Who's Listening?
Analyzing Privacy Preferences in Multi-User Smart Personal Assistants Settings

Carolina Carreira
Carnegie Mellon University
IST University of Lisbon INESC-ID

Cody Berger
Carnegie Mellon University

Khushi Shah
Carnegie Mellon University

Samridhi Agarwal
Carnegie Mellon University

Yashasvi Thakur
Carnegie Mellon University

McKenna McCall
Carnegie Mellon University

Nicolas Christin
Carnegie Mellon University

Lorrie Faith Cranor
Carnegie Mellon University

1 Motivation

Smart Personal Assistants (SPAs) are internet-connected devices that can respond to spoken commands verbally and through actions. Popular SPAs include Amazon Echo, Google Home, and Apple Homepod. SPAs may be able to answer questions using the internet, play music, control other internet-connected devices, make phone calls and more. While SPAs are often situated in communal environments such as homes and workplaces, commercially available SPAs adapt poorly to multi-user environments [4]. SPAs in multi-user environments must accommodate multiple user groups: *primary users*, which in this paper is defined as the SPA owner(s); and *secondary users*, who do not own the SPA but may use the SPA intentionally or incidentally, including children, guests, and others. While previous studies have investigated how SPAs affect privacy in multi-user environments [3, 8, 9] and primary users' privacy preferences and concerns in multi-user environments [7], secondary users' privacy preferences and access control preferences remain underresearched.

This gap means existing privacy frameworks and SPA mechanisms do not adequately cater to the diverse privacy needs of different user groups (e.g., children, guests, incidental users) within the same device context. This work-in-progress paper bridges this gap with a 90-participant survey of SPA users and non-users and an expert evaluation on commercially available SPAs to determine how pre-existing privacy settings cater to different user groups' privacy preferences. We aimed to answer the following research questions: **RQ1.** Do secondary users have different privacy preferences than primary users? If so, how are they different? and **RQ2.** How do existing smart home assistants' privacy settings align with user preferences?

USENIX Symposium on Usable Privacy and Security (SOUPS) 2024.
August 11–13, 2024, Philadelphia, PA, United States.

2 Methodology

We designed and deployed an online survey of 90 participants, targeting SPA users and non-users, and measured privacy preferences using techniques from contextual integrity. Contextual integrity can be used to describe information flows with specific parameters: data subject, sender, recipient, information type, and transmission principles. We follow a methodology similar to Abdi et al. [1] but focus on the first four contextual integrity parameters. We iteratively designed explanations for each one of the nine attributes in our scenarios. The attributes are common privacy settings users may want to change in their SPA, including voice recording or activity history (see Table 3). Using the data from our survey, we also performed heuristic-based expert evaluations inspired by Habib and Cranor [5] on commercially available SPAs to determine how pre-existing privacy settings cater to different user groups' privacy preferences.

Recruitment We recruited 90 participants through Prolific with diverse ages, genders, races, and educational backgrounds. The study and consent form were approved by the Carnegie Mellon University IRB.

2.1 Survey Design

In the main part of the survey, we measured participants' privacy preferences. This section was primarily based on Abdi et al.'s work, which uses contextual integrity to build a survey to measure [1] users' privacy preferences. We built scenarios and asked closed-ended questions using Likert scales. In this part of the survey, we also asked an open-ended question about users' privacy concerns when using SPAs.

Scenarios Based on the contextual integrity framework, we created scenarios exploring all combinations of data, attributes, and recipients as shown in Tables 1 to 3. We con-

structed 36 scenarios, categorized into two main groups to represent different user perspectives within the SPA environment: the primary and secondary users.

For the *primary user* group, each scenario positioned the users' SPA as the sender of information, with the participants serving as the subject. The data attributes can be seen in Table 3. Recipients of the data, detailed in Table 2, included close personal connections like partners or children. Additionally, each scenario for the primary user incorporated a version focused on the recipient *learning* and another on *changing* information about the primary user or about a setting in the SPA, allowing us to assess different aspects of data protection within the SPA's operational context. We provided participants with a 5-point Likert scale for each recipient.

"Assume you own a voice assistant, e.g., Amazon Alexa/Google Assistant. How acceptable is it for the smart home assistant [Attribute] to be shared with the following recipients: [Primary User Recipients]"

Conversely, the scenarios for the *secondary user* group featured a shift in dynamics, where the sender was still the SPA, but the subject was not the participant (see Table 2).

"How important is it for you to be able to change the [Attribute] on a smart voice assistant, assuming that the smart voice assistant is owned by:[Secondary User Recipients]"

We randomized primary or secondary user scenarios, the order of data attributes, and the list of potential recipients presented within each scenario. Given the extensive number of scenarios developed (36), participants were only presented with 16 scenarios during the survey: 8 scenarios related to the primary user and another 8 related to the secondary user.

2.2 Analysis.

To measure privacy preferences, we conducted quantitative and qualitative analyses of the results of this survey. To conduct the qualitative analysis, we blind double-coded the response data. We went through three rounds of coding and refining the codebook. Our final codebook in Table 5 includes 24 codes, categorized by thematic similarities and overall sentiment. Our quantitative data analysis aimed to explore user responses and identify patterns based on relevant survey variables. Responses were categorized based on these variables, and we used descriptive statistics to analyze the data due to the limited number of participants. We plan to use a binary regression model, and considering our predictors, we will expand our work to a minimum of 363 participants (considering the 15:1 ratio for regressions [6]).

We also conducted heuristic-based expert evaluations inspired by Habib and Cranor [5] to analyze the privacy settings of two widely-used SPAs: Google Home and the Apple HomePod. These evaluations focused on assessing specific attributes related to user interaction with these devices, such as voice recording and activity history (see Appendix B).

3 Preliminary Results

Participants have many concerns about SPAs. We coded participant responses (see Table 5) to the open-ended question, "*When you think about using smart home assistants in your home, what security concerns come to mind, if any?*". In Figure 1, we show the frequency of negative sentiment codes for SPA and non-SPA owners. Our results show that participants have concerns about privacy and surveillance. Participants expressed unease about SPAs listening to everything they say (*"I am always concerned that someone could be listening in"* (P32)). Another source of concern is related to unauthorized information sharing and access. This includes fears about data breaches, hacking, and the unauthorized selling of personal data (*"I am concerned my information, conversations, location, etc. can be shared."* (P36)). Some participants also mentioned stalking and abuse of power as concerns with SPAs (*"If a woman is married to an abusive man (...) can he locate her through a device that she buys after leaving him?"* (P44)).

Users seem to have different preferences for different groups. Our results suggest that participants do not wish to share control in the primary user scenario for most settings. However, sharing is more acceptable for close relationships like partners, children, and family members. Despite these results for the primary user, when participants encountered the secondary user scenario, they still showed an interest in knowing and changing settings (e.g., for voice recording), particularly if the owner was their child, housemate, or partner. This aligns with prior work findings highlighting the context-dependent nature of privacy preferences [2]. A fundamental desire across user groups is retaining control over modifying privacy settings, even within trusted relationships, while wishing for control when they are secondary users.

Our results suggest a misalignment between the privacy controls offered by current SPAs and the privacy preferences articulated by users. The expert evaluation uncovered inconsistencies and uncertainties regarding the availability and functionality of privacy settings across SPAs. Certain attributes, such as location data and third-party skills (thrid party apps that connect to the SPA) privacy settings, lacked clear and accessible multi-user controls. This lack of transparency and consistency can contribute to user concerns and mistrust, as highlighted by the qualitative analysis of participants' security concerns.

Future work Our research explored the complex world of privacy preferences around SPAs in shared living spaces. We plan to expand the participant pool so that we may conduct further statistical analyses. We also intend to expand our expert evaluations to include Amazon Alexa.

Acknowledgments This work was partially funded by the Portuguese Foundation for Science and Technology through the CMU Portugal Program (PRT/BD/153739/2021).

References

[1] N. Abdi, X. Zhan, K. M. Ramokapane, and J. Such. Privacy norms for smart home personal assistants. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–14, May 2021.

[2] 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):1–23, June 2018.

[3] George Chalhoub, Martin J Kraemer, Norbert Nthala, and Ivan Flechais. "It did not give me an option to decline"': A longitudinal analysis of the user experience of security and privacy in smart home products. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. ACM, 2021.

[4] 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*, page 268. ACM, 2019.

[5] Hana Habib and Lorrie Faith Cranor. Evaluating the usability of privacy choice mechanisms. In *SOUPS '22*, 2022.

[6] Frank E Harrell. Regression modeling strategies. *R package version*, pages 6–2, 2012.

[7] Weijia He, Nathan Reitinger, Atheer Almogbil, Yi-Shyuan Chiang, Timothy Pierson, and David Kotz. Contextualizing interpersonal data sharing in smart homes. *Proceedings on Privacy Enhancing Technologies*, 2024:295–312, 04 2024.

[8] Y. Huang, B. Obada-Obieh, and K. 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*, 2020.

[9] Eric Zeng and Franziska Roesner. Understanding and improving security and privacy in multi-user smart homes: a design exploration and in-home user study. In *28th USENIX Security Symposium (USENIX Security 19)*, pages 159–176, 2019.

A Explanations

We iteratively designed explanations for each attribute (see Table 3) in our scenarios. We built the explanations with the following structure:

Table 1: Table with the parameters for the primary and secondary user questions.

User	**Primary User**	**Secondary User**
Sender	Users' SPA	Subjects' SPA
Subject	The participant	Table 2
Attribute	Table 3	
Recipients	Table 2	The participant
Version	View Access, Modification Access	

Table 2: Table with the respective subjects or recipients for the Primary User and Secondary User questions.

	Primary U.	Secondary U.
your partner	X	X
your children	X	X
your housemates	X	X
your neighbors	X	X
housekeeper	X	
close friends	X	X
close family	X	X
your landlord	X	X
law enforcement	X	
advertising agencies	X	
Total	**10**	**7**

Table 3: Table with the 9 attributes used in the survey.

	Attribute
VR	Voice Recording History
AH	Activity History
UD	Controlling Smart Devices
CD	Changing Smart Devices
AL	Smart Home Assistant Location Data
UL	User Location Sharing
MC	Microphone Controls
TP	Third-Party Skill/Action Privacy
AA	Automation

[EXPLANATION OF CONCEPT + EXAMPLE + EXPLANATION OF CONCEPT SETTINGS WITH EXAMPLE]

For example, for the "Voice Recording History" the final explanation was:

EXPLANATION OF CONCEPT: Voice recordings are audio recordings of things you've said to a smart assistant,

EXAMPLE: Like an audio recording of you asking a smart home assistant to play a song.

EXPLANATION OF CONCEPT SETTING: Voice history settings may allow you to view voice recordings made by the smart home assistant or delete voice recordings from the smart home assistant's history.

Before answering the scenario questions, all participants saw the respective attribute explanation. We avoided jargon and tried to keep the explanations short, with less than four sentences. See below all explanations:

A.1 Voice Recording History

EXPLANATION OF CONCEPT: Voice recordings are audio recordings of things you've said to a smart assistant,

EXAMPLE: Like an audio recording of you asking a smart home assistant to play a song.

EXPLANATION OF CONCEPT SETTING: Voice history settings may allow you to view voice recordings made by the smart home assistant or delete voice recordings from the smart home assistant's history.

A.2 Activity History [activity history settings]

EXPLANATION OF CONCEPT: Activity history is a transcript of things you've done with a smart assistant,

EXAMPLE: Like having asked about the weather.

EXPLANATION OF CONCEPT SETTING: Activity history settings may let you view past interactions with the smart home assistant or delete past interactions with the smart home assistant from its history.

A.3 Use External Smart Devices

EXPLANATION OF CONCEPT: Smart devices are internet-connected everyday devices. You may be able to control these devices hands-free using only your voice with a smart home assistant.

EXAMPLE: One example of a smart device is a smart light bulb, which you can turn on/off hands-free using only your voice.

EXPLANATION OF CONCEPT SETTING: Smart device settings for using devices may allow you to specify which smart device skills can be triggered through the smart home assistant (e.g., who can turn lights on and off when talking with your assistant, or who can also ask the smart home assistant to change their color)

A.4 Configure External Smart Devices

EXPLANATION OF CONCEPT: Smart devices are internet-connected everyday devices. You may be able to control these devices hands-free using only your voice with a smart home assistant.

EXAMPLE: One example of a smart device is a smart light bulb, which you can turn on/off hands-free using only your voice.

EXPLANATION OF CONCEPT SETTING: Smart device settings for configuring devices may allow you to select which smart devices can be used through your smart home assistant, by which users (e.g., who can turn on your light when talking with your assistant).

A.5 Smart Home Assistant Location Data [smart home assistant location data settings]

EXPLANATION OF CONCEPT: A smart home assistant can know its own location,

EXAMPLE: For example the town, zip code, or map coordinates describing where the smart assistant is to provide more accurate information to the user about weather and news.

EXPLANATION OF CONCEPT SETTING: Location data settings may allow you to control if your smart assistant is allowed to know its location.

A.6 User Location Sharing [user location sharing settings]

EXPLANATION OF CONCEPT: Location sharing lets your smart assistant know where people are,

EXAMPLE: For example, a person's current location so you can know how long it will take for them to get home.

EXPLANATION OF CONCEPT SETTING: Location-sharing settings may allow you to control if your smart home assistant is allowed to know your location or other peoples' locations.

A.7 Microphone Controls [microphone controls settings]

EXPLANATION OF CONCEPT: When a smart home assistant's microphone is on, it is constantly listening for a call,

EXAMPLE: Like when you say "Hey Google" to ask for a timer.

EXPLANATION OF CONCEPT SETTING: Microphone control settings allow you to decide when your smart assistant

can listen to you, for example, enabling you to turn off the microphone when you want privacy.

A.8 Third-Party Skill/Action Privacy [third-party skill/action privacy settings]

EXPLANATION OF CONCEPT: Your smart home assistant and its software are developed by a company. Third-party skills or actions are features for your smart assistant developed by other companies, which you can add to your smart home assistant,

EXAMPLE: Like a workout application that you can use through your smart home assistant.

EXPLANATION OF CONCEPT SETTING: Third-party skill/action privacy settings may allow you to manage how these additional features use your information and what they are allowed to do, such as allowing a recipe app to know only your dietary preferences but not your full search history.

A.9 Automation [automation information]

EXPLANATION OF CONCEPT: Automations or routines are custom setups that make your smart devices work together automatically based on certain triggers

EXAMPLE: Like your smart lights automatically turning on when it senses you are getting close to your house.

EXPLANATION OF CONCEPT SETTING: Automation/routines settings let you create, manage, and delete these setups, giving you control over what happens and when. For example setting your coffee maker to start when your morning alarm goes off.

B Expert Evaluations

As part of our methodology, we conducted heuristic-based expert evaluations inspired by Habib and Cranor [5] to analyze the privacy settings of two widely used SPAs: Google Home and the Apple HomePod. These evaluations focused on assessing specific attributes related to user interaction with these devices, such as voice recording and activity history (see all attributes in Table 3). For each attribute, we established a set of two heuristics:

Hr1. Existence of Setting We first determined whether a privacy setting related to the attribute exists within the SPA's configuration options.

Hr2. Granularity of Control We then assessed whether the setting provided fine control for multiple users. This involved checking if the SPA allowed different settings or permissions for other users.

Each attribute for both SPAs was evaluated against these heuristics and coded accordingly: "exists" if the setting was present, "does not exist" if absent, or "unsure/not applicable" if the situation was ambiguous or the heuristic did not apply to the particular attribute. It is important to remember that this was not an exhaustive search. Workarounds may enable users to control a setting we categorized as nonexistent.

Table 4: Expert evaluation results see Appendix B for the full descriptions of the heuristics and details about the method used.

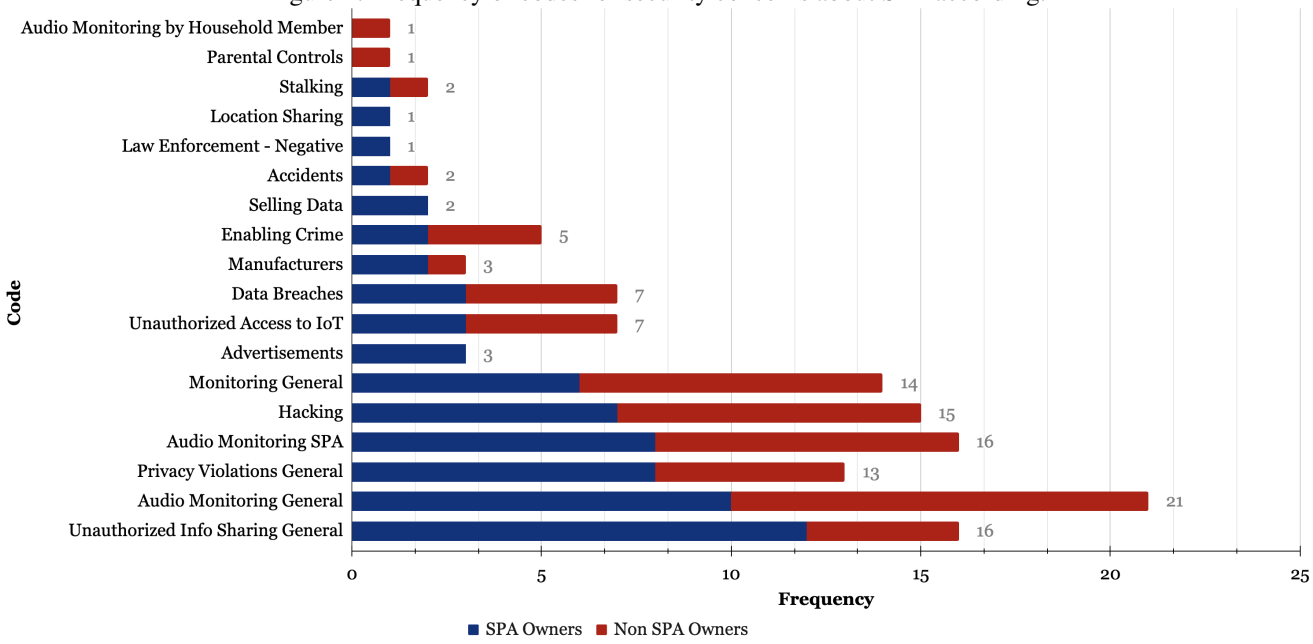
	Homepod	**Google Home**
	VR	
Hr1.	Does not exist	Exists
Hr2.	Does not exist	Does not exist
	AH	
Hr1.	Does not exist	Exists
Hr2.	Does not exist	Does not exist
	UD	
Hr1.	Exists	Exists
Hr2.	Exists	Exists
	CD	
Hr1.	Exists	Exists
Hr2.	Exists	Exists
	AL	
Hr1.	Unsure/NA	Unsure/NA
Hr2.	Unsure/NA	Unsure/NA
	UL	
Hr1.	Unsure/NA	Exists
Hr2.	Unsure/NA	Unsure/NA
	MC	
Hr1.	Exists	Exists
Hr2.	Does not exist	Exists
	TP	
Hr1.	Unsure/NA	Exists
Hr2.	Unsure/NA	Exists
	AA	
Hr1.	Exists	Exists
Hr2.	Exists	Exists

The Google Home evaluated was the Google Nest Mini 2nd Generation, controlled via the Google Home app version 3.16.64 on an iPhone 13 running iOS 17.4.1. Similarly, the Apple Home Pod's settings were managed through the 'Home' app on an iPhone 15, also operating on iOS 17.4.1.

B.1 SPAs privacy settings

A few distinct patterns seem to emerge in the availability of privacy controls. For voice recording (VR) and activity history (AH) attributes, Google Home had settings more consistently than the Apple HomePod. Both devices, however, lacked fine control for AH. In contrast, controlling smart devices (UD), changing smart devices (CD), and microphone control (MC) settings were comparably robust across both devices, with both heuristics generally satisfied. The results also highlighted areas of uncertainty or non-applicability, notably in smart home assistant location data (AL), user location (UL), and third-party skill (TP), where neither device consistently offered clear privacy settings or controls. The existence of settings without fine control in Google Home for UL and TP suggests a partial approach to privacy that may not fully meet user expectations for granularity.

Figure 1: Frequency of codes for security concerns about SPA according.



C Codebook

Table 5: Codebook and respective frequency of each code in answers to "When you think about using smart home assistants in your home, what security concerns come to mind, if any?"

Sentiment	Code Group	Code	Explanation	# SPA Owners	# Non-SPA Owners
Negative	Monitoring	Monitoring General	Audio and non-audio monitoring	6	8
		Audio Monitoring General	Unspecified agent monitoring audio	10	11
		Audio Monitoring SPA	SPA monitoring audio	8	8
		Audio Monitoring by Household Member	Household member monitoring audio	0	1
		Stalking	SPA aiding stalking	1	1
		Location Sharing	SPA knowing user location	1	0
	Unauthorized info leak	Data Breaches	Data theft	3	4
		Hacking	Mention hacking	7	8
		Parental Controls	SPA exposing children to unauthorized information	0	1
		Law Enforcement - Negative	SPA sharing info with law enforcement, legal proceedings	1	0
		Info Leak	General information leak	12	4
		Selling Data	Data being sold without authorization	2	0
		Unauthorized Access to IoT	Unauthorized access to IoT devices, including SPA	3	4
		Privacy Violations in General	Indicated nonspecific privacy concern	8	5
		Enabling Crime	SPA or its information leveraged for criminal activity	2	3
		Accidents	Accidentally triggering an unwanted action	1	1
		Advertisements	Ads and marketing	3	0
		Manufacturers	Doubting trustworthiness of the SPA manufacturer	2	1
		Smartphone	Their smartphone is already listening	2	1
Neutral		Unavoidable	Other tech makes privacy breaches unavoidable	0	1
		No Concern	Didn't indicate or describe concern	4	2
		Nothing to Hide	The participant mentions they have nothing to hide	1	0
		Unnecessary	Thinks SPAs are unnecessary	0	1
Positive		Law Enforcement - Positive	Thinks SPAs could help catch criminals	1	0