Gary Liu and Nathan Malkin*

Effects of Privacy Permissions on User Choices in Voice Assistant App Stores

Abstract: Intelligent voice assistants, and the third-party apps (aka "skills" or "actions") that power them, are increasing in popularity and beginning to experiment with the ability to continuously listen to users. This paper studies how privacy concerns related to such always-listening voice assistants might affect consumer behavior and whether certain privacy mitigations would render them more acceptable. To explore these questions with more realistic user choices, we built an interactive app store that allowed users to install apps for a hypothetical always-listening voice assistant. In a study with 214 participants, we asked users to browse the app store and install apps for different voice assistants that offered varying levels of privacy protections. We found that users were generally more willing to install continuously-listening apps when there were greater privacy protections, but this effect was not universally present. The majority did not review any permissions in detail, but still expressed a preference for stronger privacy protections. Our results suggest that privacy factors into user choice, but many people choose to skip this information.

Keywords: privacy, usability

DOI 10.56553/popets-2022-0116
Received 2022-02-28; revised 2022-06-15; accepted 2022-06-16.

1 Introduction

This paper studies people's interactions with app stores for *continuously listening voice assistants*. We see these as a potential evolution of today's voice assistants, which have reached hundreds of millions of consumers around the world [27] and provide a convenient, hands-free way to play media, make purchases, and control smart home gadgets [7], as well as access tens of thousands of voice-controlled third-party apps [4].

Gary Liu: University of California, Berkeley, E-mail: liu.garyj@berkeley.edu
***Corresponding Author: Nathan Malkin:** University of California, Berkeley, E-mail: nmalkin@cs.berkeley.edu

Current voice assistants act predominantly on commands that follow wakewords (e.g., "hey Siri"), but in order for these devices to become more helpful, this requirement may be relaxed. For example, Google Assistant can, under certain circumstances, be activated without the wakeword [32, 56]. Amazon has experimented with "pre-wakeword speech processing" [45] and released Alexa features that monitor the audio environment continuously for certain sounds [58]. Continuous listening has also found its way into wearable fitness trackers [8, 36]. Additionally, research in HCI has shown how always-listening voice assistants can benefit users through proactive health support, streamlining work environments, and automating common tasks [42, 49, 54, 55]. Furthermore, scientific studies have demonstrated consumers' interest (and additional ideas) for these features [52, 53]. We therefore believe there is a reasonable likelihood that, when the technology catches up, manufacturers will augment voice assistants with continuous listening features—always on, always listening, and ready to respond to user input without being explicitly prompted. The goal of our research is to help protect consumers who eventually choose to adopt this technology from the privacy risks it entails. Our study does this by measuring people's demand for privacy controls, comparing views on different privacy approaches, and showing when in the product usage cycle these preferences manifest themselves.

Our underlying hypothesis is that strong privacy controls and limitations on what voice assistants and apps hear would offer platforms, and their users, greater confidence to adopt continuous listening. However, at present, the privacy controls offered by voice assistants are limited. For smart speakers, the primary available control is a mute button, which is rarely used [33]. Similarly, there are not many privacy options and choices when it comes to apps. For example, Amazon Alexa requires users to review and grant permissions when installing skills that access information deemed sensitive, such as a customer's name, address, or email [5].

Would stronger privacy controls make people more willing to adopt continuously listening assistants and their apps? If so, what should those controls be? What is the best way to protect user privacy while maintain-

ing device functionality? Our study attempts to make progress on this problem by focusing on one specific aspect of it: **which privacy protection approaches make people more likely to use continuously listening assistants?** Specifically, we aim to: (1) measure the extent to which the privacy protections of a continuously listening assistant matter to potential users, (2) consider specific examples of privacy controls and how people might react to them, and (3) investigate methodological techniques for studying a device's acceptability before the full ecosystem has been developed.

To explore these questions with more realistic user choices, we presented participants in our study with a description of a particular—but as-yet hypothetical—always-listening voice assistant. To accompany this vision, we built an interactive app store that allowed users to install apps for this device. Each participant was serially presented with two different assistant designs that offered varying levels of privacy protections, and was asked to install apps from the app store for each condition. Our findings show that many people prefer enhanced privacy protections and install more apps when those are in place. However, despite over half of participants indicating that privacy was a reason for preferring one of the two conditions they were assigned, the majority did not look at the permissions for the apps they installed. We conclude that privacy controls could improve user acceptance of continuous listening devices, but implementation details are critical to ensure that users are aware of said controls. Our experiments also shed light on which types of controls could be most effective, along with their potential limitations.

2 Related work

This section surveys the literature relevant to our study.

2.1 Continuous listening assistants

Our work studies the acceptability of different versions of proactive, continuously-listening voice assistants. Today's assistants do not provide this level of monitoring, though their microphones are always on, listening for wakewords and other sounds [58]. Other continuously listening products currently on the market are also more limited in their scope. For example, Fitbit products offer snoring detection [36], while Amazon and Spotify are focused on identifying mood and emotions [8, 46]. However, academic researchers have shown the potential for more advanced technologies, such as assistants that engage their users proactively [54], for example by analyzing conversations to offer information from relevant web searches [49] or performing administrative tasks like creating reminders, updating calendars, or sending emails [42]. Researchers have also studied user perspectives on these more advanced technologies, asking about conversations they imagine having with a perfect assistant [53] and investigating features an advanced assistant such as this might offer [52]. We prepare for the eventuality that these proactive technologies may soon become real products by studying consumer privacy concerns and potential mitigations for them.

2.2 Voice assistant privacy concerns

While continuous listening assistants may still be in the future, present-day voice assistants already give rise to a variety of privacy concerns, which researchers have documented [31, 41]. Studies have compared the concerns of users and non-users [33], risks from within and without the household [21], and differences between beliefs and devices' actual behavior related to data retention [38] and data flow [1]. Researchers have also identified that special concerns arise from users whose voices are captured but who are *not* the device administrators ("passenger" or "incidental" users) [15, 28]. Our work helps extend the understanding of voice assistant privacy by studying how these concerns may impact users' decisions about installing apps.

2.3 Third-party apps, risks and mitigations

A major source of risk for voice assistants (as well as other platforms [51]) stems from their third-party features, known as "skills" for Alexa and "actions" for Google Assistant. (We refer to them using the more generic term "apps.") Researchers have created apps to execute man-in-the-middle attacks [43] and hijack interactions intended for benign apps [29]. While apps for voice assistants face a certification process [6], malicious apps have been able to bypass it [12]. The danger from malicious apps is compounded by users' sometimes-incorrect mental models about assistants [1], as they struggle to differentiate third-party apps from features of the assistant itself [37]. Researchers have sought to curb the risks from malicious apps by identifying those that ask for sensitive [48] or otherwise inappropriate

content [20]. Our work adds to the literature by proposing and examining new limitations that can be imposed on third-party apps to enhance privacy.

2.4 Install-time permissions

Our study focuses on how privacy choices during the installation process affect people's decisions about adopting apps for always-listening assistants. This is analogous to the decision users face when they are prompted to review permissions before installing apps in other situations, such as smartphone app stores. Researchers have studied these decisions, often finding low attentions and comprehension rates [17], but showing that user engagement can be improved through redesigned interfaces [24] and nudging [3]. Our work extends this research into a new domain, where install-time permissions have already started to make an appearance [5].

3 Threat model

Developing privacy-enhancing technologies necessarily raises questions about who will be doing the protection and against whom. In the context of voice assistants, the devices themselves are often seen as privacy risks [1, 33, 41] and are therefore a natural target. However, adding external privacy controls means introducing a new third party—which must be trusted—to the equation. That complicates research on this subject, especially due to the still-hypothetical nature of not only this party, but also the assistant itself.

Instead, we modeled voice assistants as the combination of (1) a platform and (2) apps that run on top of it. Under our **threat model**, the platform itself is trusted, and the privacy protections are targeted at limiting data exposure to apps.

We believe this is a realistic assumption, because existing voice assistants are already hosts to vast ecosystems of apps: third-party developers are able to write applications that add features and capabilities to the assistants, such as ordering food, arranging transportation, or playing games. Amazon Alexa, the most common assistant [9], advertises that its "skills" number in the tens of thousands [4].

The privacy risks of apps are also far from theoretical: over the years, academic and industry researchers have uncovered numerous ways in which motivated attackers could use them to gain access to user data, up to and including raw audio recordings from the devices' microphones [12, 14, 25, 43]. Furthermore, end-users are often uncertain about the boundaries between voice assistants and third-party "skills" [37]—a confusion which attackers are able to leverage [29, 57].

By assuming that only apps are distrusted—but the platform is not—we miss out on capturing some user concerns, which is a limitation of our approach. Nonetheless, we believe that our findings can be useful in protecting people's privacy against device manufacturers as well. First, we envision that, like current smartphones, most core functionality can be implemented through built-in apps, which will be subject to the same permissions as third-party apps. This allows for a smaller trusted computing base and leaves most user concerns directed at the (first-party) apps. Second, the solutions that we evaluate could also, in the future, be implemented by third-party devices (similar to those proposed in recent research [22, 35]) to limit what the assistant device (regardless of any apps) hears. We therefore believe that our results will inform those adopting a stricter and more realistic threat model.

4 Experiment design

Our goal is to determine which privacy approaches would make people more comfortable with using continuously listening voice assistants. One approach is to ask people directly: describe the assistant and the potential privacy protections, then survey self-reported preferences and willingness to adopt the technology. However, intentions do not always match actual behavior [26]. Therefore, we felt it would be more instructive to observe consumers' revealed preferences: seeing how they actually make decisions, rather than asking about how they *say* they would.

4.1 Uncovering revealed preferences

In general, there are a few ways in which preferences about voice assistants could be revealed. People who are more comfortable with continuous listening would be more likely to obtain one of these devices. They would also be more likely to install apps for it. Once the apps are installed, those more comfortable would be more willing to use them (i.e., keep the assistant enabled and have conversations around it).

Since continuously listening assistants are still hypothetical, there are no products that can be used to examine revealed preferences. However, aspects of the experience can be simulated, with varying degrees of realism. In particular, app installation might take place off-device (e.g., on the user's smartphone) and therefore does not rely on the assistant's existence. For this reason, we decided to focus our study on the app installation process for continuously listening assistants.

Framing our study around app installation enhances the realism of the preferences we collect because it mimics how users will encounter information and make decisions in the real world. After getting a new assistant device, consumers will not be required to read instructions (or complete a survey), but they *will* most likely install apps. Therefore, to make the study more realistic for our participants, we created a fully interactive app store for a hypothetical voice assistant, modeled after existing smartphone app stores, which allowed users to browse and "install" a range of apps. The privacy information displayed in the store varied depending on the condition a participant was assigned to, which enabled between-subjects comparisons.

When considering the app installation process, we conjectured that, if people are less comfortable with a device's privacy features, they will be less likely to install new apps for it, because doing so might expose their private speech to untrusted third parties. We therefore hypothesized that **when people are more comfortable with an assistant's privacy approach, they will install more apps for it**.

Based on that hypothesis, our study compares the number of apps installed in different privacy conditions. However, we caution that the link between app installation and comfort is unproven; and therefore we note that the null hypothesis (finding that there is no difference between number of apps installed in different privacy conditions) does not imply that people have no preferences between the offered privacy protections. It could be that such preferences exist, but do not translate into different numbers of installed apps.

To enable us to verify our hypothesis despite this uncertainty, we designed our study to collect people's stated preferences about the privacy conditions in addition to their quantified behavior (the number of apps they installed). To facilitate this, we added a within-subjects experiment to our study, in which we had each participant try out more than one privacy condition. This allowed us to ask participants directly about the extent to which the privacy differences influenced their decisions, as well as gain additional qualitative insights.

4.2 Privacy protections

Our next task was to determine which privacy approaches to include in our acceptability experiments. A variety of privacy-preserving approaches would be suitable for inclusion; however, because we were using novel methods with an uncertain effect size, we decided to limit the number of conditions in our study and focus our evaluation on just two approaches to permissions—the ones we considered most promising—as well as a third, control, condition. Accordingly, we defined three privacy approaches, inspired by techniques referenced in literature [22, 35, 39], that would serve as our study's independent variable:

- The **Control** condition had no privacy protections. Every app would have access to all audio heard by the voice assistant. (For example, a flight reservations app might hear and share *all* conversations happening in the home.)
- Under the topic modeling ("**Topic**") approach, only speech that falls into an app's category will be accessible to it. A category is a predefined "bucket" of speech, such as a topic or intent. (For example, the topic for a flight reservations app might be *travel*.)
- In the network-restricted approach ("**Network**"), the voice assistant processes audio locally and will only pass speech on to third-party API endpoints when necessary for an app's functionality. The permission system enforces the rule that only certain types of data may leave the device. (For example, a flight reservations app might need to share destinations and desired prices with its server, but all other speech in a conversation can stay local.)

Table 1 lists expanded descriptions of each privacy model. All three privacy models assume that apps receive raw audio (though *which* audio might be restricted by the privacy approaches), rather than transcripts of conversations. We made this choice so that the architectures we explored would be compatible with a greater variety of apps with advanced use cases that might require access to underlying audio, such distinguishing speakers, analyzing emotions, or making sense of noises. However, this choice is not inherent to these privacy approaches, and, in fact, performing speech-to-text before sharing conversations with apps could be one possible privacy enhancement for platforms.

Table 1. Descriptions of the different privacy protections, as seen by participants. ("ALVA" is the name used to describe the continuously listening voice assistant in our study.)

	Description
Control	**Every app has access to all the audio that ALVA's microphone picks up. This means that a conversation held in the same room as ALVA will likely be entirely heard by all apps, while a quiet conversation in another room behind a closed door probably won't be heard.**
Topic	**To minimize the audio shared with third-party apps, ALVA lists the topic of speech that the app has access to. Only speech that is relevant to that topic will be accessible to the app. However, all speech that falls into an app's topic will be accessible to an app, even if it is not explicitly relevant to the app's functionality.**
Network	**ALVA will be listening to all speech that its microphone picks up, but will only record speech and send it to an app when necessary for functionality. This means that, even if ALVA heard you say something, nobody would find out unless it was sent to an app. What ALVA sends to each app is restricted by the app's privacy policy, which is defined by each app's third-party developer. This policy is found in each app's permissions section, which you can view before installing, on the app's ALVA Store page.**

5 Methods

To test our hypotheses, we designed a study that had participants browse a simulated app store under different privacy conditions and answer questions about their experience. In this section, we detail our methods for collecting and analyzing the data.

5.1 Study flow overview

Our study used a survey format to allow participants to complete the experiment in a self-guided manner. The survey consisted of a mix of free-response and multiple-choice questions, in addition to two interactive activities (browsing the app store), which were embedded directly within the survey (see Figure 1). We used a fictional new voice assistant, "ALVA," as a product that the participants would be giving their opinion on.

At a high level, participants in our study went through the following steps:

1. Learning about the voice assistant
2. Learning about a specific privacy protection method
3. Installing apps from the interactive app store
4. Answering questions about their installation choices
5. Learning about another privacy protection method offered by a "different version" of the device
6. Installing apps from the interactive app store for the new version of the device
7. Answering questions comparing the privacy options

Thus, participants learned about two different privacy protections. These were assigned randomly, without replacement (i.e., no participants experienced the same approach twice).

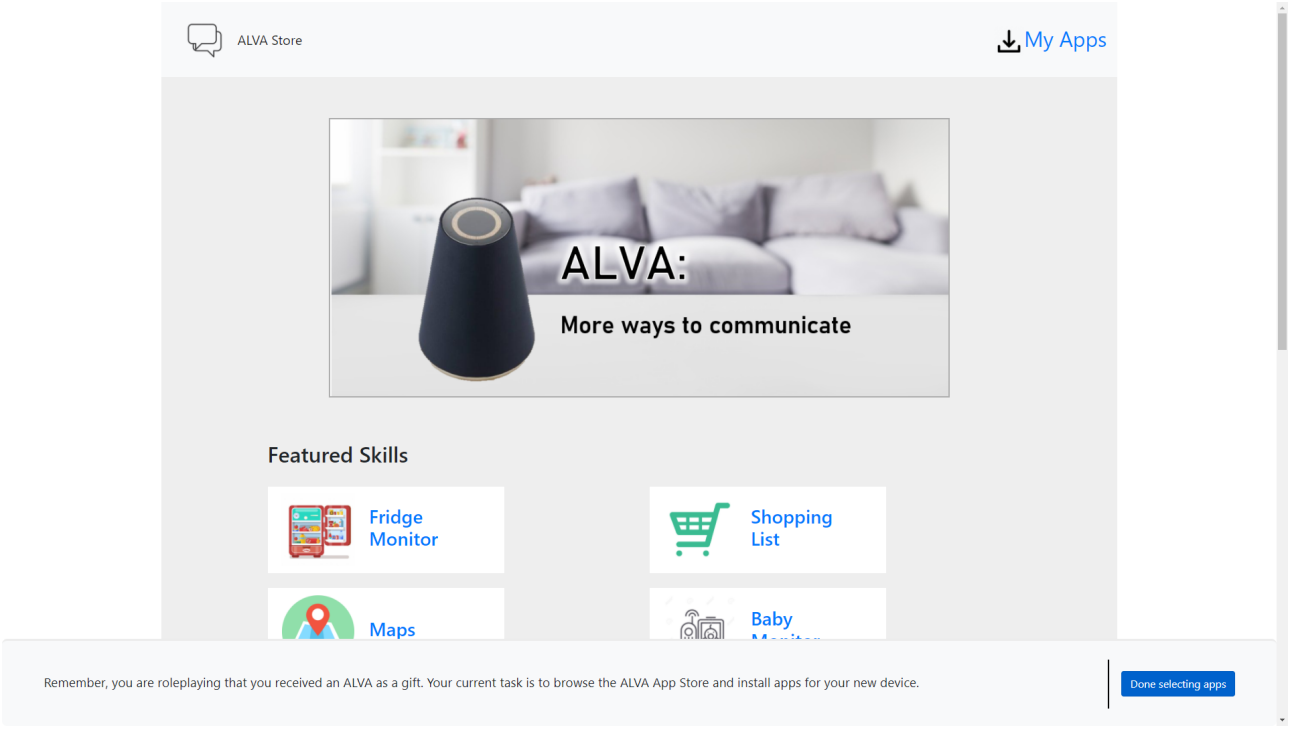
5.2 Study details

We began the survey by collecting information about participants' prior usage of existing voice assistants (e.g., Siri, Amazon Alexa, or Google Assistant) and their present attitudes towards them. (The complete survey instrument is available in Appendix A.)

Next, we introduced the ALVA device, describing it as always-listening in that—unlike existing voice assistants—it does not need a wakeword to activate. We used an attention check question to ensure participants were reading the explanations; those who failed this attention check were excluded from further participation and were not included in the reported data. We did not apply additional screening criteria. In particular, we did not limit the study to participants who had prior experience with voice assistants or who were in the market for a voice assistant; we made this choice because we felt that the unfamiliarity of a continuous listening assistant would make it hard for respondents to predict whether they would be a potential user of one of these devices, and probing the purchase decision in more detail would detract from the focus of our study.

We then introduced the idea of apps, which add functionality to the otherwise featureless base voice assistant. Participants were randomly assigned to learn about one of the three privacy protection methods (described in 4.2). At this point, we asked two comprehension check questions to ensure participants understood that they would need to install apps to use ALVA and how the privacy model of their ALVA version worked. We did not disqualify participants for failing these comprehension checks, but instead asked them to try again. We chose this strategy because we felt that our initial attention check, combined with skimming submissions for gibberish and other responses that obviously did not address questions, was sufficient for weeding out low-effort responders. (Incorrect answers among remaining participants may be indicative of confusion rather than

Fig. 1. Front page of the app store seen by participants as one of the survey pages



cheating.) Our approach helped ensure that participants achieved a baseline level of understanding of the different privacy conditions. Furthermore, later in the survey, we asked participants, "In your own words, please describe the differences between ALVA 1 and ALVA 2." We observed that responses to this question were consistent with an understanding of the differing privacy mechanisms presented in our study.

Participants' next task was to browse the app store and install any apps that they thought they would use if they were gifted an ALVA device. The app store contained various apps such as alarms, music, and a calendar. In total, 18 apps were listed on the store homepage, along with their icon and name. (The complete list of apps can be found in Table 2.) We selected these apps based on literature on existing usage patterns and projected features [7, 39, 52, 53].

The store listed every app as having been created by a "third-party developer" without specifying their name or other attributes. Additionally, all apps requested permissions appropriate to their functionality (i.e., there were no malicious or over-permissioned apps).

Participants could click on an app to view its overview page (Figure 2), which consisted of three sections: a general description of the app, a permissions section that described the app's privacy policy, and an examples section with sample relevant phrases that would trigger a response from ALVA. Details in the "Permissions" section corresponded to the privacy protection method based on the participant's assigned condition.

We recorded the list of apps installed, the pages visited, and the time spent on each page and section. We used this data to ask participants follow-up questions and, after the survey, analyze what participants had been paying attention to.

We did not enforce a minimum amount of time spent browsing the store, nor did we require participants to install any apps. (If they chose to install none, we prompted them if they were sure, in order to prevent accidental submissions.)

After participants indicated that they had installed all the apps they were interested in, we proceeded to ask about their experience browsing the app store. If they had not installed any apps, we asked why this was the case. Otherwise, we randomly selected an app the participant had opened and asked them about their motivation for installing it or for deciding not to install it.

To conclude the first half of the survey, we asked participants how likely they would be to use the hypothetical voice assistant: "If you received an ALVA smart speaker as a gift, how likely would you be to set it up in your home and use it?" Participants could respond

Fig. 2. Screenshots of the app store interface, including app page description, permissions, and examples sections

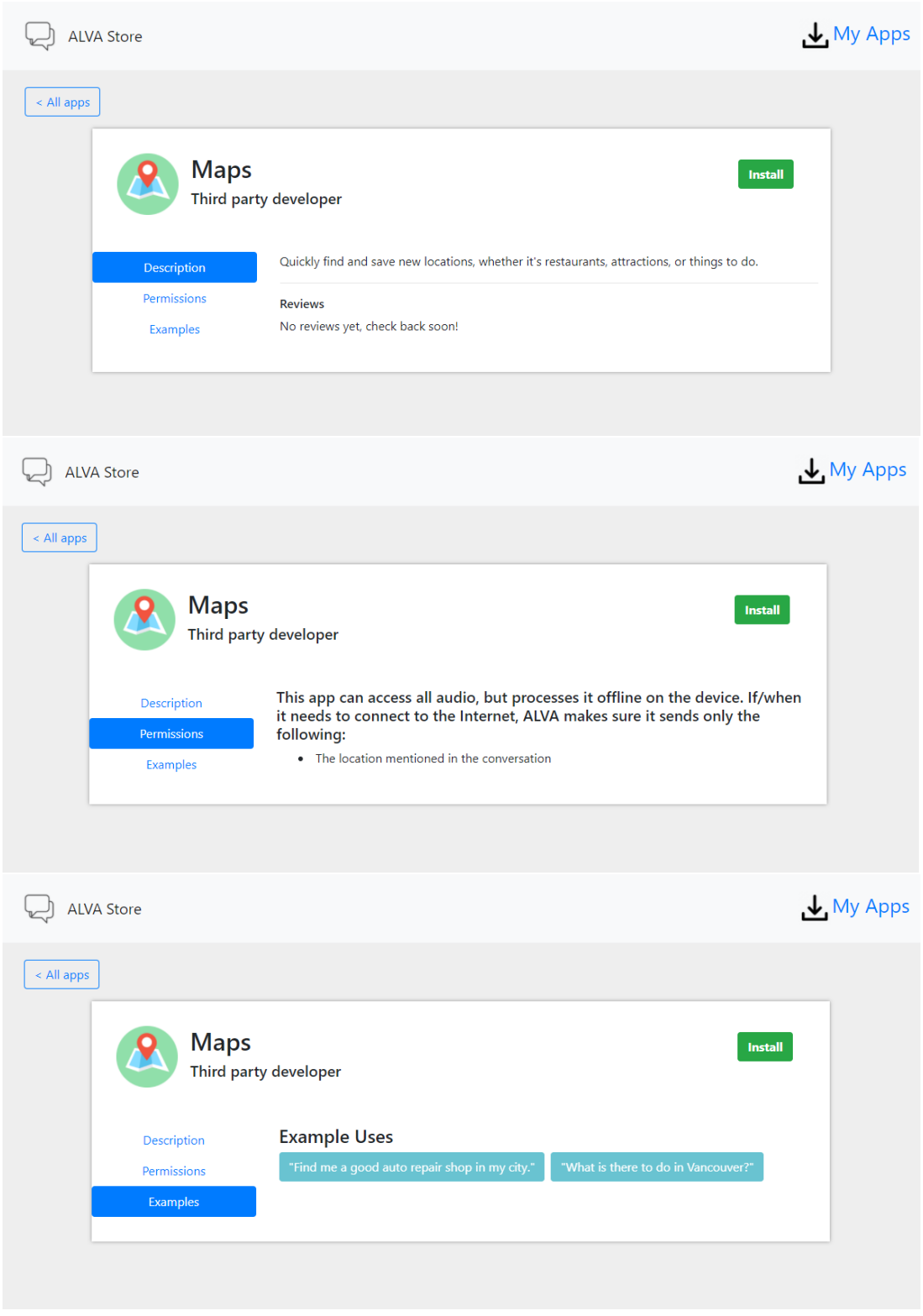


Table 2. ALVA apps: the apps our participants could install, along with their descriptions that were included in the app store

Name	Description
Search	Queries you would ask Google/Siri/etc.
Calendar	Automatically add planned meetings and appointments to your calendar.
Reminders	Detects when you want to remember something and automatically save it, in order to remind you at an appropriate time.
Shopping List	Adds items to your shopping list based on your conversations.
Fridge Monitor	Keep track of items in the refrigerator and when they are likely to expire.
Recipe Search	Helps you quickly find recipes and dictates them to you, hands-free.
Baby Monitor	Detects and alerts you when your child is crying.
Alarm	Creates and sets custom alarms.
Coffee Scheduler	Prepares a custom cup of coffee in advance.
Weather	Get the weather forecast. This app can also sense when a conversation discusses plans that may be affected by the weather, and offer relevant advice.
Scorekeeper	This app keeps track of running total scores. For instance, this app can gamify household chores by keeping track of children's scores.
Swear Jar	This app keeps track of the number of times someone says an expletive. You also have the option of donating a certain amount to a charity of your choice after swearing too many times.
Trivia	This app lets you play a game of trivia against other people or ALVA.
Audiobooks	Play audiobooks out loud. Connect to third-party services to get access to all your digital books. Pick your voice, leave bookmarks, and easily switch to the part you want to hear, all hands-free.
Email	Check and send emails. Connect to your existing email accounts and stay productive, hands-free.
Maps	Quickly find and save new locations, whether it's restaurants, attractions, or things to do.
Meditation Instructor	Whether you need a few minutes to relax or want to practice daily mindfulness, a guided meditation instructor can help make this process smooth and simple. Choose your desired session length and what you want to focus on.
Music	Easily play music from your favorite artist and discover new music. Can tell you what the current song is.

on a five-point Likert-style scale ranging from "very unlikely" to "very likely." We also asked how easy they found it to control information shared with ALVA and third party apps, by rating the following statement on a five-point scale ranging from "strongly disagree" to "strongly agree": "I think I can easily control the information I provide to ALVA and its third party apps." Respondents also had to explain why they felt that way.

After sharing their thoughts about this first iteration of the smart speaker, we asked participants to imagine a different version of the assistant device, ALVA 2, that employed one of the two remaining privacy protection methods (Control, Topic, or Network). They repeated the process of learning about the privacy approach, browsing the app store, and giving their thoughts on the privacy protection. After this second round, participants answered, if applicable, why they installed different apps in Round 1 and Round 2. They again described their likelihood of using ALVA and the perceived effectiveness of the privacy protection.

5.3 Analysis

We formulated several hypotheses about the effect that the availability of privacy controls would have on people's propensity to use the voice assistant.

1. The number of apps installed by participants will be greater in conditions with some sort of privacy protection than those with no privacy protection.
2. Participants will show greater willingness to use ALVA if it includes privacy protections.
3. The effectiveness ratings of the privacy-preserving approaches will be higher than that of the control condition.

To test the first hypothesis, our analysis examined the mean number of apps participants installed for each privacy protection. We used permutation tests [18] to compare the difference in the mean number of apps installed between the three conditions. We chose this test over a Generalized Linear Mixed Model (GLMM) or the Kruskal-Wallis/Mann-Whitney U as it better modeled the statistical assumptions of our study and did not require knowledge of the underlying distribution of the random variables. However, we also verified our results with these more traditional statistical tests.

The null hypothesis was that the number of apps installed was the same in each condition (i.e., had the same distribution). To test this, we separated participants into groups based on round and condition. We chose to analyze the rounds separately, as the data points are not independent, and the order in which the participants receive the conditions has an effect.

To perform the permutation test, we first calculated a base difference in sample means $d = \bar{X}_1 - \bar{X}_2$, with respective sample sizes n_1 and n_2, for each pair of conditions (Control–Topic, Control–Network, Network–Topic). We then pooled the data into a single distribution. For 100,000 iterations, we permuted the data and computed the sample means for the first n_1 observations and the remaining n_2 observations. We then calculated the difference in these sample means, δ_i. The proportion of the permuted differences in sample means greater than our base difference $\delta_i > d$ is our p-value.

To test the second and third hypotheses, we compared how likely participants would be to use ALVA for each condition and how effective participants found each protection by using a Wilcoxon signed-rank test on each possible pair of conditions among all participants who experienced those conditions in any order.

In order to further understand participants' thinking and reasoning, we used thematic analysis [10] to analyze responses to our open-ended questions. Two researchers read every one of the 214 answers to each question (text responses were required, so no participants skipped questions), iteratively defined a codebook based on the major themes, and then independently categorized every response. Because a response could contain multiple sentiments, the codes were not mutually exclusive (i.e., a response could be labeled with multiple codes), so we used Kupper and Hafner's statistic for computing interrater reliability [30].

5.4 Limitations

Several factors could have affected the validity of our results. While we modeled a real-life decision for our participants, its ecological validity is necessarily limited by the nature of the simulation, as well as the fact that the product we are studying is currently hypothetical. Had the device been real, users may have perceived its risks as more tangible. However, responses from participants—in particular, reactions to the Control condition without any protections—suggest that they realized the risks of even a hypothetical device. Nonetheless, participants' choices lacked actual privacy consequences for them. Additionally, people may feel differently if they are actually considering purchasing a voice assistant, and future work may consider investigating whether prospective owners make different choices compared with less likely users.

People's decisions may have also been affected by aspects of the simulation being incomplete or not fully articulated. For example, we attributed apps to a "third-party developer" rather than specific entities people may have been familiar with. We also did not provide information about additional privacy controls voice assistants are likely to carry over from their present-day versions, such as mute buttons, data deletion options, and any sort of app review process.

Furthermore, the specific mechanics of our study may have affected the results. In particular, the two-phase design of our survey, and the task repetition it entailed, may have fatigued some participants, affecting their judgment or recall. Like with all surveys, ours may suffer from experimenter demand effects [59] or social desirability bias [34] (e.g., perceiving the private option as more "acceptable"). However, we note that our methodology focused participants on a decision (installing apps), so the effect would have to manifest in people choosing different—but specific—apps, which is a higher barrier than a typical survey. Respondents' free-text answers further illustrate their thinking.

Most generally, while we seek to tackle the broad problem of privacy for always-listening devices, this study offers a narrow perspective by using the lens of install-time permissions to study this issue. In order to gain a complete picture, it must be complemented by a multifaceted approach that takes into account other aspects of the system and user perspectives on those. Furthermore, the scope of our work may limit generalizability in other ways. For example, since we focused on always-listening assistants, our results *may* be applicable, but should not be assumed to generalize, to the trigger-based voice assistants that are popular today. Because continuously listening assistants are currently hypothetical, any eventual products may have different designs from those we posited; it is therefore possible that our results may apply only to the designs in our study. However, as detailed in §4 and §5, we worked to control, as much as possible, for factors other than the privacy mechanisms. Consequently, we believe that our findings can still yield valuable insights about people's behavior faced with novel devices, even if not all of our assumptions about future voice assistants are realized.

5.5 Participants

We recruited participants from the Prolific platform [44] to take an online survey. In total, we used data from 214 participants. The majority (57%) self-identified as male, the average age was 35, and the average household size was 2.8. Most (87%) of the participants had some prior experience with voice assistants. The survey took approximately 15 minutes to complete, and participants were compensated $3.75 for their time.

Ethics Participants provided informed consent before starting the survey. They were made aware of study procedures (the study involved no deception), the data being collected (survey responses and interactions with the interface; no personally identifying information was obtained), and how it would be handled (all data was collected anonymously, as the recruitment service provides only anonymous identifiers, and stored privately). The study procedures, including data handling practices, were reviewed and approved by our IRB. We therefore believe our study conforms with current best practices for survey-based human subjects research.

6 Results

In this section, we report the results that address each of our research questions.

6.1 Do people install more apps when protections are stronger?

We hypothesized that people would install more apps in conditions in which they are more comfortable with the privacy protection. To account for ordering effects, we separated the data into two phases; we refer to Round 1 as the first privacy approach the participants experienced and Round 2 as the second. As seen in Table 3, the mean number of apps installed in Round 1 across conditions was essentially the same, with a standard deviation of $\sigma_{A1} = 0.123$. However, we begin to observe greater differences across conditions in Round 2, with a standard deviation of $\sigma_{A2} = 0.804$ apps installed.

Table 3. Mean number of apps installed by condition, all participants

	Control	Topic	Network
Round 1	5.5	5.8	5.7
Round 2	3.3	4.7	5.2

Table 4. Were there significant differences in the number of apps installed *by all participants* between each pair of conditions? Permutation test p-values

	Control-Topic	Control-Network	Network-Topic
Round 1	0.48	0.74	0.72
Round 2	0.015	0.0010**	0.31

To assess whether the difference in the number of apps installed in different conditions was significant, we used a permutation test (Table 4). Among all participants, the differences between conditions in Round 1 were not significant. In Round 2, we did find that participants installed significantly more apps in the Topic and Network conditions, as compared with the Control. There were no significant differences between the Network and Topic conditions.

6.2 How does user attention affect install decisions?

Observing the discrepancies between Round 1 and Round 2, we decided to explore whether the telemetry we collected could shed light on what was happening. We hypothesized that differences in behavior could be driven by whether a user examined permissions before installing apps. We therefore defined a heuristic for whether someone looked at permissions: they had to open the Permissions tab for at least one app in either round. Using this metric, we found that only 29% of participants in our study ever looked at any permissions.

Next, we divided participants into subgroups based on whether they had looked at permissions and examined whether these subgroups differed in behavior. Among people who *did look at permissions* (Table 5), the differences in the number of apps installed across conditions in Round 2 are even more pronounced compared to the entire population ($\sigma_{A2} = 1.314$).

Table 5. Mean number of apps installed by participants who viewed permissions ($n = 62$)

	Control	Topic	Network
Round 1	6.0	5.6	5.6
Round 2	2.5	3.8	5.7

Table 6. Mean number of apps installed by participants who did not view permissions ($n = 152$)

	Control	Topic	Network
Round 1	5.4	6.0	5.7
Round 2	3.9	4.8	5.0

The number of apps installed by the remaining participants—those who *never looked at any permissions*—is shown in Table 6. Among them, the differences between the two rounds are smaller, when compared with participants who did review permissions during installation. Furthermore, the differences across conditions in Round 2 are smaller than in the combined population ($\sigma_{A2} = 0.478$). Correspondingly, we found no significant differences between the conditions in either Round 1 or Round 2 (Table 7). We therefore conclude that privacy considerations influenced the app install choices of those who paid attention to permissions, but not of those who skipped them.

Our results so far are somewhat ambiguous: there are not always differences between people's behavior under different privacy conditions. Is it because they do not care about privacy, or is something else going on? We explore this question next.

6.3 Do people prefer privacy protections?

We asked participants how likely they were to use ALVA (Figure 3) and how effective they found the privacy protections (Figure 4) after each round of browsing the store.[1] The majority stated that they would be likely to use the device with the privacy protections in both the Topic and Network conditions; not so in the control condition. Furthermore, more than twice as many found these controls to be effective, compared with the baseline of the control condition.

Table 7. Were there significant differences in the number of apps installed *by participants who did not look at permissions* between each pair of conditions? Permutation test p-values

	Control-Topic	Control-Network	Network-Topic
Round 1	0.18	0.52	0.64
Round 2	0.19	0.14	0.77

1 Five participants were excluded from these questions due to a data collection error.

Fig. 3. Likelihood of using device: *If you received an ALVA smart speaker as a gift, how likely would you be to set it up in your home and use it?* Answers by participants in each condition

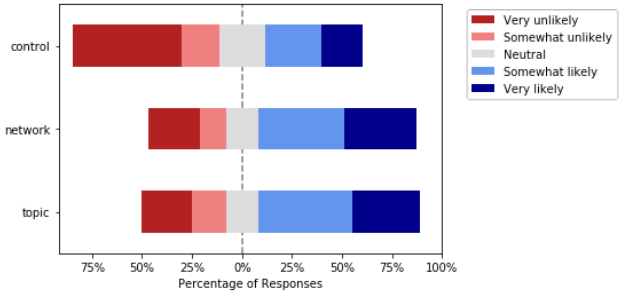


Table 8. Were there differences in effectiveness ratings and willingness to adopt the privacy modes between each pair of conditions? Wilcoxon Signed-Rank Test p-values

	Control-Topic	Control-Network	Network-Topic
Likelihood	$p < 0.001$***	$p < 0.001$***	0.036*
Effectiveness	$p < 0.001$***	$p < 0.001$***	0.13

To verify these results, we tested whether the observed differences between the preferences were statistically significant. Table 8 shows the results of the Wilcoxon signed-rank test for both questions and each pair of conditions. All three condition pairs showed significant differences for the likelihood question. We observed significant differences for the effectiveness question for Control-Topic and Control-Network pairings. On average, participants found the Network protection slightly more effective than the Topic protection in controlling information being accessed by third parties (a difference of 0.25 on a scale of 5 possible discrete ratings), but this difference was not significant.

Fig. 4. Effectiveness: *I think I can easily control the information I provide to ALVA and its third party apps.* Answers by participants in each condition

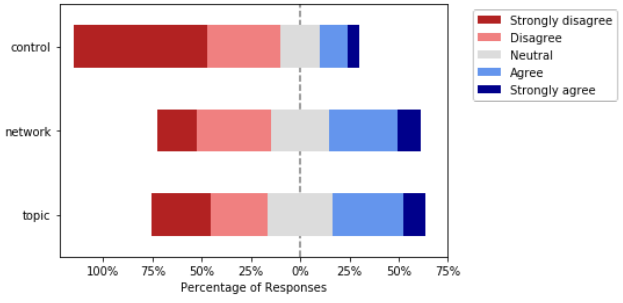


Table 9. Common reasons for privacy protection preferences ($IRR = 0.74$)

Reason	Count	Frequency
Prefer one for increased privacy features	121	57%
Not interested in these devices	24	11%
Neither model has sufficient privacy protections	18	8.4%
Both models do the same thing	17	7.9%
Prefer one for increased sense of security	16	7.5%
Unsure of preference	10	4.7%
Don't like always-listening	10	4.7%
Can't trust third-party developers	6	2.8%
Gave other reason	5	2.3%
Prefer one for being more transparent	4	1.9%
Not interested because inconvenient for end user	3	1.4%
Both models have sufficient privacy protections	3	1.4%
Not worried about privacy	1	0.4%

We asked participants for their preferences between the two privacy approaches they experienced (Figure 5). Most people (52 participants, 69.3%) said they preferred or strongly preferred the Topic privacy protections to the Control (which lacked any protections). A similar proportion (49 participants, 70%) said they preferred or strongly preferred the Network approach to the Control. Between Network and Topic, 48.4% said they preferred Network, 40.6% had no preference, and only 10.9% preferred the Topic approach.

We then asked participants to explain why they preferred (or were indifferent between) the protection methods in Round 1 or Round 2. Table 9 lists the common themes that emerged from our thematic analysis. The majority of participants (57%) attributed their preference to privacy. For example, P28 stated: "I strongly prefer [ALVA with Network controls] as it gives the user more control over what data is monitored, stored, and collected." Respondents who preferred Network to Topic protections typically perceived the former as offering greater privacy, but did not offer concrete details about why they felt more protected: "I am more comfortable with how the ALVA 2 works. It seems like you have more control over your privacy and what is recorded and sent to third party apps" (P193).

The next most common sentiments were a lack of interest in voice assistants ("I wouldn't care for either device because I wouldn't be interested in anything that required as much set up as they do," P104), a feeling that neither approach offered sufficient privacy protections ("I don't particularly like either of these models because of the privacy concerns," P71), and the opinion that the approaches were in effect the same ("I think they are similar enough to say that I can use either one," P215). Only one participant explicitly stated that they were unconcerned with the possibility that their data could be obtained by third parties: "Honestly, I have no fear of info getting out regardless, so whether it hears what I say or not doesn't bother me" (P155).

Table 10. Explanations for installing different apps across conditions ($IRR = 0.85$)

Reason	Count	Frequency
Considered difference in privacy approaches	85	40%
Installed same apps in both stores	41	19%
Mistakenly installed different apps	30	14%
Irrelevant explanation[2]	28	13%
Wanted to try different apps	21	9.8%
Changed mind about installing an app	16	7.5%

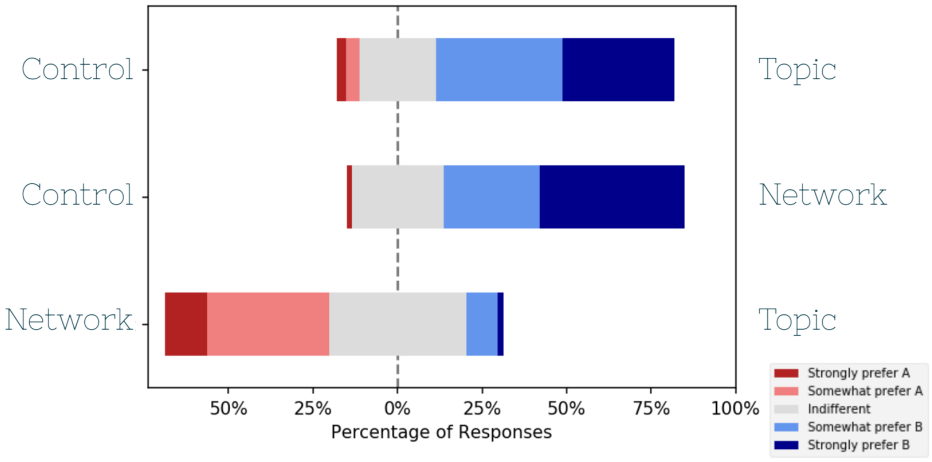
As part of our aim to understand differences between the different privacy conditions, if a participant installed different apps in Round 1 and Round 2, we asked them why this was the case (Table 10). A plurality (40%) of participants cited concern about privacy protections as a reason for installing different apps. These sentiments were common even among those who, according to our analysis, ignored privacy information while browsing the app store: of the 185 participants who did not look at app permissions when installing apps, 55 of them (29.7%) cited privacy as a reason for installing different apps. Other reasons commonly cited for installing different apps included making a mistake wanting to try different apps in different rounds, or changing their minds about downloading an app for reasons other than privacy.

7 Discussion

Our study's interactive app store experience provided a unique means of measuring consumer sentiment in a scenario modeling real life. Using both quantitative and qualitative analysis, we learned about people's views on privacy approaches for continuously listening voice assistants. While we witnessed some outright rejection

2 These answers did not directly answer the question about why their choices were different, e.g., "I just chose what I think I would use on each device" (P37).

Fig. 5. Preference between privacy approaches: Do you prefer Condition A or Condition B?



(see Table 9), many were open to these products, which is consistent with findings in related work [52, 53]. However, a vast majority of our participants agreed that privacy controls were an essential prerequisite for them adopting an always-listening device.

The results of our study demonstrate that, as a whole, people are generally concerned about the privacy protections, or lack thereof, offered by always-listening voice assistants. We recognize that the privacy protections we studied may be simplistic or incomplete, but users still preferred them over nothing. Our findings show that consumers do seek to make choices to protect their privacy when considering new technologies, as demonstrated by the number of apps they installed, and is confirmed by the preferences they expressed after browsing the store.

7.1 People install more apps when there are privacy controls

A key hypothesis of our study was that users will install significantly more apps for their voice assistant when it has some form of privacy protections. Our results provide some evidence in support of our hypothesis, as the participants in the Network and Topic conditions installed significantly more apps than the Control condition in Round 2. On average, participants in Round 2 installed the most apps in the Network condition, followed the Topic condition, and the fewest in the Control.

Yet, the picture is complicated by the absence of a similar effect in Round 1, where participants installed approximately equal numbers of apps across conditions. This is especially concerning as the first round, where participants had no prior experience with the app store, may be more representative of naive users.

7.2 Privacy choices may depend on focusing effects

In our results, we observed that only in Round 2 did participant behavior differ based on the privacy controls, whereas they had no effect in Round 1. Why was this the case? There may be several possible explanations of this phenomenon. One hypothesis is that, at first, people focus on the exploration task—choosing apps that are interesting to them. Privacy does not factor into their decision, either because it is not top-of-mind, or because they were told to assume that they already have the always-listening device, and they may believe that installing apps may not significantly alter their privacy exposure. Round 2, by virtue of only differing in the device's privacy controls, may bring those front and center—priming people to think about privacy [13]. Alternately, the effects may be deliberate rather than unconscious: the highlighting of the privacy controls may cause participants to be more careful or reflective about their choices, where they had not been previously [2]. Finally, some participants may not have comprehended the privacy scenario presented in the first round, but this could have changed in the second round, when

they could mentally compare the second privacy scenario to the first one. Future research could test these suppositions, for example by interrogating participants' assumptions in greater detail, or emphasizing privacy choices and their effects earlier (i.e., in Round 1) and examining the effects of doing so.

7.3 Most do not pay attention to permissions

Another open question is whether Round 1 or Round 2 is more reflective of people's real-world behavior; regardless of the answer, a clear implication of our study is that many people skip—and therefore miss—privacy information available when installing apps. As part of our analysis, we divided people into groups based on whether they examined individual apps' permissions. Less than a third of our participants (29%) fell into the group that did look at permissions, but it was among them that we saw significant differences in the number of apps installed between privacy conditions (Table 3). In contrast, those who skipped the permissions, while accounting for the majority of our participants, showed no significant effects in their data. While one might think that these people simply did not care about their privacy, their open-ended responses suggest that this is not the case. The discrepancy between people's stated preferences and their actual behavior is a well-studied phenomenon in privacy literature [26], and it is often explained by a lack of available or usable alternatives [50]. Accordingly, we conclude that it is specifically the user experience of install-time permissions that leads people to skip them, rather than a lack of privacy preferences or concerns. This confirms prior work studying install-time permissions in smartphone app stores, which likewise found low levels of attention, engagement, and understanding [17, 23], resulting in recommendations against this privacy mechanism [16, 47] and the switch among smartphone platforms to runtime permissions [11, 19]. Our results may therefore hold implications for existing voice assistant app stores, which currently rely on install-time permissions [5].

7.4 Privacy controls improve the perception of the assistant

We investigated how likely participants would be to use the continuously-listening assistant in each condition. The results support our original hypothesis that privacy controls make people more willing to use always-listening assistants, as participants were significantly more likely to say that they would use ALVA when it featured the Network or Topic protections, as compared with the Control condition. We conclude that privacy protections not only increase people's comfort in installing individual apps, but also improve the general perception of the device itself.

We also investigated how effective participants found each privacy approach. Again, the results support our original hypothesis, as participants thought they could more easily control the information they shared with ALVA under the Network and Topic approaches.

7.5 People show slight preference for offline-first voice assistants

Between the two privacy protection methods we tested in our study, participants found the Network-restricted approach to be more effective than the Topic approach. However, they were not significantly more likely to say that they would use the Network model over the Topic model, and the mean number of apps installed was also not significantly different between the two. This similarity indicates that neither approach is robust enough by itself to be a clear winner. A more complete and granular solution is required to protect user privacy while maximizing device capabilities.

8 Future work

A number of open questions remain about the best ways to protect user privacy in continuously-listening voice assistant ecosystems. Our work has demonstrated a clear demand for privacy solutions, but it has only begun to explore which solutions may be optimal. The approaches used as examples in our study—topic-based restrictions and limits on network communication—showed promise. However, future work could refine and iterate on these ideas to produce more concrete and deployable privacy controls.

Another natural next step is to apply our study's methodology to explore the acceptability of other privacy-enhancing techniques for continuously listening voice assistants. After comparing many different privacy mechanisms, the most favored one could be determined. Different protections can also be combined to create stronger, more complete privacy solutions.

A major observation from our study is that the most common sentiments show concern about how privacy might be violated, but the majority of participants did not look at app permissions. Exploring this disconnect is another area for future work. Why did participants not engage with the permission information? To what extent is this an artifact of the study, versus a behavior that would be replicated in the real world? Especially in the latter case, research needs to focus on more effective ways of presenting privacy information to users, either by making it more understandable and actionable, or by exploring alternate modalities and timing, such as runtime permissions [40].

9 Conclusion

To explore people's privacy expectations for continuously listening voice assistants, we designed an interactive voice assistant app store experience to use in concert with a traditional survey. Participants indicated a preference for voice assistants with privacy protections and, in aggregate, installed more apps for them—though only after their privacy properties were foregrounded. Furthermore, despite over half of participants indicating that privacy was a reason for preferring one of the two protections they were assigned, the majority of participants did not look at the permissions for the apps they installed. We conclude that privacy is an important feature for potential device users, and the protection approaches we tested—topic modeling and restrictions on network communications—show promise. However, install-time permissions suffer from usability problems, including lack of user attention, making them a potentially inadequate choice for voice assistant app stores.

Acknowledgments

We are grateful to Runjing (Bryan) Liu and Jake Soloff for their advice on statistical methods, Conor Gilsenan for edits to the paper, and David Wagner for feedback on a draft. We would also like to thank members of the BLUES lab for feedback on the study design and Serge Egelman for support. This work was supported by NSF grant CNS-1801501 and the Center for Long-Term Cybersecurity at UC Berkeley.

References

[1] 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.

[2] Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian Schaub, Manya Sleeper, Yang Wang, and Shomir Wilson. Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online. *ACM Comput. Surv.*, 50(3):44:1–44:41, August 2017.

[3] 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.

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

[5] 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.

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

[7] Tawfiq Ammari, Jofish Kaye, Janice Y. Tsai, and Frank Bentley. Music, search, and IoT: How people (really) use voice assistants. *ACM Transactions on Computer-Human Interaction*, 26(3), April 2019.

[8] Samuel Axon. Amazon Halo will charge a subscription fee to monitor the tone of your voice. *Ars Technica*, August 2020.

[9] Todd Bishop. Amazon maintains big lead over Google and Apple in U.S. smart speaker market, new study says. *GeekWire*, August 2021.

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

[11] 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.

[12] 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.

[13] Isis Chong, Huangyi Ge, Ninghui Li, and Robert W. Proctor. Influence of privacy priming and security framing on mobile app selection. *Computers & Security*, 78:143–154, September 2018.

[14] Catalin Cimpanu. Academics smuggle 234 policy-violating skills on the Alexa Skills Store. *ZDNet*, July 2020.

[15] 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.

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

[17] 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.

[18] R.A. Fisher. *The Design of Experiments*. Oliver & Boyd, Edinburgh, 1935.

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

[20] Zhixiu Guo, Zijin Lin, Pan Li, and Kai Chen. SkillExplorer: Understanding the behavior of skills in large scale. In *29th USENIX Security Symposium (USENIX Security 20)*, pages 2649–2666. USENIX Association, August 2020.

[21] 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.

[22] Md Tamzeed Islam, Bashima Islam, and Shahriar Nirjon. SoundSifter: Mitigating overhearing of continuous listening devices. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '17, pages 29–41, New York, NY, USA, 2017. Association for Computing Machinery.

[23] 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.

[24] 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.

[25] Sean Michael Kerner. Researchers Find Amazon Alexa Can Be Hacked to Record Users. *eWeek*, April 2018.

[26] Spyros Kokolakis. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, 64:122–134, January 2017.

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

[28] 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.

[29] 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.

[30] Lawrence L. Kupper and Kerry B. Hafner. On Assessing Interrater Agreement for Multiple Attribute Responses. *Biometrics*, 45(3):957, September 1989.

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

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

[33] 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.

[34] Paul Lavrakas. Social Desirability. In *Encyclopedia of Survey Research Methods*. Sage Publications, Inc., Thousand Oaks, California, 2008.

[35] Yuchen Liu, Ziyu Xiang, Eun Ji Seong, Apu Kapadia, and Donald S. Williamson. Defending Against Microphone-Based Attacks with Personalized Noise. *Proceedings on Privacy Enhancing Technologies*, 2021(2):130–150, April 2021.

[36] Kim Lyons. Fitbit may soon be adding snoring detection to its devices. *The Verge*, May 2021.

[37] 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.

[38] 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.

[39] Nathan Malkin, Serge Egelman, and David Wagner. Privacy controls for always-listening devices. In *Proceedings of the New Security Paradigms Workshop*, NSPW '19, pages 78–91, New York, NY, USA, 2019. Association for Computing Machinery.

[40] Nathan Malkin, David Wagner, and Serge Egelman. Runtime Permissions for Privacy in Proactive Intelligent Assistants. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*, Boston, MA, August 2022. USENIX Association.

[41] 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.

[42] 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.

[43] 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.

[44] Eyal Peer, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70:153–163, May 2017.

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

[46] Mark Savage. Spotify wants to suggest songs based on your emotions. *BBC News*, January 2021.

[47] 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.

[48] Faysal Hossain Shezan, Hang Hu, Jiamin Wang, Gang Wang, and Yuan Tian. Read between the lines: An empirical measurement of sensitive applications of voice personal assistant systems. In *Proceedings of the Web Conference 2020*, WWW '20, pages 1006–1017, New York, NY, USA, 2020. Association for Computing Machinery.

[49] 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.

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

[51] Iraklis Symeonidis, Gergely Biczók, Fatemeh Shirazi, Cristina Pérez-Solà, Jessica Schroers, and Bart Preneel. Collateral damage of Facebook third-party applications: A comprehensive study. *Computers & Security*, 77:179–208, August 2018.

[52] 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.

[53] 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.

[54] Jing Wei, Tilman Dingler, Enying Gong, Brian Oldenburg, and Vassilis Kostakos. Proactive smart speakers for chronic disease management: Challenges and opportunities. *"Mapping Grand Challenges for the Conversational User Interface Community" workshop, CHI 2020*, 2020.

[55] 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.

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

[57] Nan Zhang, Xianghang Mi, Xuan Feng, XiaoFeng Wang, Yuan Tian, and Feng Qian. Understanding and Mitigating the Security Risks of Voice-Controlled Third-Party Skills on Amazon Alexa and Google Home. *arXiv:1805.01525 [cs]*, May 2018.

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

[59] Daniel John Zizzo. Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1):75–98, March 2010.

A Survey instrument

1. Which of the following voice assistants do you use regularly (at least several times a week)?
   - Amazon Alexa
   - Apple Siri
   - Google Assistant
   - Microsoft Cortana
   - Samsung Bixby
   - Other assistant
   - I don't use any intelligent voice assistants
2. How do you usually interact with your voice assistant?
   - Through my smartphone
   - Through a smart speaker (such as Amazon Echo or Google Home)
   - Through a smart watch (Apple Watch, etc.)
   - In my car
   - Through another device that has the voice assistant built in
   - I don't use any intelligent voice assistants
3. If you received a smart speaker as a gift, how likely would you be to set it up in your home and use it?
   - Very unlikely
   - Somewhat unlikely
   - Neutral
   - Somewhat likely
   - Very likely
4. If a close friend or family member received a smart speaker as a gift, how likely would you be to recommend that they set it up in their home and use it?
   - Very unlikely
   - Somewhat unlikely
   - Neutral
   - Somewhat likely
   - Very likely

5. Imagine that there's a new voice assistant on the market: ALVA. Unlike today's devices, you don't need to say specific words to wake up ALVA, because it is always ready to help you. ALVA can also provide services and suggestions based on conversations you have with other people in your household. For example:
   - Instead of saying "hey ALVA, tell my smart light bulb, turn on," you could just say, "turn on the lights."
   - Instead of saying "Ok ALVA, order some pizza for delivery," you could say, "let's order pizza," and ALVA could prepare to order for you and ask you for confirmation.
   - Instead of saying "ALVA, what's the weather in Aspen?", ALVA could automatically respond when you say, "What's the weather in Aspen?"
6. Based on the description above, which of the following is true of the device in this survey?
   - The device is always on and can provide services and recommendations based on your current conversation, even if you don't say its name.
   - The device reacts to your conversation only when explicitly addressed, for example when you say a specific word, such as "Alexa", "Siri", "Ok Google."
   - The device adds a video streaming feature and allows you to watch your favorite movies and shows on a big screen.
   - The device is waterproof and can be used in a bathroom or swimming pool.
7. How useful do you think you would find ALVA's functionality?
   - Not at all useful
   - Not very useful
   - Neutral
   - Somewhat useful
   - Very useful
8. Please explain your answer [free response]
9. **ALVA's apps.** Similarly to how you download different apps on your phone's app store, you can download new apps to add functionality to ALVA by visiting the ALVA App Store. *ALVA does not come with any built-in functionality.*

   Apps in the ALVA App Store, like smartphone apps in Apple's App Store or Google Play, are usually created by third-party developers who are not affiliated with the manufacturers of ALVA. Because of this, ALVA cannot guarantee that apps work as described.
10. **Comprehension check 1:** Based on the explanation above, which of the following is true of ALVA?
    - All of the smart features are already built in to ALVA. You never need to download anything new.
    - To add features to ALVA, you need to download apps from the App store. Apps are created by third-party developers.
    - A special vetting process guarantees that the behavior of each app matches its description.
    - I haven't read the explanation above about ALVA and third-party apps.
11. **How ALVA protects your privacy** [condition-specific explanation and **Comprehension check 2**]
12. For the next part of the survey, you will be *role-playing* that you received a new ALVA device and are *looking to install* some new apps to add features to the device. The ALVA App Store works just like your phone's app store. You can browse apps, click on apps to learn more about them, install, and uninstall apps. Please install any apps that you think you would install if you owned an ALVA device, regardless of your current opinion on smart speakers.
13. *[Interactive component]*
14. You installed an app named _. Please tell us why.
15. If you received an ALVA smart speaker as a gift, how likely would you be to set it up in your home and use it?
    - Very unlikely
    - Somewhat unlikely
    - Neutral
    - Somewhat likely
    - Very likely
16. Imagine that you have already begun using ALVA. Please rate how much you agree with the following statement: "I think I can easily control the information I provide to ALVA and its third party apps.
    - Strongly disagree
    - Disagree
    - Neutral
    - Agree
    - Strongly agree
17. Please explain your choices for the two questions above.
18. *[Interactive component 2]*
19. You installed the following apps for ALVA 1 (but not for ALVA 2): _. And you installed the following apps for ALVA 2 (but not ALVA 1): _. Please explain why you chose to install different apps on these devices.

20. If you received an ALVA 2 smart speaker as a gift, how likely would you be to set it up in your home and use it?
    - Very unlikely
    - Somewhat unlikely
    - Neutral
    - Somewhat likely
    - Very likely
21. Imagine that you have already begun using ALVA 2. Please rate how much you agree with the following statement: "I think I can easily control the information I provide to ALVA 2 and its third party apps.
    - Strongly disagree
    - Disagree
    - Neutral
    - Agree
    - Strongly agree
22. Please explain your choices for the two questions above.
23. In your own words, please describe the differences between ALVA 1 and ALVA 2.
24. Do you prefer ALVA 1 or ALVA 2?
    - Strongly prefer ALVA 1
    - Prefer ALVA 1
    - Indifferent between ALVA 1 and ALVA 2
    - Prefer ALVA 2
    - Strongly prefer ALVA 2
25. Please explain your answer to the question above: why do you prefer ALVA 1 or ALVA 2 (or are indifferent)?
26. How did the ability to explore the App Stores affect your perception of ALVA? Do you have any further comments about the App Store experience or interface?
27. What is your gender?
28. What is your age?
29. How many people are in your household?
30. How many children under the age of 18 are in your household?