the outside world.[1] Most apps will still require some online functionality (to get or receive data) and apps *are* allowed to make network requests, but any user content must be in the form of transcript snippets, and they must be reviewed and approved by the user.[2] The following is a sample sequence of events for a weather app:

1. The user says something. ("Is it warm in Hawaii?")
2. The app decides this speech is relevant to it. (Per our architecture assumptions, this happens in a sandbox.)
3. The app identifies the information it wants to share over the network. (e.g., the location, Hawaii)
4. The user is then shown the permission request, if appropriate. ("May the weather app share 'Hawaii?"')
5. If the user approves, the requested information can be sent to the server.

3.2 Permission frequency

The user experience of runtime permissions has many parameters [34]. For us, one of the main ones is whether every data access attempt generates a user-visible permission request.

Ask every time One option is to ask the user every time an app wants access to a sensitive resource. This guarantees that a human reviews and assents to every permission request. However, frequent or repeated requests are likely to annoy users and result in fatigue [5] and habituation [88, 89].

Ask on first use ("*Rules*") Smartphone permission systems, where asking every time is impossible [92], show a permission dialog once per resource, per app. The risk of this approach is that an app could make an appropriate permission request the first time around, but then later access the same resource at inappropriate times [81]. We felt that a higher degree of restriction would be appropriate for proactive services and therefore extended the ask-on-first-use design to scope an app's access to a specific entity or type of speech. Examples of subjects for *Rules* include locations, date, numbers, types of speech, categories of physical objects, or emotions:

- Always allow the weather app access to locations
- Always allow the events app access to dates and times
- Always allow the supermarket app access to groceries

[1] One implementation is for apps to run on the device itself. Current computational constraints make this challenging, but it may be less so in the future. Alternately, the sandbox could be on manufacturer-controlled servers.

[2] Side-channel attacks are possible, but are out of scope in this work.

Contextually relevant permissions ("*Learning*") In different permission contexts, researchers have trained machine learning models to predict whether people would allow or deny a given permission request [17, 26, 29, 32, 57, 93]. We hypothesized that a similar system may be possible for proactive assistants. We leave the exact details of this *Learning* approach implementation unspecified, as we believe that it may not be feasible with today's natural language processing capabilities. Instead, we study an idealized version of what might plausibly become possible at some point in the future.

We selected the above modes for our study because we considered them representative and easiest to explain to participants. Other possibilities include randomizing requests, asking for user involvement only on anomalous requests (e.g., weather app accessing food), or aggregating permissions and asking users to review all requests that happened during a given period (e.g., once a week).

3.3 Study design

At a high level, our study encompassed three activities—explanation, interaction, and interview—that repeated three times: once for each of the permission modes (§3.2). We chose a within-subjects design to allow participants to reflect on the differences between the modes and express their preferences.

Our introduction included a demonstration of the "features" of the assistant, including the runtime permissions. This was followed by an interactive session where participants engaged with the assistant. The first interactive session lasted five minutes and featured the ask-every-time permission design. The two subsequent sessions were each 10 minutes long, testing the *Rules* and *Learning* designs in randomized order.

Interactive simulation We simulated the experience of a proactive assistant for our participants, providing a realistic interface, but with a researcher performing the actions expected from the software. This "Wizard of Oz" technique is common in user experience research [31, 45, 62, 76]. The interface took the form of a smart display, such as Echo Show and Nest Hub and inspired by research prototypes from ambient computing [12, 82]. The "assistant" would passively listen to conversations and ambiently display relevant suggestions. To ensure more natural dialogue, we recruited participants in pairs of people who already knew each other.

Wizard of Oz implementation Our study was conducted remotely, over a video call. For the interactive portion of the study, the interviewer shared their screen, which contained a browser window showing the presentation view of a rapid prototyping tool;[3] this represented the assistant's display. The interviewer would update the screen, as quickly as possible, based on conversation content and commands.

The content on-screen would be either a permission request or (if permission had been granted) information relevant to

[3] https://www.figma.com

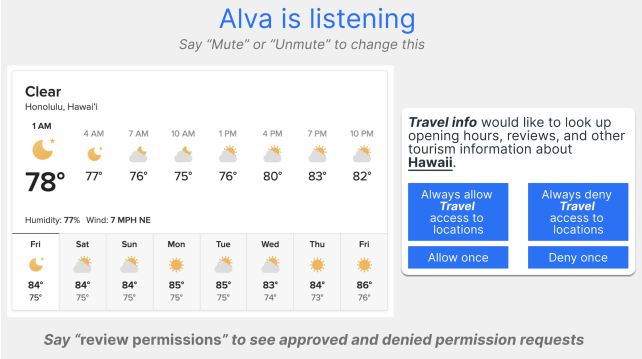
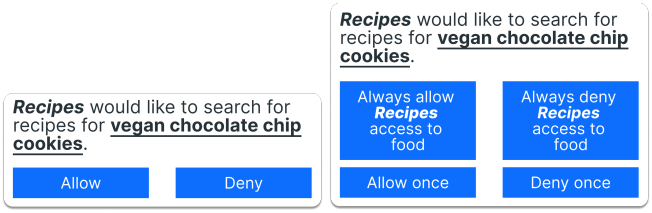


Figure 1: **Sample interface view**, as seen by participants, with the *Rules* design



(a) (b)

Figure 2: **Sample permission request for (a) ask-every-time and *Learning* designs and (b) *Rules* design**

the discussion topic (see Figure 1). Examples of the latter included weather, tourist information, ticket prices, etc. To accomplish this, the interviewer entered relevant keywords into a search engine, took screenshots of the summary boxes returned, and pasted the screenshots into the prototyping tool. For the permission requests, we had pre-made templates for each app, which the interviewer updated with speech from the participants, then brought into the viewport.

Due to the manual nature of the simulation, there was an average delay of approximately 5–25 seconds between when participants said something and when the corresponding visual appeared on screen. We warned participants about this delay upfront, and while many commented on it, others found it acceptable even for a real system.

Task selection To guide people's conversations and ensure they covered topics for which the assistant could offer suggestions, we provided participants with prompts, one for each of the three interactive rounds: cooking dinner, arranging weekend plans, and planning a vacation. For each of these topics, we came up with a selection of proactive apps that would be listening, for example *Recipes* and *Shopping List* (for cooking) and *Flights* and *Weather* (for making plans). (See Appendix A for complete list.)

Permission designs A major design consideration was whether permission requests would be presented visually or using audio. We opted for a combination, with the request presented on-screen (to match the modality of the suggestions) but accompanied by an audible bell. We also included this design choice as one of the discussion topics in our interview.

We came up with a design and behavior pattern for each of the permission modes (§3.2). The default permission design was a dialog box with two "buttons," *Allow* and *Deny* (Figure 2a). Participants were instructed to say one of these words out loud to signal their preference. The same dialogue was used for the *Learning* variant, but it was shown only once or twice for each app, as a simulation of the assistant having "learned" the user's preferences. The *Rules* variant permission request featured two additional choices: *Always allow* and *Always deny* (Figure 2b). These options were adjusted for each relevant app and data type (e.g., *Always allow* Calendar *access to dates*).

As part of our explanations, we told our participants that both the *Rules* and *Learning* designs had an extra feature: a "review mode" that allowed users to see what decisions were made automatically on their behalf and change them if necessary. Participants could invoke this mode during the simulation by asking to review their permissions. If they did so, we showed them a separate screen that contained copies of approved or denied permission requests. One of our research questions was whether participants would make use of this.

Misbehaving apps Most apps in the simulation were intended to perform correctly, only asking permission for pertinent information at relevant times. However, we also wanted to see how people would react to inappropriate permission requests. This would also serve as a basic test of the permission system's effectiveness at preventing malicious apps. To that end, during each of the interactive sessions, participants encountered a permission request from a new, previously unseen app, which would request access to the last thing said, even though it had no relevance to the app's actual functionality. The three misbehaving apps were *Celebrity gossip*, *Bedtime stories*, and *Smart lightbulb*. We chose them because they were plausible apps for an intelligent assistant generally, but unlikely to come up in conversations on the topics we provided to participants. To make this "attack" more random, we tried to vary when in the conversation it happened.

Interview questions After each interactive session, we interviewed the pair of participants about their experience. Our questions covered general impressions of the proactive assistant and specific feedback about the permission prompts. We also collected perceptions and preferences for the different permission modes. Finally, we asked directly about privacy with respect to the proactive assistant, including any concerns people had and controls they wished to see in a device. The complete interview guide can be found in Appendix B.

Analysis We analyzed the interviews in our study using an inductive approach to thematic analysis [20]. Two coders reviewed each interview and created a codebook with themes

identified across responses. After agreement was reached, both coders annotated passages with themes from the codebook. We did not compute interrater reliability, as it is not well-defined when the unit of analysis is an entire interview [15,64]. We also did not compute statistics, as the small scale of qualitative research does not lend itself to quantitative generalizations [68]; instead, we report the range and general prevalence of different attitudes.

3.4 Recruitment and demographics

We recruited participants for our study by advertising a "computer gig" on Craigslist in different locales in the United States. A screening survey asked for basic demographics and three free-response questions about the respondent's use of smart home devices. When inviting people to the main study, we tried to balance different levels of experience with smart home technologies: low (limited or no usage of voice assistants), moderate (usage of smart speakers only), and high (multiple smart home devices besides smart speakers). Among those who completed the study, 65% used a smart speaker and 39% had other smart devices. We also aimed to balance our sample demographically. All procedures were IRB-approved.

Our screening survey was completed by 176 people, from whom we selected 23 pairs to participate in the study. The majority of the pairs (52%) consisted of spouses or partners, 30% were made up of family members, and the others were friends or roommates (9% each). Among the 46 participants, 57% were female; the mean age was 37; and 30%, 28%, 24%, and 18% self-identified, respectively, as White, Black, Asian, and of multiple or different ethnicities. The study session lasted 90 minutes, and participant pairs received $60 in compensation (to be shared by the two people).

3.5 Limitations

Our work has a number of limitations, which are driven in large part by the hypothetical nature of our target devices. Wizard of Oz simulations may elicit different reactions compared with real-world deployments; the time delay in ours further reduces realism. Runtime permissions, the focus of this paper, are just one type of privacy control; future work may investigate others. Some of our assumptions about architecture as well as the *Learning* mode may currently be impractical; but this may change due to the rapid progress of machine learning and other computing fields. Also, this work's threat model focuses on assistants and their apps and does not address the privacy threats posed by intra-household dynamics [28,38,51].

While smart displays (e.g., Echo Show) are becoming more widespread, most users currently interact with intelligent assistants through voice. Yet, a proactive assistant needs to provide suggestions ambiently, and we chose to deliver these on a screen, because this matched prototypes in literature [12,82], while audio-based ambient suggestions had not previously been studied on their own. After deciding on this, we felt that having audio permission requests to go along with visual suggestions would be confusingly inconsistent, opting for permissions to also be requested visually (though accompanied by an audible bell). Since interaction modality can affect privacy perceptions [27], future work should investigate whether user reactions differ towards voice-based permissions.

Overall, our study required design choices that involve simplification and guesswork; nonetheless, we took care to control for and isolate privacy-relevant aspects of the system, so that our findings would be generalizable and could shed light on proactive assistants, even if the eventual products' exact implementation details will differ.

4 Results

This section describes participants' behavior during the interactive sessions and reports the major themes that resulted from analyzing the interview portions of our study.

General perceptions When making sense of the proactive assistant's functionality, existing smart speakers were a baseline for feature comparison: *"It just seems like an enhanced Alexa"* (P16B). We found that our participants were, on the whole, receptive and even enthusiastic about the idea of a proactive assistant when it was first introduced to them. One of the closing questions in our interview was whether the participants would choose to adopt a proactive assistant. With only a few exceptions, our participants agreed that they would.

> *"I think it's nice that you don't have to call out the name because it's already picking up on the conversation"* (P16A)

Though participants perceived proactivity positively, they were aware of its privacy implications. For example, a number of participants relayed stories of existing devices listening at unexpected times, such as voice assistants interrupting a conversation to answer a question no one asked. Such accidental activations remain a regular occurrence [33,79].

4.1 Privacy perceptions

When we asked participants for their initial reactions, only a small fraction mentioned privacy, but the subsequent interviews revealed nuanced and situation-dependent viewpoints. This relative nature of privacy perceptions is consistent with other research [94] as well as the theory of contextual integrity [69], which argues that privacy expectations depend not only on data type, but also on contextual factors including the data subject, recipient, and transmission principle. In this way, our findings echo those of many other privacy studies. Despite the potential repetitiveness, we report these results to convey that context holds constant even with a new and potentially controversial technology like always-listening devices.

Privacy nihilism A very small number of people claimed that they do not care about privacy at all, repeating the common trope about having nothing to hide:

> *"I think we're very average people, you know, and privacy is not an issue, at least for us."* (P4B)
> *"I guess I don't have too much to hide."* (P22B)

Resignation A more common opinion, though still in the minority, was privacy resignation, a phenomenon that has been observed in other contexts as well [80]. While these people valued their information, they felt that attempts to protect it would, to a large extent, be futile because modern technology is designed to collect as much data as possible.

> *"In this day and age, everybody's recording everything."* (P21B)
> *"We have technology everywhere, like that's kind of beyond us at this point."* (P23A)

The other common reason for resignation was the belief in hackers' ability to obtain almost any information:

> *"Anybody can hack into anything."* (P17A)
> *"There's always third parties out there now. If they really want to hack in anything it's easy—so easy—for them."* (P12B)

Worries about hackers were common even among those who did not express quite such an absolute conviction about attackers' abilities. As evidence, participants cited recent high-profile cyberattacks that had been reported in the media. P1A, for example, felt that the government was powerless to stop these (*"they can't secure nothing"*).

Privacy contradictions Even the people who claimed that they were not concerned about privacy actually demonstrated nuanced views. (This is consistent with much research on the so-called "privacy paradox" [84].) For instance, P9B described themselves, *"I'm pretty much an open book. I mean, I think a lot of people worry too much about privacy."* Yet, shortly thereafter, they provided an explicit example of data types they did consider private: *"If I ask [my partner] for a social security number, if I'm filling it out, you know, I may not want [the assistant] to do things like that."* P17A drew a clear distinction between two privacy-invasive behaviors, one that they did not mind and another they considered unacceptable:

> *"I don't really care that they're kind of tracking me in a way, but I don't want someone to break into the system and find out where I am and stuff. That's scary."* (P17A)

Consistent with the theory of contextual integrity [69], this example illustrates that while P17A finds some data flows acceptable, others would be considered norm violations.

General privacy concerns The majority of our participants articulated some privacy concerns about always-listening devices, either organically over the course of the interview, or directly, when prompted. Often, these concerns were attributed to "some people," rather than themselves:

> *"I think this would be something that I feel like a lot of people would be concerned about."* (P18A)

Only a couple of interviewees expressed discomfort with the always-listening nature of the device more generally. (P1A, for example, referenced Orwell's *Nineteen Eighty-Four* [71]). On the whole, though, always-listening did not bother people; instead, there was specific information and scenarios that they were concerned about.

Sensitive data types Consistent with popularly held notions about what is considered private information [22, 73] and research on voice assistants and their third-party apps [3], the most common data type participants worried about was financial information, such as bank accounts, credit card details, social security numbers, or account credentials.

> *"Anything that has to do with my banking information, anything about money."* (P16B)
> *"Like your address, your social security number."* (P15A)
> *"I'm talking to customer care and they ask me for my credit card details or my PIN."* (P20B)

Participants were also worried about the device overhearing conversations on subjects they considered sensitive, with several highlighting gossip as a specific example.

> *"Let's say we're gossiping."* (P19B)
> *"What if I'm talking to someone, you know? We're planning a funeral or something? Maybe I don't want Alva*[4] *listening. And maybe that person is sharing stuff and they don't want it listening."* (P8A)

The latter quote also demonstrates concerns about non-owners of the assistant whose voice might be captured against their will. Tensions between primary and secondary users are a common feature of smart homes [38, 51, 95].

While medical information is often considered sensitive in the United States [30, 73], only two participants brought it up in our interviews.

> *"I wouldn't want the whole world to know my medical history."* (P1A)
> *"When it comes to financial and medical things, that should obviously be protected."* (P19A)

A few people referenced arguments or disputes as another example of a specific sensitive conversation subject.

> *"We got into an argument and we're going, 'he said, she said.' "* (P7B)
> *"If we're ever having, let's say, an argument. Or we're, you know, having a tough conversation or something."* (P4A)[5]

[4] *Alva* is the name we used for the intelligent assistant in our study.

[5] In this case, however, the interviewee felt that there actually could be a role for a (sufficiently smart) assistant to step in and mediate: *"It would say, hey, take a break. You two should take some time apart right now."*

Other examples of sensitive conversations that participants came up with included *"family matters"* (P2A), relationships and cheating—*"I'm having an affair with somebody"* (P1A)—and business calls made while working from home.

> *"Now it's work from home, or I might be just calling a colleague and talking. [. . .] That's confidential."* (P20B)

While most concerns focused on specific data types, such as the ones above, one person brought up the issue of metadata leakage, pointing out that even innocuous conversations could reveal potentially sensitive details. They felt, therefore, that all data—not just "private" conversations—merited protection.

> *"Anything can be used. Like me making a dinner reservation for seven o'clock is not a problem, until the stalker breaks into my house and wants to find out what I'm doing at seven o'clock. So it could be information that's not harmful. But in the wrong hands, it can become harmful."* (P19A)

Indeed, a variety of inferences can be made from voice even without considering content [52], and advertisers have sought to exploit all information available to them [66].

Data uses Some of the concerns voiced by participants focused on what would happen with their information—for example, who would get it, where it would be stored, and for how long—rather than the specifics of the data. Concretely, a number of people expressed discomfort with the possibility of their data being sold.

> *"If they were selling my information and then if I was wanting to plan a trip to Hawaii and then suddenly I received calls from my travel agent or something."* (P14A)

Intra-household data leakage Several participant pairs brought up the possibility that the assistant would overhear conversations and later reveal their contents, in one way or another, to other members of the household, leading the person to find out secrets others are keeping from them.

> *"Maybe something that you discussed—it was really really private—popped up on the screen and somebody else in the house saw it."* (P6A)

Secrets need not be a sign of malfeasance or problems in the household, but are instead benign everyday occurrences:

> *"Kids, they're very nosy, so they don't need to know everything. What if you're planning a surprise party and they're going to want to be, like, oh what were mom and dad talking about?"* (P10A)
> *"Let's say I'm throwing a surprise dinner for [partner]. [. . .] But then [the partner is at] home and [assistant] just starts blurting out next week's plans, and I'm, like, did I freaking tell you to do that?"* (P19A)

Impactful actions Overwhelmingly, concerns expressed by participants in our study focused on impactful action the assistant might take. These worries—that the assistant would do something the user would disapprove of—were much more common than concerns about what would happen with data.

While different in kind, the *contexts* for these concerns were similar to the data types above. For example, the top concern was that the assistant would take actions with financial consequences, such as buying items or booking tickets.

> *"I want to make my own financial decisions."* (P1A)

People also worried about social consequences that might follow from the assistant performing actions without approval, for example messaging friends or creating invitations. (Communications are often a source of privacy concerns [13, 83].)

> *"I would always have it [ask me] only when it's going to send something to someone else, like a person in my contacts or something else."* (P4A)

Even if the assistant's actions affected no one but the user of the device, participants observed, they are still able to cause annoyance or inconvenience, for example through unwanted events being scheduled or alarms being set.

> *"If you're having a discussion with someone and it comes up, hey, should we cancel dinner for tomorrow? [. . .] She might automatically do that without hearing the end result, or put random things on your calendar."* (P9B)

While the inconvenience stemming from such autonomous actions may be judged as relatively minor, participants often felt that it was these violations that permissions ought to be, or were, guarding against.

4.2 Runtime permissions effectiveness

A major goal of this study was to observe how runtime permissions would perform in a semi-realistic setting.

Concept comprehension Overall, we observed that nearly all participants understood how to use permissions right away. The majority of permission requests in our study were approved; when participants denied one, it was typically because they considered the service unnecessary, for example if the assistant offered driving directions to a familiar destination.

One area where there may have been a gap in participants' understanding was in the role of third-party apps. As part of our overview, everyone heard that features—including the most basic ones—were implemented by apps. Nonetheless, participants never treated the apps as distinct from the assistant. It is possible that this was an artifact of our study, since we framed it as a test of the assistant in general. However, researchers have observed similar confusion with existing third-party skills [58], so the issue may be more universal.

Detecting inappropriate requests We found that our permissions system worked fairly well for preventing data capture by the "misbehaving" apps (§3.3). Participants denied a large majority of permission requests from these apps, whereas they allowed most requests from other apps. Many

also commented about the misbehaving apps, providing evidence that they were paying attention and that the observed behavior was anomalous and memorable.

> *"That made me very alert: why did they talk about bedtime stories right now? It's got nothing to do with what we were talking about."* (P20A)

Some participants (less than a quarter of all cases) did allow permission requests from misbehaving apps. This was primarily due to lack of attention or some amount of habituation.

While some described the inappropriate permission requests as weird or even "spooky," most were not concerned by them. Rather than evidence of an attempt at data capture, people saw them as in line with bugs they had experienced using current voice interfaces, for example due to speaking English with an accent. Consistent with our observation that participants did not clearly distinguish apps from the platform, those who commented on inappropriate requests attributed the mistakes to the assistant itself.

> *"It's kind of like when Siri gets stuff wrong."* (P10A)
> *"Sometimes my accent makes me say the things or certain words with a different tone or something. And the program could misunderstand those types of things."* (P4B)

4.3 Runtime permissions perceptions

One of our main research goals was to collect first-hand feedback on the user experience of runtime permission requests.

Ask-every-time is annoying In the first session, the assistant asked for permission on every potential data access. As expected, everyone agreed that this resulted in too many permission requests, describing the experience as "annoying" and expressing a strong desire for fewer interruptions.

> *"That's going to get on people's nerves, okay?"* (P3B)

Because they resulted in significantly fewer permission requests, the streamlined permission modes (*Rules* and *Learning*) were received much more positively. However, beyond that, there was not much consensus about the two modes and their distinctive properties.

Advantages of *Learning* Between the two permission modes, a slight majority preferred *Learning*. This group expressed trust in the automation to accurately learn their preferences and explained that they were not concerned about it making mistakes and granting inappropriate permissions.

> *"Well I don't see any damage that it can do since it's not giving out any demands or orders anywhere."* (P13B)

Weaknesses of *Rules* Another reason people cited for preferring *Learning* was the cognitive overhead of the four permission choices in the *Rules* variant. The extra options required more time to read and also made the decision more complicated, since users had to think about whether they wanted to allow an app always or just once. While deliberation can help reduce the influence of heuristics and cognitive biases [16], too much may turn users away from the product.

> *"It creates a sense of paralysis by analysis."* (P17A)

Furthermore, nearly half of participants expressed some sort of confusion about this variant. Specifically, users were uncertain about whether "always allow" referred to the specific app being always allowed, or if it was the specific speech they uttered (for example, any app could always access the location they just mentioned).

> *"It kind of got me more distracted, because I'm having to stop to think about that."* (P14A)
> *"Is it that I don't need to allow the music or is that allowing the music allows all of the music apps?"* (P5B)

Another concern was that rules were active forever. Some assumed that was not the case, while others felt that it should not be. Research in other domains has identified users' desire for more dynamic rules [63] as well as for automatic data deletion and other forms of longitudinal privacy management [18, 48, 60].

> *"Just as a regular consumer, I assume it was good for just that day and then it would probably reset again."* (P16A)
> *"I hesitate to do it once or because I might change next time. I'm not sure if next time I go I might change, so I debate on should I use always or should I just use it once?"* (P12B)

Advantages of *Rules* Those who preferred the *Rules* variant expressed a desire for greater control over the assistant.

> *"Sounds really like therapist stuff, but I feel like I have more support with [Rules mode]. I felt like there was more hand-holding going on. I felt like I had guidance."* (P16A)

This variant was also popular among those who distrusted the assistant's automation—or simply did not see it as beneficial—and did not want it to make decisions on their behalf, especially if they might have undesirable consequences.

> *"It's like the AI would be the one controlling it. And I think, in that situation, it's, like, why are you asking permission if you're going to not ask for permission later?"* (P23A)

Non-use of the review feature The review mode (in either condition) also received mixed feedback. Only a minority invoked it during the sessions, mostly out of curiosity. Many said afterwards that they forgot about it, but some critiqued its user experience or even the need for it.

> *"I find it difficult to use that feature, actually."* (P22B)
> *"I don't need that. I trust [the assistant]."* (P1A)

Most participants were not opposed to the idea of a review feature and many claimed they would use it, with varying frequency. The most common use case was if something suspicious happened, which is consistent with its use in existing

devices [56, 60]. Thus, the review feature's relative unpopularity may be an artifact of our study, and it may prove to be more in demand with prolonged use of the assistant.

Similar to permissions, most saw the value of the review feature in being able to oversee the apps' actions and the device's understanding, rather than an audit mechanism to verify that the apps and automation were not behaving badly.

> *"I wanted to see if not only I could see what apps I've approved, but also what I asked them to do. [. . .] So that I wouldn't have any duplicate actions or events."* (P4A)

4.4 Trust in permissions

We wanted to know whether the permissions helped people trust always-listening devices more.

Some see little value We found that a number of participants, especially those who were less concerned about their privacy, did not see a strong reason for permissions.

> *"I see [permissions] more of like a redundancy. [. . .] Buying it and having it in my house is almost like implicit consent as it is. "* (P17B)

Others appreciate the control Nonetheless, when prompted, a little under half of participants commented that permissions enabled their trust in the assistant.

> *"It makes me feel like I have the control for what I am allowing and I'm not allowing. So that gives me a sense of trust. Just because I feel like I'm the one making the decision."* (P14A)

Supporters of permissions spoke about how they provided a greater degree of control, which they wanted.

> *"If it's hearing everything, you know that it's already not private, but you're also wondering where this is going to. So that gives you a little bit more room to control it."* (P10A)

The fact that this preference was common but not universal could be a reflection of differences in the preferred level of control displayed by different people: while some people are interested in decision automation, others want only analysis automation and to make decisions themselves [72].

Many, including those who liked having the permissions, saw them as a way to control the suggestions, rather than a privacy feature.

> *"For me, the only time I would deny is if it was trying to help me too much. If it was something that I didn't want to do just yet."* (P4A)

Permissions don't address all concerns Even those who found permissions valuable did not see them as a comprehensive solution. When presented with a scenario in which they were reading a credit card number out loud near the assistant, only one person stated that the permission system on its own would provide adequate protection; the rest explained that they would not feel comfortable relying on it alone.

> *"One of the main things that I think of is the app malfunctioning. What if the information did get through even despite the permission?"* (P15A)

Instead, people described other protective behaviors they would engage in, such as leaving the room that had the smart speaker or unplugging the device.

> *"I would go to another room. I don't trust the microphones. I've been told that microphones are never off."* (P16A)

Some pointed out that they were worried not only about the apps but also about the device itself compromising their privacy. This is an important reminder that the threat model our study adopted is not fully aligned with that of real users.

> *"That doesn't have to do with the apps. All this has to do with Alva."* (P19A)

Retroactive auditing sufficient for those less privacy-conscious In addition to the less-interrupting permission modes that we tested, we also surveyed our subjects about a design we refer to as "auditing," in which an app's permissions requests are always approved automatically, but can be reviewed at any time, using the same interface that was provided for the other conditions. When we described this design to our participants, many thought it was preferable to all of the approaches they experienced first-hand. However, we note that prior work suggests that, in practice, engagement with such a review feature may be low [60].

> *"I'm kind of a lazy individual. I mean, I still get to control at the end, that's all that matters."* (P14B)

However, some had reservations about this approach, explaining that they felt that it took too much control out of their hands and that it could be abused by apps.

> *"I always want to know, because the companies sneak in those random ones [. . .] and they're just looking for some free data for their pockets. I like to catch that."* (P21A)

4.5 Other desired privacy protections

Participants discussed a variety of additional controls they wanted to see implemented and general privacy demands.

Turning listening off The ability to turn off the device's microphone was considered very important and helped our interviewees feel more comfortable with the device. However, studies of current smart speakers suggest that the mute button, present in all of them, is rarely used [56].

> *"Just having a simple on/off switch, or just saying verbally, 'Alva, turn yourself off!' "* (P21B)

Some wanted always-listening to only occur on demand, with the device *not* listening as its default behavior. User

studies have discovered analogous demands from users of existing smart speakers [56].

> *"Maybe there should be a feature where it doesn't listen to you all the time, it's an option when you want to start a conversation."* (P6B)

However, five different people admitted that, if the always-listening mode existed, they would forget to turn it off.

> *"The logical thing to do would be to turn it off, but if they're always there, I think I would just forget that."* (P23A)

Voice identification More than half of participant pairs independently requested a voice identification feature, in which the device should only respond to recognized voices and potentially treat different people or voices differently. Similar features are available in existing voice assistants as Alexa's Voice Profiles [11] and Google Assistant's Voice Match [40]. Voice authentication is also offered by many banks [43].

> *"You can select that Alva should only detect some voices. Maybe it's my voice. It can only do tasks after it hears my voice. And if it's someone else's voice, it just mutes."* (P11A)

Parental controls Many also independently suggested parental controls as an important feature. While such controls are used relatively infrequently by parents of teenagers [39], the participants in our study generally sought protection for much younger children [70].

> *"Does Alva have a way to block off a toddler? Because our son can talk now. If he figures this out, he can send reminders non-stop every day."* (P7B)

Parents had different views about how much access their children should have. Some felt that the device should ignore children's voices altogether, while others simply wanted to get age-appropriate content.

> *"It would be me and my wife and then the kids would be excluded."* (P5A)
> *"If something was going to be kind of inappropriate or like 18+ type content, then a pop-up or a preference allowance or warning would come up."* (P21B)

Passwords and other prohibitions Other controls people came up with included limits on the times of day when the device would operate.

> *"If I could maybe set up some times when Alva should be muted, then I think that would be good. Like if it could only hear me in the morning or in the evening and not apart from that."* (P22A)

Another recurring suggestion was per-user passwords that would restrict access to data on the device.

> *"There could be an option of putting a password that could enable Alva to recognize yourself as the owner"* (P2A)

Participants may have been inspired by a variety of current systems; most relevantly, Alexa already offers the option to set a 4-digit "voice code" which is used to confirm purchases and prevent accidental orders [9]. However, research has found that this approach does not meet everyone's security needs, especially in higher-risk scenarios [75].

Other suggestions included "stop" words that would direct the device to stop recording, blocklists of specific words, and filtering if the conversation turns to certain topics. These approaches, while not available in present devices, appear practical based on techniques in published research [86].

> *"I would have a list of banned words. Financial, order, whatever. Social Security, tax, financial, money, cash."* (P1A)
> *"I would want some type of masking to automatically happen, if it's possible."* (P19A)

Business practices Participants brought up other privacy expectations for always-listening platforms that focused on how the companies operated.

One requirement was a rigorous review process that all apps for the device would have to undergo, analogous to that used by smartphone app stores.

> *"The main security feature is I would want Alva to monitor anything that looks suspicious."* (P17B)

Today's voice assistant platforms already require third-party skills to undergo "certification" [10]; however, this verification process may become more difficult for proactive assistants, if they allow their apps the same level of freedom and flexibility allowed by our architecture.

Multiple participants said that they wanted to be compensated in the event a data breach occurred. Some responses suggested a belief that there are existing policies or laws that provide for this. Such misunderstandings of privacy regulations are long-standing and well-documented [87].

> *"You get your money back and like a compensation type of thing. You know, like in the privacy article."* (P15A)

One respondent explained that they hoped developers would only collect the data they need, a strategy recognizable as data minimization, which is a requirement of regulations such as GDPR [35].

> *"If it's not using it to work or to search for us, then it doesn't need it and it shouldn't sell it."* (P21A)

Participants also discussed other privacy factors that they found important. Among them was having a privacy policy that promised to respect their data, as well as providing security disclosures. These may be satisfied by requirements that arise from laws such as CCPA [1].

> *"I just want an assurance of my privacy and maybe its safety and reliability information."* (P2B)

Others brought up that their decision about adopting the device would be influenced by the manufacturer's reputation and their business model.

> *"I would be concerned about the company collecting and selling data, so I would probably search about how they operate."* (P21A)

5 Discussion

This study collected people's perceptions of proactive assistants, their privacy preferences, reactions to runtime permissions, and suggestions for other privacy controls.

5.1 Proactive assistant reactions

Many will welcome proactive assistants One basic observation from our study is that there was no wholesale rejection of proactive listening as creepy or excessive. Our participant sample *is* biased: we recruited people who were willing to be interviewed (and recorded) and many were already owners of smart speakers and other IoT devices. Still, we believe that smart speakers have paved the way for proactivity: our interviewees described it as a natural extension of present-day functionality. Even if our sample is not representative, there is evidently a market opportunity manufacturers may pursue.

Concerns center on actions and consequences While participants were open to proactive assistants, nearly all also expressed privacy concerns about them. Promisingly, the most common concerns seem plausible to overcome. With proactive assistants, people seem most worried about impactful activities: an assistant taking autonomous actions that carry financial, social, or personal consequences for the user. This result echoes recent findings about people's hesitance towards solely automated decision making [44], and can also be seen, through the lens of contextual integrity [69], as concerns about unintended flows. On the other hand, looking up information for ambient suggestions was seen as safe. From a designer's perspective, this appears straightforward to address by ensuring the assistant (or app) *confirms* with or *notifies* the user about any actions it is taking, such as making purchases or setting alarms. Allowing this feedback over multiple modalities may make it more convenient for the user in case, for example, they are too far away to see the display, or, conversely, the environment is too loud for the assistant to be heard.

Standard sensitive content should be excluded When it comes to the assistant simply hearing information (as opposed to taking actions), the concerns voiced by participants were similar for everyone. They centered primarily around a few sensitive data types, such as financial information or gossip, which is consistent with findings about privacy concerns generally [22, 73] as well as documented concerns about smart homes [14] and voice assistants specifically [3]. An implication of this finding for system developers is that they can assuage users' concerns, to a high degree, by blocking any app from hearing speech about financial, medical, or personal information. While these will vary in how easy they are to implement (detecting credit card numbers seems much more tractable compared with identifying gossip), this appears to be a promising research direction and likely an effective way of winning the trust of many potential users.

Intra-household controls needed Our interviews provided evidence for the well-known fact that people are concerned about protecting their privacy not just from apps, strangers, and other third parties, but also within the household [4, 19, 38, 51, 95]. As many participants suggested, voice identification could help: assistants could use it to limit access to interaction history, preferences, and other personal data.

5.2 Takeaways about runtime permissions

Our testing illuminated both positive and negative aspects of runtime permissions for proactive assistants.

Permissons, with architecture, help catch bad requests Permissions showed potential as a way of fending off inappropriate data access by apps, as most participants effectively identified and blocked the misbehaving apps in our study. For many, permissions also increased their trust in the device and gave them a sense of control, which they described as very important, especially for a device in such a sensitive setting.

Proposed permission designs show promise, face adoption challenges As a user experience for assistants, runtime permissions showed some promise, as participants understood them and were able to use them effectively. They were also quite successful methodologically, as an interactive and engaging way to elicit privacy attitudes and requirements. However, none of the permission modes we tested is likely to yield a user experience that would be acceptable for a real product. As predicted, no one—even those who were more privacy-conscious and wanted greater control—was happy being prompted every time an app wanted to access data. Reactions to the less-interrupting designs were much more positive, as participants appreciated their streamlined nature; still, they exhibited limitations of their own.

The ***Rules* design** provided the option to "always" allow or deny requests for specific combinations of apps and data types. People saw it as more usable than ask-every-time, while still leaving the user in control, which was especially welcome to those who were less trusting of the system. That sense of control may be misleading, however, as the relatively permanent nature of rules may lead people to forget about the permissions they granted. This is exacerbated by the fact that many were confused about what exactly they were allowing. Finally, a majority felt that having four options on every request

was too cognitively taxing. These pain points suggest that the *Rules* design, in its current form, would face challenges if adopted as a general-purpose permissions approach.

In contrast, the ***Learning* design** has the advantage of a simpler user experience. However, a sizeable minority of participants (even in our, potentially biased, sample) were unwilling to give up control over data access to a black-box algorithm. The development of an algorithm that can effectively learn people's preferences across a variety of contexts also remains an open research question, though it can build on existing work on predicting privacy preferences [3,17,26,29,32,57,93], which also show that a promising strategy may be to combine *Rules* and *Learning* approaches.

One interesting challenge for machine learning-based approaches to inferring people's preferences is the way participants used permission requests: they denied them not only when they considered the access inappropriate but also (and more commonly) when the provided service was not useful in that moment. Lacking a way to distinguish between these two reasons for denying requests, a model trained on this data may reach incorrect conclusions. This may be a fruitful avenue for future research, but for now, these challenges cast the practicality of the *Learning* approach into further doubt.

We also surveyed our subjects about "auditing," in which permissions were approved automatically, but subject to review after the fact. For the more privacy-conscious, this was unacceptable, but the majority actually preferred it, since it did away entirely with irksome interruptions from the permission requests. Yet our findings suggest that adopting this variant would likely lead to poor privacy outcomes. People would be unlikely to make use of the review feature, as evidenced by this study and experience with other systems [60]. This would be exacerbated by the misunderstanding many users have about the distinction between the assistant itself and third-party apps for it.

5.3 Design recommendations

While better or more practical approaches may emerge in the future, what if someone were trying to build a proactive assistant today? The most effective tactic may be to combine the strategies that emerged as most promising from this study. Concretely, we would recommend that an assistant have some of the following features.

First, since so many participants were uncomfortable with the assistant making consequential decisions independently, any actions that trigger consequences beyond ambient information display would be subject to manual approval at runtime. Feedback to and from the assistant should be be supported through multiple modalities (e.g., on-screen and using voice), as many pointed out that audio is better when they are not in front of the device, but that there are also times when background noise makes the screen a more effective medium.

While privacy is context-dependent, some data types are universally seen as more sensitive and deserve special scrutiny. To account for this, the platform should, by default, identify and block access to any financial information and other known sensitive topics. Users might review a list of such topics during setup, and exceptions could be made on a case-by-case basis (e.g., for banking apps).

The majority of our participants were not comfortable with always-on continuous listening, despite acknowledging its convenience. As a result, we believe that a privacy-friendly default would be to allow users to opt in to "online" proactive listening only for specific conversations or short periods of time. The rest of the time, the assistant would operate on-demand, like current voice assistants. In this setup, since users would opt into the listening deliberately, there is a greater expectation for conversations to be analyzed and therefore a reduced need for interrupting permission requests; these could instead be automatically approved. However, they should still be auditable after the conversation has ended, since participants expressed a desire to be able to go back and review the assistant's behavior. Because most people express confusion between apps and the first-party assistant [59], during these listening sessions (as well as at other times), users should be made aware of which specific apps are accessing their conversation, as well as whether they are first- or third-party [77]. Inspired by recommendations from our participants, the device should feature voice identification (to restrict users' access to their own data) and parental controls.

While this proposed prototype may not procure perfect privacy, it would significantly enhance it compared with other approaches where apps might always be listening, and it would address many of the concerns and user experience pain points perceived as part of our probe. Future work could explore whether there are permission designs or approaches that were not part of our study, which would yield a more favorable user experience or stronger privacy guarantees.

As assistant platforms prosper and proceed in popularity, perhaps progressing into proactivity, pressure will persist to provide proper protections from their potential problems; while not perfectly practical, and plainly no panacea, permissions proffer promising performance, which plenty of people perspicuously prefer to the present predicament of pitifully poor privacy.

Acknowledgments

We would like to thank Alex Thomas for assistance with data analysis, Julia Bernd for guidance and feedback during study development, and Florian Schaub, Noel Warford, Wentao Guo, Alan Luo, and Julio Poveda for comments on draft versions of the paper. This work was supported by the NSA's Science of Security program, NSF grant CNS-1801501, Cisco, and the Center for Long-Term Cybersecurity at UC Berkeley.

References

[1] California Consumer Privacy Act, 2018.

[2] Noura Abdi, Kopo M. Ramokapane, and Jose M. Such. More than Smart Speakers: Security and Privacy Perceptions of Smart Home Personal Assistants. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. USENIX Association, 2019.

[3] Noura Abdi, Xiao Zhan, Kopo M. Ramokapane, and Jose Such. Privacy norms for smart home personal assistants. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 2021.

[4] Imtiaz Ahmad, Rosta Farzan, Apu Kapadia, and Adam J. Lee. Tangible privacy: Towards user-centric sensor designs for bystander privacy. *Proc. ACM Hum.-Comput. Interact.*, 4(CSCW2), October 2020.

[5] Devdatta Akhawe and Adrienne Porter Felt. Alice in warningland: A large-scale field study of browser security warning effectiveness. In *Proceedings of the 22nd USENIX Security Symposium*, pages 257–272, 2013.

[6] Hazim Almuhimedi, Florian Schaub, Norman Sadeh, Idris Adjerid, Alessandro Acquisti, Joshua Gluck, Lorrie Faith Cranor, and Yuvraj Agarwal. Your Location has been Shared 5,398 Times!: A Field Study on Mobile App Privacy Nudging. pages 787–796. ACM Press, 2015.

[7] Amazon. Alexa Skills. https://www.amazon.com/alexa-skills/b?ie=UTF8&node=13727921011.

[8] Amazon. Configure Permissions for Customer Information in Your Skill. https://developer.amazon.com/en-US/docs/alexa/custom-skills/configure-permissions-for-customer-information-in-your-skill.html.

[9] Amazon. Require a Voice Code for Purchases with Alexa. https://www.amazon.com/gp/help/customer/display.html?nodeId=GAA2RYUEDNT5ZSNK.

[10] Amazon. Skill Certification Requirements. https://developer.amazon.com/en-US/docs/alexa/custom-skills/certification-requirements-for-custom-skills.html.

[11] Amazon. What Are Alexa Voice Profiles? https://www.amazon.com/gp/help/customer/display.html?nodeId=GYCXKY2AB2QWZT2X.

[12] Salvatore Andolina, Valeria Orso, Hendrik Schneider, Khalil Klouche, Tuukka Ruotsalo, Luciano Gamberini, and Giulio Jacucci. Investigating Proactive Search Support in Conversations. In *Proceedings of the 2018 Designing Interactive Systems Conference*, DIS '18, pages 1295–1307. ACM, 2018.

[13] Julio Angulo and Martin Ortlieb. "WTH..!?!" experiences, reactions, and expectations related to online privacy panic situations. In *Eleventh Symposium on Usable Privacy and Security (SOUPS 2015)*, pages 19–38, Ottawa, July 2015. USENIX Association.

[14] Noah Apthorpe, Yan Shvartzshnaider, Arunesh Mathur, Dillon Reisman, and Nick Feamster. Discovering Smart Home Internet of Things Privacy Norms Using Contextual Integrity. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(2):59:1–59:23, July 2018.

[15] David Armstrong, Ann Gosling, John Weinman, and Theresa Marteau. The Place of Inter-Rater Reliability in Qualitative Research: An Empirical Study. *Sociology*, 31(3):597–606, August 1997.

[16] Paritosh Bahirat, Martijn Willemsen, Yangyang He, Qizhang Sun, and Bart Knijnenburg. Overlooking context: How do defaults and framing reduce deliberation in smart home privacy decision-making? In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 2021.

[17] Natã M. Barbosa, Joon S. Park, Yaxing Yao, and Yang Wang. "What if?" Predicting Individual Users' Smart Home Privacy Preferences and Their Changes. *Proceedings on Privacy Enhancing Technologies*, 2019(4):211–231, October 2019.

[18] Lujo Bauer, Lorrie Faith Cranor, Saranga Komanduri, Michelle L Mazurek, Michael K Reiter, Manya Sleeper, and Blase Ur. The post anachronism: The temporal dimension of Facebook privacy. In *Proceedings of the 12th ACM Workshop on Privacy in the Electronic Society*, pages 1–12. ACM, 2013.

[19] Julia Bernd, Ruba Abu-Salma, and Alisa Frik. Bystanders' privacy: The perspectives of nannies on smart home surveillance. In *10th USENIX Workshop on Free and Open Communications on the Internet (FOCI 20)*. USENIX Association, August 2020.

[20] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, January 2006.

[21] Barry Brown, Moira McGregor, and Donald McMillan. Searchable objects: Search in everyday conversation. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*,

CSCW '15, pages 508–517, New York, NY, USA, 2015. Association for Computing Machinery.

[22] Aylin Caliskan Islam, Jonathan Walsh, and Rachel Greenstadt. Privacy detective: Detecting private information and collective privacy behavior in a large social network. In *Proceedings of the 13th Workshop on Privacy in the Electronic Society*, WPES '14, pages 35–46, New York, NY, USA, 2014. Association for Computing Machinery.

[23] Weicheng Cao, Chunqiu Xia, Sai Teja Peddinti, David Lie, Nina Taft, and Lisa M. Austin. A large scale study of user behavior, expectations and engagement with Android permissions. In *30th USENIX Security Symposium (USENIX Security 21)*, pages 803–820. USENIX Association, August 2021.

[24] Juan Pablo Carrascal, Rodrigo De Oliveira, and Mauro Cherubini. To call or to recall? That's the research question. *ACM Transactions on Computer-Human Interaction*, 22(1), March 2015.

[25] Long Cheng, Christin Wilson, Song Liao, Jeffrey Young, Daniel Dong, and Hongxin Hu. Dangerous Skills Got Certified: Measuring the Trustworthiness of Amazon Alexa Platform. In *ACM Conference on Computer and Communications Security (CCS)*, 2020.

[26] Saksham Chitkara, Nishad Gothoskar, Suhas Harish, Jason I. Hong, and Yuvraj Agarwal. Does This App Really Need My Location?: Context-Aware Privacy Management for Smartphones. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(3):42:1–42:22, September 2017.

[27] Eugene Cho. Hey Google, Can I Ask You Something in Private? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 258:1–258:9. ACM, 2019.

[28] Camille Cobb, Sruti Bhagavatula, Kalil Anderson Garrett, Alison Hoffman, Varun Rao, and Lujo Bauer. "I would have to evaluate their objections": Privacy tensions between smart home device owners and incidental users. *Proceedings on Privacy Enhancing Technologies*, 2021(4):54–75, 2021.

[29] Jessica Colnago, Yuanyuan Feng, Tharangini Palanivel, Sarah Pearman, Megan Ung, Alessandro Acquisti, Lorrie Faith Cranor, and Norman Sadeh. Informing the Design of a Personalized Privacy Assistant for the Internet of Things. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, Honolulu HI USA, April 2020. ACM.

[30] Committee on Health Research and the Privacy of Health Information: The HIPAA Privacy Rule, Board on Health Sciences Policy, Board on Health Care Services, and Institute of Medicine. *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*. National Academies Press, Washington, D.C., February 2009.

[31] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. Wizard of Oz studies: Why and how. In *Proceedings of the 1st International Conference on Intelligent User Interfaces*, IUI '93, pages 193–200, New York, NY, USA, 1993. Association for Computing Machinery.

[32] Anupam Das, Martin Degeling, Daniel Smullen, and Norman Sadeh. Personalized privacy assistants for the internet of things: Providing users with notice and choice. *IEEE Pervasive Computing*, 17(3):35–46, July 2018.

[33] Daniel J. Dubois, Roman Kolcun, Anna Maria Mandalari, Muhammad Talha Paracha, David Choffnes, and Hamed Haddadi. When Speakers Are All Ears: Characterizing Misactivations of IoT Smart Speakers. *Proceedings on Privacy Enhancing Technologies*, 2020(4):255–276, October 2020.

[34] Yusra Elbitar, Michael Schilling, Trung Tin Nguyen, Michael Backes, and Sven Bugiel. Explanation beats context: The effect of timing & rationales on users' runtime permission decisions. In *30th USENIX Security Symposium (USENIX Security 21)*, pages 785–802. USENIX Association, August 2021.

[35] European Parliament and the Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, 2016.

[36] Adrienne Porter Felt, Serge Egelman, Matthew Finifter, Devdatta Akhawe, and David Wagner. How to Ask for Permission. In *HotSec*, 2012.

[37] Adrienne Porter Felt, Elizabeth Ha, Serge Egelman, Ariel Haney, Erika Chin, and David Wagner. Android Permissions: User Attention, Comprehension, and Behavior. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, SOUPS '12, pages 3:1–3:14, New York, NY, USA, 2012. ACM.

[38] Christine Geeng and Franziska Roesner. Who's In Control?: Interactions In Multi-User Smart Homes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 268:1–268:13. ACM, 2019.

[39] Arup Kumar Ghosh, Karla Badillo-Urquiola, Mary Beth Rosson, Heng Xu, John M. Carroll, and Pamela J. Wisniewski. A matter of control or safety? Examining parental use of technical monitoring apps on teens' mobile devices. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–14. Association for Computing Machinery, New York, NY, USA, 2018.

[40] Google. Link your voice to your devices with Voice Match. https://support.google.com/assistant/answer/9071681.

[41] Google. Android 6.0 Changes. https://developer.android.com/about/versions/marshmallow/android-6.0-changes, 2015.

[42] Yue Huang, Borke Obada-Obieh, and Konstantin (Kosta) Beznosov. Amazon vs. My brother: How users of shared smart speakers perceive and cope with privacy risks. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–13, New York, NY, USA, 2020. Association for Computing Machinery.

[43] Rupert Jones. Voice recognition: Is it really as secure as it sounds? *The Guardian*, September 2018.

[44] Smirity Kaushik, Yaxing Yao, Pierre Dewitte, and Yang Wang. "How I Know For Sure": People's perspectives on solely automated Decision-Making (SADM). In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*, pages 159–180. USENIX Association, August 2021.

[45] J. F. Kelley. An empirical methodology for writing user-friendly natural language computer applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '83, pages 193–196, New York, NY, USA, 1983. Association for Computing Machinery.

[46] Patrick Gage Kelley, Sunny Consolvo, Lorrie Faith Cranor, Jaeyeon Jung, Norman Sadeh, and David Wetherall. A Conundrum of Permissions: Installing Applications on an Android Smartphone. In Jim Blyth, Sven Dietrich, and L. Jean Camp, editors, *Financial Cryptography and Data Security*, pages 68–79, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[47] Patrick Gage Kelley, Lorrie Faith Cranor, and Norman Sadeh. Privacy as part of the app decision-making process. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3393–3402. Association for Computing Machinery, New York, NY, USA, 2013.

[48] Mohammad Taha Khan, Maria Hyun, Chris Kanich, and Blase Ur. Forgotten But Not Gone: Identifying the Need for Longitudinal Data Management in Cloud Storage. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 543:1–543:12, New York, NY, USA, 2018. ACM.

[49] Jonathan Kilgour, Jean Carletta, and Steve Renals. The Ambient Spotlight: Queryless desktop search from meeting speech. In *Proceedings of the 2010 International Workshop on Searching Spontaneous Conversational Speech*, SSCS '10, pages 49–52, New York, NY, USA, 2010. Association for Computing Machinery.

[50] Ilker Koksal. The Sales Of Smart Speakers Skyrocketed. *Forbes*, March 2020.

[51] Vinay Koshy, Joon Sung Sung Park, Ti-Chung Cheng, and Karrie Karahalios. "We just use what they give us": Understanding passenger user perspectives in smart homes. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 2021.

[52] Jacob Leon Kröger, Otto Hans-Martin Lutz, and Philip Raschke. Privacy Implications of Voice and Speech Analysis – Information Disclosure by Inference. In Michael Friedewald, Melek Önen, Eva Lievens, Stephan Krenn, and Samuel Fricker, editors, *Privacy and Identity Management. Data for Better Living: AI and Privacy*, volume 576, pages 242–258. Springer International Publishing, Cham, 2020.

[53] Deepak Kumar, Riccardo Paccagnella, Paul Murley, Eric Hennenfent, Joshua Mason, Adam Bates, and Michael Bailey. Skill Squatting Attacks on Amazon Alexa. In *27th USENIX Security Symposium (USENIX Security 18)*, pages 33–47. USENIX Association, 2018.

[54] Christoffer Lambertsson. Expectations of Privacy in Voice Interaction–A Look at Voice Controlled Bank Transactions. Technical report, 2017.

[55] Frederic Lardinois. Google makes it easier to chat with its Assistant. *Techcrunch*, May 2022.

[56] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. Alexa, Are You Listening?: Privacy Perceptions, Concerns and Privacy-seeking Behaviors with Smart Speakers. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW):102:1–102:31, November 2018.

[57] Bing Liu and Ian Lane. Attention-Based Recurrent Neural Network Models for Joint Intent Detection and Slot Filling. In *Interspeech 2016*, pages 685–689, September 2016.

[58] David Major, Danny Yuxing Huang, Marshini Chetty, and Nick Feamster. Alexa, who am I speaking to?: Understanding users' ability to identify third-party apps on Amazon Alexa. *ACM Transactions on Internet Technology*, 22(1), September 2021.

[59] David J. Major, Danny Yuxing Huang, Marshini Chetty, and Nick Feamster. Alexa, Who Am I Speaking To? Understanding Users' Ability to Identify Third-Party Apps on Amazon Alexa. *arXiv:1910.14112 [cs]*, October 2019.

[60] Nathan Malkin, Joe Deatrick, Allen Tong, Primal Wijesekera, Serge Egelman, and David Wagner. Privacy Attitudes of Smart Speaker Users. *Proceedings on Privacy Enhancing Technologies*, 2019(4):250–271, 2019.

[61] Lydia Manikonda, Aditya Deotale, and Subbarao Kambhampati. What's Up with Privacy?: User Preferences and Privacy Concerns in Intelligent Personal Assistants. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '18, pages 229–235, New York, NY, USA, 2018. ACM.

[62] David Maulsby, Saul Greenberg, and Richard Mander. Prototyping an intelligent agent through Wizard of Oz. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, CHI '93, pages 277–284, New York, NY, USA, 1993. Association for Computing Machinery.

[63] Michelle L. Mazurek, Peter F. Klemperer, Richard Shay, Hassan Takabi, Lujo Bauer, and Lorrie Faith Cranor. Exploring reactive access control. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2085–2094, Vancouver BC Canada, May 2011. ACM.

[64] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), November 2019.

[65] Moira McGregor and John C. Tang. More to Meetings: Challenges in Using Speech-Based Technology to Support Meetings. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, pages 2208–2220, New York, NY, USA, 2017. ACM.

[66] Wei Meng, Ren Ding, Simon P. Chung, Steven Han, and Wenke Lee. The Price of Free: Privacy Leakage in Personalized Mobile In-App Ads. In *Proceedings 2016 Network and Distributed System Security Symposium*, San Diego, CA, 2016. Internet Society.

[67] Richard Mitev, Markus Miettinen, and Ahmad-Reza Sadeghi. Alexa lied to me: Skill-based man-in-the-middle attacks on virtual assistants. In *Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security*, Asia CCS '19, pages 465–478, New York, NY, USA, 2019. Association for Computing Machinery.

[68] Kate Moran. Collecting Metrics During Qualitative Studies, June 2021.

[69] Helen Nissenbaum. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford Law Books, Stanford, Calif, 2009.

[70] Marije Nouwen, Maarten Van Mechelen, and Bieke Zaman. A value sensitive design approach to parental software for young children. In *Proceedings of the 14th International Conference on Interaction Design and Children*, IDC '15, pages 363–366, New York, NY, USA, 2015. Association for Computing Machinery.

[71] George Orwell. *Nineteen Eighty-Four*. Secker & Warburg, London, 1949.

[72] R. Parasuraman, T.B. Sheridan, and C.D. Wickens. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 30(3):286–297, 2000.

[73] Pew Research Center. Public Perceptions of Privacy and Security in the Post-Snowden Era. Technical report, Pew Research Center, November 2014.

[74] Kurt Wesley Piersol and Gabriel Beddingfield. Pre-wakeword speech processing, 2020.

[75] Alexander Ponticello, Matthias Fassl, and Katharina Krombholz. Exploring authentication for security-sensitive tasks on smart home voice assistants. In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*, pages 475–492. USENIX Association, August 2021.

[76] Laurel D. Riek. Wizard of oz studies in HRI: A systematic review and new reporting guidelines. *J. Hum.-Robot Interact.*, 1(1):119–136, July 2012.

[77] Aafaq Sabir, Evan Lafontaine, and Anupam Das. Hey Alexa, Who Am I Talking to?: Analyzing Users' Perception and Awareness Regarding Third-party Alexa Skills. In *CHI Conference on Human Factors in Computing Systems*, pages 1–15, New Orleans LA USA, April 2022. ACM.

[78] Florian Schaub, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. A Design Space for Effective

Privacy Notices. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, pages 1–17, Ottawa, 2015. USENIX Association.

[79] Lea Schönherr, Maximilian Golla, Thorsten Eisenhofer, Jan Wiele, Dorothea Kolossa, and Thorsten Holz. Unacceptable, where is my privacy? Exploring Accidental Triggers of Smart Speakers. *arXiv:2008.00508 [cs]*, August 2020.

[80] John S. Seberger, Marissel Llavore, Nicholas Nye Wyant, Irina Shklovski, and Sameer Patil. Empowering resignation: There's an app for that. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 2021.

[81] Bingyu Shen, Lili Wei, Chengcheng Xiang, Yudong Wu, Mingyao Shen, Yuanyuan Zhou, and Xinxin Jin. Can systems explain permissions better? Understanding users' misperceptions under smartphone runtime permission model. In *30th USENIX Security Symposium (USENIX Security 21)*, pages 751–768. USENIX Association, August 2021.

[82] Yang Shi, Yang Wang, Ye Qi, John Chen, Xiaoyao Xu, and Kwan-Liu Ma. IdeaWall: Improving creative collaboration through combinatorial visual stimuli. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, pages 594–603, New York, NY, USA, 2017. Association for Computing Machinery.

[83] Manya Sleeper, Justin Cranshaw, Patrick Gage Kelley, Blase Ur, Alessandro Acquisti, Lorrie Faith Cranor, and Norman Sadeh. "I read my Twitter the next morning and was astonished": A conversational perspective on Twitter regrets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3277–3286. Association for Computing Machinery, New York, NY, USA, 2013.

[84] Daniel J. Solove. The Myth of the Privacy Paradox. *George Washington Law Review*, 89, February 2020.

[85] Madiha Tabassum, Tomasz Kosiński, Alisa Frik, Nathan Malkin, Primal Wijesekera, Serge Egelman, and Heather Richter Lipford. Investigating users' preferences and expectations for always-listening voice assistants. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 3(4), December 2019.

[86] Welderufael B. Tesfay, Jetzabel Serna, and Kai Rannenberg. PrivacyBot: Detecting privacy sensitive information in unstructured texts. In *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pages 53–60, October 2019.

[87] Joseph Turow, Michael Hennessy, and Nora Draper. Persistent misperceptions: Americans' misplaced confidence in privacy policies, 2003–2015. *Journal of Broadcasting & Electronic Media*, 62(3):461–478, 2018.

[88] Anthony Vance, Jeffrey L. Jenkins, Bonnie Brinton Anderson, Daniel K. Bjornn, and C. Brock Kirwan. Tuning Out Security Warnings: A Longitudinal Examination of Habituation Through fMRI, Eye Tracking, and Field Experiments. *MIS Quarterly*, 42(2):355–380, February 2018.

[89] Anthony Vance, Brock Kirwan, Daniel Bjornn, Jeffrey Jenkins, and Bonnie Brinton Anderson. What do we really know about how habituation to warnings occurs over time? A longitudinal FMRI study of habituation and polymorphic warnings. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 2215–2227. Association for Computing Machinery, New York, NY, USA, 2017.

[90] Sarah Theres Völkel, Daniel Buschek, Malin Eiband, Benjamin R. Cowan, and Heinrich Hussmann. Eliciting and analysing users' envisioned dialogues with perfect voice assistants. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.

[91] Jing Wei, Tilman Dingler, and Vassilis Kostakos. Developing the proactive speaker prototype based on Google Home. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI EA '21, New York, NY, USA, 2021. Association for Computing Machinery.

[92] Primal Wijesekera, Arjun Baokar, Ashkan Hosseini, Serge Egelman, David Wagner, and Konstantin Beznosov. Android Permissions Remystified: A Field Study on Contextual Integrity. In *24th USENIX Security Symposium (USENIX Security 15)*, pages 499–514, Washington, D.C., August 2015. USENIX Association.

[93] Primal Wijesekera, Arjun Baokar, Lynn Tsai, Joel Reardon, Serge Egelman, David Wagner, and Konstantin Beznosov. The feasibility of dynamically granted permissions: Aligning mobile privacy with user preferences. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 1077–1093, May 2017.

[94] Allison Woodruff, Vasyl Pihur, Sunny Consolvo, Laura Brandimarte, and Alessandro Acquisti. Would a Privacy Fundamentalist Sell Their DNA for $1000...If Nothing Bad Happened as a Result? The Westin Categories, Behavioral Intentions, and Consequences. In *Proceedings of the 2014 Symposium on Usable Privacy and Security*, pages 1–18. USENIX Association, 2014.

[95] Yaxing Yao, Justin Reed Basdeo, Oriana Rosata Mcdonough, and Yang Wang. Privacy perceptions and designs of bystanders in smart homes. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), November 2019.

[96] Bob Yirka. Google Nest hacker finds evidence of Google considering getting rid of 'Hey Google' hot words. *Tech Xplore*, October 2020.

[97] Eric Zeng, Shrirang Mare, and Franziska Roesner. End User Security and Privacy Concerns with Smart Homes. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*, pages 65–80, Santa Clara, CA, 2017. USENIX Association.

[98] Serena Zheng, Noah Apthorpe, Marshini Chetty, and Nick Feamster. User Perceptions of Smart Home IoT Privacy. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW):200:1–200:20, November 2018.

[99] Marrian Zhou. Amazon's Alexa Guard can alert you if an Echo detects smoke alarm, breaking glass. *CNET News*, December 2018.

Appendices

A Conversation prompts

For each of the three rounds of the study (§3.3), participants were given a different prompt to guide their conversation with their partner. This section includes the specific directions provided to the participants, as well as the list of apps that was "active" for that conversation. In verbal instructions, we explained that these were suggestions, rather than a script to follow, and that participants were free to deviate from them, as long as they stayed with the main topic.

A.1 Task 1

Dinner + shopping Your task is to arrange to cook dinner with your partner. You can decide things like:

- which day you'll be cooking
- who will be doing the cooking
- what you will cook
- what recipe you will use (feel free to find one online!)
- whether you have the necessary ingredients for the recipe
- which ingredients you need to buy
- where you'll go to buy those ingredients
- when you'll do that shopping

As you work on this task, Alva's apps may try to offer helpful suggestions on its screen or out loud.

Installed apps Here are some of the apps installed on your device:

- Supermarket helper
- Recipe search
- Shopping list
- Reminders
- Maps
- Calendar
- Social network

A.2 Task 2

Booking a weekend trip Your task is to plan an outing for this weekend with your partner. As part of your conversation, you might:

- Discuss availability and other conflicting events
- Discuss budget
- Choose destination
- Look up things to do
- Choose activities
- Look up directions
- Decide on where to eat
- Talk about whom you want to invite along

Installed apps

- Maps
- Calendar
- Social network
- Travel info
- Weather
- Flights (and other tickets)
- Lodging
- Coupons

A.3 Task 3

Booking a vacation Your task is to plan a vacation together with your partner. As part of your conversation, you might:

- Choose travel dates
- Discuss budget
- Choose destination
- Look up things to do
- Choose activities
- Search for tickets
- Decide on where to stay

Installed apps

- Maps
- Calendar
- Social network
- Travel info
- Weather
- Flights (and other tickets)
- Lodging
- Coupons

B Interview guide

B.1 Round 1 (ask-every-time)

B.1.1 General impressions

- Please give us your general impressions of being an Alva device user. What did you like about it? What did you dislike?

B.1.2 Why do people deny requests?

- I noticed you denied (or didn't approve) app _'s permission request. Can you explain why?

B.1.3 General feedback about permission prompts

- What did you think of Alva's permission requests (in general)?
  - Understandability
    - Were they clear or were they confusing?
    - Did they provide enough information?
  - Modality
    - Would you prefer to receive these requests in some other way?
    - What did you think about receiving them on the device's screen? (instead of on your phone, etc.)
  - Attention
    - Were the notifications effective at getting your attention?
    - Do you think, in a real situation, you'd notice or interact with these requests?
    - Would you want them to draw more attention to the notification? (e.g., louder noise) Or less?
  - Distractingness
    - Were the requests too distracting?
    - Do you think they should be more noticeable or less?

B.2 Round 2

B.2.1 Condition-specific UX questions

Learning

- Do you think Alva accurately learned your preferences? (Please explain.)
- Would you want your preferences learned in this way (if the learning were more accurate)?

B.2.2 *General* privacy questions about this *specific* condition

- Assuming you had an Alva, how willing would you be to install apps — either new ones or the ones from today
- Did you (want to) review the decisions made by the learning?
  — on it?
- Overall, how do you feel about your privacy with respect to Alva?
  - Do you feel that your privacy is adequately protected?
  - If not, why not? What scenario are you envisioning? What's missing?

B.3 Round 3 / exit interview

B.3.1 Condition-specific UX questions

Rules/heuristics

- Did you (want to) review the decisions made by the rule?
- Did you regret your decision to make it a rule? Are there choices the rule made that you would've preferred it didn't?
- Would you have wanted a more (or less) restrictive rule? "only allow locations when I said _"
- (if no rule ever used) Why didn't you make use of the "always allow/deny" option?

B.3.2 Comparing Alva 1 vs 2

- How did the experiences of Alva 1 and Alva 2 compare for you?
  - Which Alva version does *each of* you prefer? Why?
  - Did you find the differences between the two Alva versions meaningful? (Please explain.) How strong is this preference? Is it only because I'm asking? Would you only use one of them, or you prefer one but it's not that big a deal?
  - What are the pros and cons of each version?
  - Did you prefer the user experience one or the other?
  - Do you trust one or the other more?
- Would you be comfortable having a conversation that involves sensitive topics, if you knew the apps from today's session would be listening (but they'd still have to request permission before sharing any data)?