Privacy Norms for Smart Home Personal Assistants

Noura Abdi*
noura.abdi@kcl.ac.uk
King's College London
London, UK

Xiao Zhan*
xiao.zhan@kcl.ac.uk
King's College London
London, UK

Kopo M Ramokapane
marvin.ramokapane@bristol.ac.uk
University of Bristol
Bristol, UK

Jose M Such
jose.such@kcl.ac.uk
King's College London
London, UK

ABSTRACT

Smart Home Personal Assistants (SPA) have a complex ecosystem that enables them to carry out various tasks on behalf of the user with just voice commands. SPA capabilities are continually growing, with over a hundred thousand third-party skills in Amazon Alexa, covering several categories, from tasks within the home (e.g. managing smart devices) to tasks beyond the boundaries of the home (e.g. purchasing online, booking a ride). In the SPA ecosystem, information flows through several entities including SPA providers, third-party skills providers, providers of Smart Devices, other users and external parties. Prior studies have not explored privacy norms in the SPA ecosystem, i.e., the acceptability of these information flows. In this paper, we study privacy norms in SPAs based on Contextual Integrity through a large-scale study with 1,738 participants. We also study the influence that the Contextual Integrity parameters and personal factors have on the privacy norms. Further, we identify the similarities in terms of the Contextual Integrity parameters of the privacy norms studied to to distill more general privacy norms, which could be useful, for instance, to establish suitable privacy defaults in SPA. We finally provide recommendations for SPA and third-party skill providers based on the privacy norms studied.

CCS CONCEPTS

• **Security and privacy** → **Privacy protections**; **Usability in security and privacy**; • **Human-centered computing** → *Empirical studies in HCI*.

KEYWORDS

Smart Home Personal Assistants, AI Assistants, Voice Assistants, Privacy, Norms, Contextual Integrity, Amazon Alexa, Google Assistant

1 INTRODUCTION

Smart Home Personal Assistants (SPA) can do many tasks *on behalf of* their users, including to purchase goods and food, manage to-do lists, reply to knowledge questions, play music, plan holidays, control other smart home devices, send messages, make calls and many more [8, 25, 62]. SPAs are becoming widespread, with 147 million units sold in 2019 [65], of which 26.2% were Amazon Echo/Alexa units and 20% Google Home/Assistant, which are the two that dominate the market by a large margin over the rest. Despite this, SPAs have been shown to entail privacy risks, with systematic studies exploring privacy risks in specific SPA parts, such as the 'always listening' risk exacerbated by proven misactivations [24], and across the complex SPA ecosystem [20, 25, 37, 63]. Previous works showed that despite having incomplete mental models of SPA and their ecosystem, some users have privacy concerns and use coping strategies in the absence of better ways to manage their privacy in SPA [1, 36, 41, 45, 67].

Recently, prior works have focused on addressing or mitigating privacy issues of SPAs from the smart speaker point of view. For instance, gaze [49, 51] and voice volume [51] were proposed for the smart speaker to detect whether it is being spoken to, and hence it should be in listening mode. Other more radical approaches include jamming the audio signal with a wearable device [17], so that the user can control when the smart speaker should receive usable audio, or blocking the reception of sensitive conversations by using filters [16].

While these works pave the way towards having more control over the level of data that reaches the smart speaker, they usually focus on a small part of SPA ecosystem. SPAs are much more than just smart speakers: they have a complex architecture involving many different stakeholders [20, 25, 62]. This complex architecture has the smart speaker part, e.g. Amazon Echo, but this is connected to a cloud-based voice assistant, e.g. Amazon Alexa, where speech and intent recognition takes place. This is where the best course of action to serve the user's voice command is decided. In fact, the smart speaker only has minimal capabilities—only being able to recognize the wake-up word [74], and when recognized, it acts as a relay, forwarding all the recorded data to the cloud for analysis.

Once the cloud-based SPA backend recognizes the intent behind the user command, it matches the command with the *skill* that can better serve it [39] and then forwards the command to the skill. SPA skills, known as *Actions* in Google Assistant and *Skills* in Amazon Alexa, can be developed by third parties and hosted anywhere on the Internet. The number of third-party skills in Amazon Alexa has already surpassed 100,000 [71]. This implies many information flows across the complex SPA ecosystem, which involves several parties: the SPA provider (e.g., Amazon/Google), multiple third-party providers of skills or actions that SPA can request following users' voice commands (e.g., playing music through Spotify, ordering a ride through Uber, have emails read through Myemail, etc.).

Though recent work on SPAs started considering users' privacy perceptions and mental models beyond smart speakers [1, 36, 45],

*Both authors contributed equally to this research.

no previous work has actually looked at the acceptability of information flows across the whole SPA ecosystem, including users, SPA providers, providers of third-party skills, and other third parties. Another stream of research studied privacy norms in smart homes in general [10–12], but they did not consider the particularities and the complex ecosystem of SPA, as introduced above (e.g., they did not consider third-party skills, typical of SPA). Our work is the first to elicit and study the privacy norms across the whole SPA ecosystem. This is important, because eliciting and studying the privacy norms across the SPA ecosystem could help towards: i) SPA providers and regulators identifying and contextualizing usage patterns and when/where privacy violations may be happening; ii) checking if existing SPA privacy controls effectively align with privacy norms; iii) allowing providers and designers of SPA and skills to develop with users' expectations and desires in mind with regards to privacy and iv) devising suitable defaults for privacy controls in SPA based on the privacy norms.

We particularly study the following research questions:

RQ1 What are the privacy norms that should govern information flows across the SPA ecosystem?

RQ2 What contextual and personal factors influence these privacy norms the most?

RQ3 What are the subset of most general, representative privacy norms across the SPA ecosystem?

To answer these questions, we conducted a large-scale empirical study with 1,738 participants. We grounded the study based on the theory of Contextual Integrity (CI) [14, 53, 54], which provides a well-established framework for studying privacy norms and expectations, eliciting privacy norms across the SPA ecosystem. We also use data mining techniques (association rule mining) to elicit highly representative, general privacy norms. We make the following contributions:

- We elicit the privacy norms for data flows across the SPA ecosystem based on CI, considering different parameters that affect acceptability: sender, subject, recipient, data types, and transmission principles. In terms of recipients, we consider the whole SPA ecosystem, other users, SPA provider, third-party skill providers, and external parties.
- We study the influence that the CI parameters and personal factors have on the privacy norms. We show that privacy norms are mostly influenced by CI parameters, with only a few personal factors playing a (less important) role.
- We identify the similarities between the elicited privacy norms based on the CI parameters and data flow acceptability to distill more general privacy norms. Importantly, these more general privacy norms could help to establish privacy defaults for SPA.
- We provide recommendations for SPA and third-party skill providers based on the privacy norms studied.

2 BACKGROUND AND RELATED WORK

This section provides an insight into the privacy of SPAs. We provide the needed background on Contextual Integrity theory and how it can elicit privacy norms. We then discuss related work on privacy in the Smart Home in general and provide a more detailed account of the related work that has focused on SPA's specific and distinct characteristics to understand, study, and mitigate privacy issues in SPA.

2.1 Contextual Integrity and its Application to Elicit Privacy Norms

The theory of Contextual Integrity (CI) provides a well-established framework for studying privacy norms and expectations [14, 53, 54]. It defines privacy as the appropriateness of information flows based on social and cultural norms in a specific given context. CI describes the information flows using five parameters: (1) the sender of the information, (2) the attribute or information type, (3) the subject of the information that is being transferred, (4) the recipient of the information, and (5) the transmission principle or condition imposed on the transfer of the information from the sender to the recipient.

Contextual Integrity has been known as an appropriate framework to elicit expected privacy norms [10, 11, 47, 64]. This is usually done through factorial vignette surveys varying the contextual parameters in Contextual Integrity, asking for the acceptability of the information flows described in a vignette using the contextual parameters. That is, how acceptable it is that one specific sender sends information of a distinct type and specific subject to a specific recipient using (not using) a set of transmission principles (e.g., for a given purpose).

This methodology has been used to elicit expected privacy norms in general across online environments [47] or in specific domains such as data sharing in an education context [64]. It has also been applied, as discussed further below, to show that privacy norms could be possible to elicit meaningful privacy norms for some Smart Home devices [10], and to elicit privacy norms for particular types of Smart IoT devices like IoT Toys [11].

We based on this well-known vignette-based survey method in our study to elicit privacy norms across the SPA ecosystem, considering the instantiation of the Contextual Integrity framework to the SPA ecosystem. We also added a novel way to analyze the elicited information w.r.t. previous works that used this CI. Finally, this paper is the first to propose using data mining techniques (association rule mining) to find similarities between privacy norms, which allows us to elicit highly representative, general privacy norms that could be used as suitable defaults.

2.2 Privacy in the Smart Home

Extensive research has been conducted to date on the privacy of smart homes. From an HCI perspective, prior work studied users' mental models for smart home devices. For instance, [76, 79] conducted semi-structured interviews on smart home users to understand their mental models and privacy concerns. They found smart homeowners prioritize convenience over privacy, while [26, 55] explore security perceptions of smart homes, i.e., identifying factors that influence security decisions such as perceived competence and trust and what concerns users share prior and after purchasing an IoT device. In addition, Tabassum et al. [68] explored the preferences of users such as giving others not living with them access to smart devices, showing that users did want to give access to users not living with them (e.g. family) for certain purposes.

Other scholars studied access-control methods for IoT-enabled smart home devices [21, 35, 46]. For example, He et al. [35] examined

how access control policies differ regarding different contextual factors such as relationships between the users and different device capabilities in an IoT smart home. Colnago et al. [21] studied IoT Personalised Privacy assistants (PPAs) to help users discover and control data collection by nearby smart devices, and Zeng et al. [77] built a prototype and evaluated the usability of an access control app. Manandhar et al. [46] proposed Hєlion, which can analyze users routines and assist them design policies to secure their smart home based on them.

In the most similar research to our paper, Apthorpe et al. [10, 11] implemented the Contextual Integrity framework to examine the acceptability of information sharing in IoT devices, both sampling examples from smart home devices in general [10] and looking specifically at privacy norms for IoT toys [11]. Similarly, Barbosa et al. [12] follow a contextual approach to capture privacy preferences in smart homes, such as considering different attributes (e.g., recipient of the information) and privacy attitudes, and they propose machine learning models for predicting personalized privacy preferences for users. While we use a similar contextual-integrity-based methodology, with some key novelties such as data mining as mentioned in the previous section (Section 2.1), our focus is on eliciting privacy norms for the SPA, which is a distinct smart home technology with the open nature of its voice channel and a specific and complex ecosystem bringing its own privacy challenges [1, 20, 25, 37, 63]. For instance, these previous works do not consider the challenge of the ever-growing third-party skills in SPA, which already surpassed 100,000 in Amazon Alexa [71], and are crucial for SPA functionality. However, these skills are developed by third parties (not the SPA developers: Amazon, Google, etc.) and hence complicate information flows involving thousands of different entities. The user can only interact with the skills indirectly via voice, as they run remotely (in the third-party provider servers), not installed in the local smart speaker, and accessed via the SPA provider cloud.

2.3 Privacy in SPA

Several works studied the privacy perceptions and concerns users have regarding SPA [1, 18, 19, 29, 36, 41, 45, 50, 67]. Malkin et al. [45] explored users' perception regarding what happened to their voice recordings, and showed that the majority of their participants were not aware of the storage mode and management options of voice recordings. Cho [18] focused on the use of SPA to retrieve health related information, showing users did have concerns about using SPA for this and that the actual modality (voice/text) to interact with the SPA did not affect these perceptions. Other works [41, 67] highlighted SPA users' reasons for adopting SPA despite having privacy concerns, and found that consumers trade the benefits and convenience for privacy. Some works also showed the inaccurate mental models users have of SPA, and the coping strategies they use to mitigate their privacy concerns considering different parties in the SPA ecosystem [1, 36]. For instance, they showed that users may avoid certain features of SPA to protect themselves. Also they showed that most users did not really know how to protect themselves leading to ineffective privacy risk management. Finally, Cho et al. [19] studied the effect of enabling privacy settings in Alexa on trust. They found that if privacy settings customization goes accompanied with the option to also customize the content users may access through the assistant (e.g. information source) then trust increases. Although these studies have shed light on perceptions, concerns, mental models and coping strategies, they have not systematically investigate the acceptability of information flows and the privacy norms that should govern them across the SPA ecosystem.

Previous work has also attempted to address or mitigate privacy issues in SPA. One research stream in particular has focused on mechanisms to provide users with control over when the smart speaker should listen to them, as even when in theory smart speakers should only react to the wake-up word, misactivations are known to happen [24]. Chen et al. [17] developed a wearable microphone jammer capable of disabling microphones from all directions. Champion et al. [16] introduced an intermediary device that intelligently filters sensitive conversations from being recorded by the smart speaker. Other works proposed methods for the smart speaker to detect when it is spoken to [49, 51]. For this, they used gaze [49, 51] and/or voice volume [51] to detect when the user was addressing towards the smart speaker. As explained above, however, these works tend to focus on the smart speaker part of the SPA ecosystem and do not consider the further information flows that will happen once a smart speaker records what the user is saying. Alternative architectures for SPA have been proposed (such as Snips [22]) that make all computation to happen offline, but they need all the functionality to be predefined before deployment, something impractical for mainstream SPA.

3 METHODOLOGY

We conducted an online survey [70] based on Contextual Integrity to understand SPA users' security and privacy preferences regarding information flows in the SPA ecosystem. The study was reviewed and approved by our IRB. All the participants were recruited and compensated through Prolific [2] as detailed below.

3.1 Contextual Integrity in the SPA Ecosystem

Contextual integrity considers five main parameters: sender, recipient, information type (topic, attributes), information subject, and transmission principles. Previous research on (see Section 2.3) already looked at ways for the user to have more control over when the SPA is listening, but not about what the SPA does with the data after receiving it.This is crucial, as the SPA ecosystem is complex, and the functionality SPA offer entails several providers and entities. If the data remained and was processed within the smart speaker, this would not be a problem. However, the problem occurs because data flows through the complex SPA ecosystem. Therefore, our focus is on the information flows that originate from the SPA. We consider the SPA as the *sender*, as it automatically collects everything the user says after the wake up word [24] or after any misactivations [24], and send it to other parts of the SPA ecosystem, e.g., to the SPA provider for speech/intent recognition. As a result, we consider the information *subject* to be the user that is speaking to the SPA[1]. This means that in our study, in order to elicit

[1]Note here that it may be that the information a user speaks about to the SPA could be of other people. To simplify and reduce the resulting experiment design's dimensionality, we consider the information subject as a single individual. Exciting follow-up work could use the same framework and our results as a basis to see what would happen in the case where information belongs to more than one subject, as discussed as part of limitation later on.

privacy norms across the SPA ecosystem, we mainly vary three of the Contextual Integrity parameters: recipients, data attribute, and transmission principles, as detailed below. The instantiation of all Contextual Integrity parameters to the case of the SPA ecosystem is summarized in Table 1.

Data Attributes (Information Type). We aimed to maximize coverage of the type of information flows in the SPA ecosystem considering the categories that both Amazon and Google have of Skills [5] and Actions [31] for Alexa and Google Assistant, respectively, distinct end-uses involving different elements of the SPA ecosystem (e.g., queries, services, smart devices, shopping) [1], and the sensitivity of the data in question. We selected 15 data attributes across 11 categories of Skill/Action categories covering the different elements of the ecosystem, with the authors' perception as an initial guide for maximizing a variety of sensitivity levels. We then conducted a pre-survey with a convenience sample with 23 participants from different backgrounds to rate the sensitivity of the data attributes selected (using a 5-point Likert scale, 1 being the least sensitive and 5 being highly sensitive). Note that we did not use the sensitivity obtained through the pre-survey; we only used these values to have further assurances of covering data attributes of varying sensitivity beyond the authors' perceptions. Table 2 summarizes the categories, data type, and specific data collected, together with the mean and std. dev. of the sensitivity rating from the short survey. Finally note that, while aiming to cover different types of data, use-cases and sensitivities, all of the categories included a substantial number of skills. For instance, the Business & Finance category contains thousands of skills in Amazon Alexa at the time of writing.

Recipients. We considered the different recipients of information in the SPA ecosystem [20, 25, 36], including users and non-users: SPA providers, skill providers, and external entities. Regarding *users*, we considered the users, other subjects that may get the information from the SPA (e.g., by asking it). Here, we covered different types of relationships, some of them more associated with living with the subject of the information (partners, children, housemates) or just visiting (neighbors, close friends), but also others where it may not be that clear whether they live or not with the subject. This is because it has been shown that, even for some relationships (siblings, parents, housekeepers, etc.) who may not leave with the subject, there may be an appetite to share access to smart devices for specific purposes [68]. Regarding *non-user* recipients, we considered the SPA provider (e.g., Amazon, Google, etc.), who provide the cloud-based SPA backend; Skill providers, who provide extra functionalities to the SPA, a selection of which can be seen in the column "Category" of Table 2 (e.g. listening to music, shopping, managing other smart devices etc.); and external entities, who may be given access to the SPA ecosystem, such as Advertising Agencies [28] and Law Enforcement Agencies [72].

Transmission Principles. In Contextual Integrity, transmission principles condition the flow of information from party to party [14, 54]. The first condition we considered was the purpose for which the data shared would be used, which can apply to user and non-user recipients. Also, we considered conditions related to the processing and storage of information, which are more associated with the non-user recipients, such as the SPA provider, as they usually have data processing capabilities. In particular, we considered six conditions related to notice and confidentiality [14, 54], anonymity and retention period [10, 11], and review and deletion [66]: *1) If you are notified; 2) If the data is anonymous; 3) If the data is kept confidential, i.e., not shared with others; 4) If the data is stored as long as necessary for the purpose; 5) If you can review or delete the data.*

3.2 Survey Instrument

The study instrument was a questionnaire divided into three parts. The first part of the questionnaire explained the study, requested consent and collected information about participants SPAs. The second part of the questionnaire contained scenarios representing information flows in the SPA ecosystem and questioned about their acceptability (detailed below). The third part focused on participants privacy and security attitudes, this was measured through 10-item IUIPC scale [44] which covers three dimensions: Collection, awareness and control. To measure security attitudes, SA-6 was used, a six-item scale [27] for assessing users' self-reported security attitudes.

3.2.1 Scenarios. Vignette-based surveys based on Contextual Integrity [10, 11, 47, 64] have shown to be a successful method to elicit privacy norms in other domains as discussed in Section 2.1. Following a similar method, we created scenarios in the form of vignettes, where we explored all combinations of data attributes, recipients, and transmission principles in the SPA ecosystem, as introduced in Section 3.1. This led to a total of 120 different scenarios. In particular, and for each of the 15 different data types, we created eight scenarios. The eight scenarios per data type include two scenarios with user recipients, one with purpose and one without (conditions on data only concern data processors); and six scenarios with non-user recipients (data processors). The six scenarios with non-user recipients correspond to five with a purpose and varying conditions per recipient: SPA provider, relevant skill, non-relevant skill, advertising and law enforcement; and one scenario with all non-user recipients without transmission principles (no purpose/conditions). Note that relevant and non-relevant skills differed in that the relevant skill would be one from the category of the 15 datatypes specified while non-relevant would be a skill type that has no relevancy to having access to such data. This was important to study to explore whether acceptability is also dependent on the skill type as previous research has suggested that functionality may be an important factor to account for privacy in SPA [41, 67]. Finally, we measured each scenario's acceptability using a 5-point Likert scale, 1 to 5, completely unacceptable and completely acceptable, respectively. We presented the 5-point Likert scale from negative to positive for better feedback and reduce completion time [32, 43, 61]. The full details for all of the 120 scenarios used can be found in the Supplementary Materials. For illustration purposes, we show the wording used for two of those scenarios.

First, we show one example scenario with user recipients and with purpose:

"Assume that you play music through a voice assistant e.g. Amazon Echo/Alexa, Google Home/Assistant. How acceptable is it for your frequently played music to be shared with the following recipients, for the purpose of playing your favourite music:

- *Your partner* [5-point likert]
- ...

Table 1: Contextual Integrity framework instantiated in the SPA ecosystem.

CI Parameter	Value	# Description
Sender	The SPA	*The SPA being used, e.g. Amazon Echo/Alexa, Google Home/Assistant.*
Subject	The SPA User	*The User speaking to the SPA.*
Attribute	The 15 attributes in Table 2	*Represents a variety of information types flowing across the SPA ecosystem.*
Recipients	- Users - SPA provider - Skill provider - External Parties	*Partners, children, housemates, neighbors, house keeper, visitors, etc..* *The company that provides the SPA service, e.g., Amazon, Google, etc..* *The provider of third-party Skills. See Table 2 for Skill categories.* *Advertising Agencies, Law Enforcement Agencies.*
Transmission Principles	- No purpose, No condition - Purpose, No Condition - Purpose, Condition	Purpose: *the purpose for which data is collected was stated.* Condition: *1) If you are notified; 2) If the data is anonymous; 3) If the data is kept confidential, not shared with others; 4) If the data is stored as long as necessary for the purpose; 5) If you can review or delete the data.*

Table 2: Sensitivity of Data Attributes

Skill Category	DataType	Data Collected	Sensitivity	
			Mean	SD
Smart Home	**Smart Door Locker**	*Door lock state*	*4.0*	*1.2*
	Smart Thermostat	*History log*	*2.8*	*1.1*
	Smart Camera	*Home surveillance*	*4.4*	*1.0*
Business & Finance	**Banking**	*Bank account details*	*3.8*	*1.2*
Social & Communication	**Email**	*Email content*	*4.0*	*1.0*
	Call Assistant	*Contacts*	*3.5*	*1.0*
	Video calls	*Video calls data*	*3.7*	*1.0*
Health & Fitness	**Healthcare**	*Diagnosis results*	*4.1*	*1.0*
	Sleep Aid	*Sleeping hours*	*2.6*	*1.0*
Music & Audio	**Playlists**	*Frequently played music*	*2.3*	*1.3*
Shopping	**Online Shopping**	*Shopping history*	*2.8*	*1.2*
Productivity	**To do lists**	*Reminders*	*3.0*	*1.4*
Weather	**Weather forecast**	*Weather updates*	*1.8*	*1.1*
Travel & Transportation	**Ride services i.e. Uber**	*User location*	*3.0*	*1.0*
Non-skill SPA data	**Voice recordings**	*Command history*	*3.7*	*1.5*

- *Visitors in general* [5-point likert]"

Second, we show one example scenario with non-user recipient (a relevant skill), with purpose, and with different conditions as transmission principles:

"Assume that you order a ride such as Uber through a voice assistant e.g. Amazon Echo/Alexa, Google Home/Assistant. How acceptable is it for the data used for this, such as your home address, to be shared with providers of Skills or Actions in the Travel & Transportation category (e.g. Uber Skill) so they can know where to send the driver to pick you up, under the following conditions:

- *No condition* [5-point likert]
- *If you are notified* [5-point likert]
- ....
- *If you can review or delete the data* [5-point likert]"

3.3 Procedure

Our survey was created and hosted on Qualtrics[2], and we recruited our participants through Prolific[3]. We conducted two pilot studies to refine our survey instrument. Twenty (20) participants completed the first pilot study. We used the pilot study to explore whether the phrasing of the scenarios and the layout made sense to participants. Based on the feedback, we rephrased some of the scenarios, added the no condition, and changed the skills scenarios by making the skill category explicit. The second pre-test focused on the phrasing of the scenarios again, time taken, and the number of scenarios participants should answer. Fifty (50) people took the second pre-test survey. As a result, we reduced the number of scenarios to be administered to each participant and included the phrase "skills/actions" as an alternative to just using "skills" to help participants using Google devices understand scenarios better. All the data collected from both pre-test studies were only used to improve the survey but excluded from the final analysis.

The final version of our study was administered to over 2,017 Prolific workers. Each participant answered 24 scenarios, which were selected randomly as follows. Each participant got randomly assigned to six data attributes and got four random scenarios per attribute. We controlled for SPA users/non-users via pre-screening in Prolific to balance the sample to include a similar number of SPA users and non-users. We studied SPA users & non-users to observe differences in their privacy norms to see whether the privacy norms of SPA users could also be helpful for non-users of SPA, as some of those non-users may avoid SPA because of privacy concerns [41]. We also excluded all the participants who took part in

[2]www.qualtrics.com
[3]www.prolific.co

Table 3: Demographics of the survey participants.

		SPA Users	SPA Non-users
Gender	Male	474	416
	Female	397	438
	Prefer not to say	8	5
Age	18 – 24	340	339
	25 – 34	294	280
	35 – 44	134	93
	45 – 54	66	60
	55 – 64	37	60
	65+	8	27
Education	PhD	25	24
	Masters	163	173
	Undergraduate	315	250
	College/ A-levels	315	339
	Secondary Education	49	54
	No Formal Education	5	8
	Prefer not to say	7	11
Employment	Full time	428	274
	Part-time	153	165
	Unemployed	140	200
	Other	74	127
	Retired/Homemaker	72	78
	Prefer not to say	12	15
Student	Yes	333	348
	No	541	511
	Prefer not to say	5	-
Technology use	Not at all	166	254
	Daily	533	421
	Weekly	180	184
SPA brand	Amazon Echo/Alexa	465	
	Amazon Echo/Alexa + others	132	
	Google Home	261	
	Google Home + others	11	
	Microsoft Invoke/Cortana	5	
	Apple HomePod/Siri	3	
	Other	2	
TOTAL		**879**	**859**

both pre-test studies from taking part in the final study to prevent biases. Participants were compensated $2.00, and the survey took an average of 14 minutes to complete. Not all participants did all questions; participants were randomly assigned 4 scenarios across 6 datatypes to balance the sample size. Finally the list of options per scenario (e.g. the order in which user recipients were shown) was also randomized.

3.3.1 Data Quality. To maximize data quality, we employed three methods: attention checkers, randomization, and workers' previous performance (completed tasks and approval) rate) [34, 38, 48, 57, 60]. At the beginning of the study, we informed participants about answering the survey questions to the best of their ability, and then included two attention check questions spread across the 24 scenarios each participant had to answer (one after every eight scenarios). As stated above, we also randomized questions to ensure that each participant answered the survey randomly and that each scenario gets the same number of participants. In addition to randomized questions, we also randomized the way scenarios and options were presented to participants to prevent ordering effects. We also selected participants with at least 50 submissions and an approval rate of 95% or more during recruitment, as suggested in previous literature [60].

3.3.2 Participants. In total, 2,017 participants completed our study, and 268 failed one or both attention check questions. We were unable to retrieve demographics information from Prolific for twelve participants. All these participants and those who failed one or both attention check questions were removed from the analysis. In the end, we analyzed 1,738 responses, 879 SPA users, and 859 non-SPA users. Table 3 summarizes the demographics and the technical background of the participants used for analysis. We had a minimum number of 378, a maximum of 559, and a mean of 411 responses per scenario. As each of the 120 scenarios had acceptabilities for different values of one CI parameter (e.g. for each user recipient scenario with a data type and a purpose/no purpose participants reported acceptabilities for ten different relationships), the total number of instances with a complete specification of contextual integrity parameters was 292,478.

4 RESULTS

4.1 Overview of Privacy Norms

We first show an overview of the privacy norms elicited (RQ1) by reporting the acceptability of the different scenarios presented to the users. For this, we use heatmaps of acceptability depending on the Contextual Integrity parameters we varied, which is one way to visualize the acceptability of information flows based on Contextual Integrity parameters, as used in previous works eliciting privacy norms in other applications to visualize them [10, 11].

In Figure 1, we present the average acceptability scores for all the data types, user recipients, and transmission principles stating/not stating the purpose. Regarding data attributes, we observed that the actual data type seems to play a role, with some data attributes, which may seem a priori very sensitive (Banking, Email), leading to very low acceptability for most of the scenarios that involve them (i.e. regardless of the recipient and whether there is purpose or not). However, not in all cases, is acceptability dependent on the data type exclusively. For example, for data types that could also be considered rather sensitive like door locker and smart camera (and were rated as such in our initial pre-survey study of data attributes to include in the study—see Table 2), acceptability is clearly dependent on the type of recipient, e.g., when the recipient is partner or parents, it is much more acceptable than if the recipient is neighbors or visitors in general. In addition, some general trends can also be observed regarding transmission principles, with acceptability increasing when the purpose of sharing a data type with a recipient is stated.

Regarding scenarios with non-users recipients, of which the mean acceptability is shown in Figure 2, weather has the highest acceptability score, while banking has the lowest score, similar to with user recipients (as detailed in the previous paragraph). Regarding recipients, advertising agencies had the lowest acceptability scores, particularly when conditions were not declared. This was also prevalent regarding data types and transmission principles where the condition is not mentioned; acceptability scores are at the lowest. Information flow towards relevant skill providers had the highest acceptability score of all recipients, higher than the SPA provider, especially when the purpose is declared, and the transmission principle is "If you can review or delete the data". However, there is less variation between SPA providers and Law enforcement agencies as recipients.

4.2 Regression Analysis

While the acceptability figures in the previous section are useful in giving a birds-eye view of the mean acceptability across scenarios, we also sought to understand a bit more the effect that the different

Figure 1: Average Acceptability for Information Flows with User Recipients

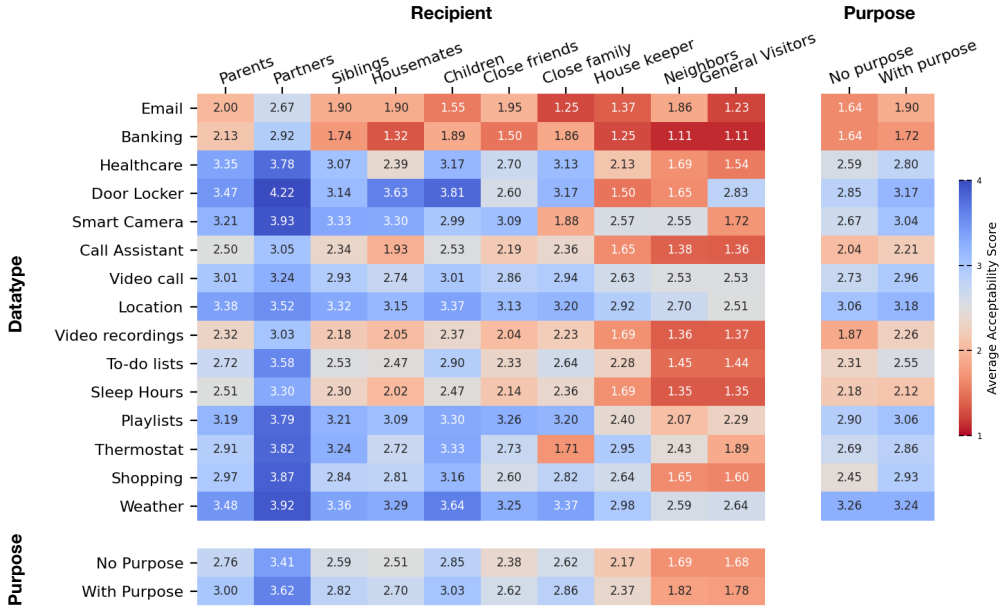
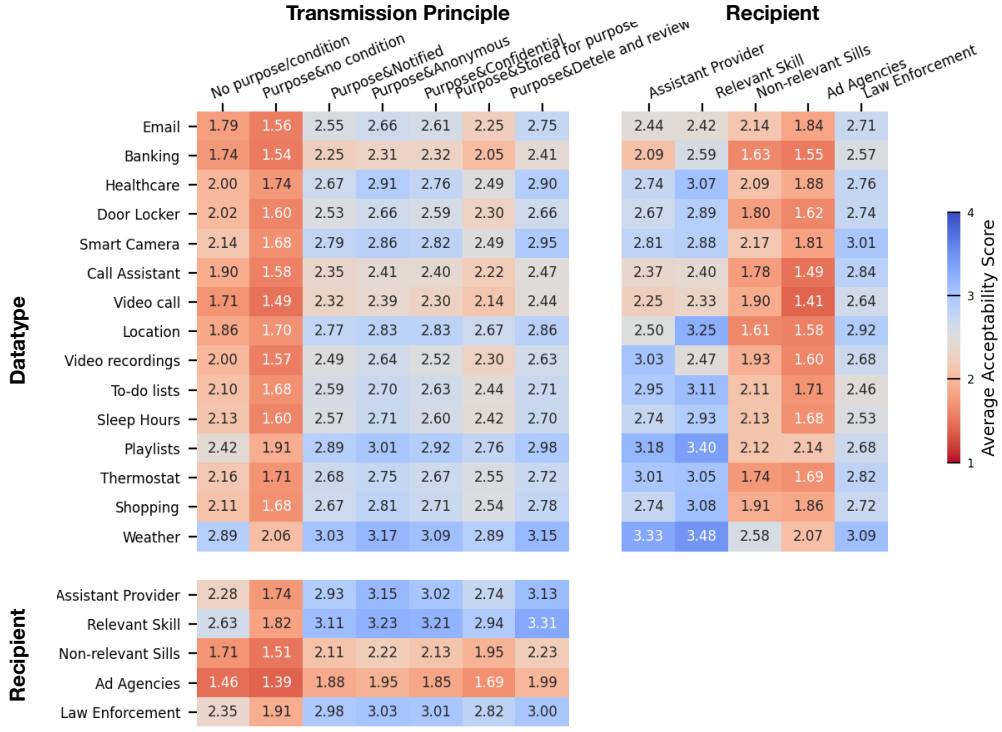


Figure 2: Average Acceptability for Information Flows with Non-User Recipients.



Contextual Integrity parameters, as well as personal characteristics of the participants, influence the acceptability of information flows (RQ2). For this purpose, we conducted regression analysis. In particular, rather than looking at a degree of acceptability, we were very interested in understanding the actual practical implications, i.e., whether an information flow would be considered a privacy violation or would be acceptable depending on the factors involved. Therefore, we binarized the acceptability of scenarios, separating between unacceptable and acceptable, and removing the

Table 4: User recipients Regression Model: Parameter Estimates. For categorical variables, the reference categories (not in the model) are Datatype=[Weather]; Recipient=[Visitors in general]; Transmission Principle=[Without Purpose]; SPA use=[Yes]; Employment=[Not in paid work]; Gender=[Female]; Technology use=[Daily].

		B	S.E.	Wald	df	Sig.			B	S.E.	Wald	df	Sig.
Datatype	[Email]	-2.703	.049	3001.795	1	.000	**Recipient**	[Parents]	2.214	.040	3003.790	1	.000
	[Banking]	-2.762	.051	2938.022	1	.000		[Partners]	3.250	.042	6051.599	1	.000
	[Healthcare]	-.962	.042	518.533	1	.000		[Siblings]	1.924	.041	2252.274	1	.000
	[Door Locker]	-.440	.043	106.415	1	.000		[Housemates]	1.783	.040	1939.825	1	.000
	[Smart Camera]	-.698	.043	267.300	1	.000		[Children]	2.330	.041	3296.858	1	.000
	[Call Assistant]	-1.946	.046	1816.590	1	.000		[Neighbours]	.056	.046	1.485	1	.223
	[Video Call]	-.813	.042	371.837	1	.000		[Close friends]	1.541	.041	1430.666	1	.000
	[Location]	-.396	.043	86.744	1	.000		[Close family]	2.007	.040	2465.254	1	.000
	[Voice Recordings]	-2.086	.046	2031.086	1	.000		[Housekeeper]	1.177	.041	807.676	1	.000
	[To-do-lists]	-1.408	.041	1197.194	1	.000	**Employment**	[Full-time]	-.158	.034	21.765	1	.000
	[Sleep Hours]	-1.923	.045	1815.239	1	.000		[Part-time]	-.048	.036	1.814	1	.178
	[Playlists]	-.482	.043	123.009	1	.000		[Rather not say]	.007	.038	.030	1	.863
	[Thermostat]	-.829	.044	363.330	1	.000		[Unemployed]	-.024	.035	.463	1	.496
	[Shopping]	-.963	.042	516.248	1	.000	**Continuous**	IUIPC Collection	-.154	.008	345.026	1	.000
Trans. Principle	[Purpose]	.388	.016	590.950	1	.000	**Variables**	IUIPC Control	-.088	.010	82.983	1	.000
SPA use	[No]	.204	.016	153.672	1	.000		SA_6	.101	.011	82.713	1	.000
Gender	[Male]	-.163	.017	95.545	1	.000		Age	-.015	.001	394.845	1	.000
Technology	[Not at all]	.050	.022	5.481	1	.019		Education Lev.	-.048	.006	56.664	1	.000
Use	[Weekly]	.105	.022	22.578	1	.000	**Constant**		.101	.095	1.123	1	.289

Table 5: User Model: Step Summary

	Improvement			Model			Correct	Variable
Step	**Chi square**	**df**	**Sig.**	**Chi-square**	**df**	**Sig.**	**Class %**	
1	12269.199	9	.000	12269.199	9	.000	66.3%	IN: Recipient
2	9961.801	14	.000	22231.000	23	.000	71.6%	IN: Datatype
3	921.541	1	.000	23152.541	24	.000	72.2%	IN: Age
4	642.157	1	.000	23794.698	25	.000	72.3%	IN: IUIPC_collection
5	582.788	1	.000	24377.486	26	.000	72.6%	IN: Transmission_principle
6	140.271	1	.000	24517.757	27	.000	72.7%	IN: SPA_use
7	148.094	1	.000	24665.851	28	.000	72.8%	IN: Education_level
8	104.442	1	.000	24770.294	29	.000	72.9%	IN: Gender
9	66.989	1	.000	24837.282	30	.000	72.9%	IN: IUIPC_control
10	72.820	1	.000	24910.102	31	.000	72.9%	IN: SA_6
11	79.067	4	.000	24989.169	35	.000	72.9%	IN: Employment
12	23.417	2	.000	25012.586	37	.000	72.9%	IN: Technology_use

neutral cases. We then conducted *Binary Logistic Regression*, using the forward stepwise method for model selection [40], in which the variables are progressively added to the model until there is no improvement. Beyond the Contextual Integrity parameters, we also considered the questions about demographics, SPA use, and privacy concerns/security attitudes in the model, so we could observe whether personal factors also play a role in the acceptability of information flows. We created two models based on the type of recipient, whether the recipients were other users (e.g. partners, neighbors, etc.) and whether recipients were non-users (SPA providers, Skill providers, etc.).

4.2.1 User recipients Model. The results for the regression model for scenarios with user recipients are summarized in Table 4 and Table 5. Regarding the quality of the model, we can see in Table 5 that model correctly classifies 72.9% of the cases (with an almost equal split between classes), which is a significant improvement over a null model (without any explanatory variables) as confirmed with an Omnibus Tests of Model Coefficients being highly significant ($\chi^2 = 25762.992, p = 0.000$). Although the purpose of the model was not to predict but to measure the effect of each variable, we also checked for the potential of overfitting. First, we had much more than the ten events per variable suggested as a minimum by previous literature [59] to avoid overfitting. Second, we also performed a random 70/30 split for building the model and predicting, giving a similar 72% accuracy.

Regarding the influence of each variable, we can see in Table 5 the order in which the model included them at each step, with student status and IUIPC_awareness being the only ones not shown, as the model rejected them because they were not providing any improvement after step 12. Based on the χ^2 quantification of improvement, we can see that the Recipient seems to be the most influential variable, followed closely by the data attribute. After this, and offering a similar improvement, we find two personal variables (age and IUIPC_collection) and the transmission principle (in this case, whether a purpose was given or not). The rest of the variables, while still providing an improvement, they do not contribute as much as the previous ones. Therefore, we can conclude that the contextual variables play the most important role. That is, the particular context is what is most likely to determine whether a flow is considered acceptable or unacceptable. From the two personal variables that play a more significant role, we see that the collection dimension of IUIPC is much more important than control and awareness (the latter not even being included in the model), which makes sense, as we focus on explicit data flows, i.e., who will be able to *collect* data as a result, and their acceptability.

Regarding the effects that the values of the variables have, Table 4 shows, among others, the value of the β coefficient for all the variables (including their categories for categorical variables) as well as their level of significance. Note: the interpretation of the coefficients and the significance of categories within variables needs to be done concerning the reference category (reference categories are listed in the caption of Table 4). For instance, we can see that all data types coefficients are significant and negative, which means that any of the attributes lead to unacceptability more often than the reference category, which in this case, is Weather. This makes sense, as Weather was reported as the least sensitive in the pre-survey we conducted to select attributes (see Section 2.1). Also, it can be seen that data types like Banking have a larger (negative)

Table 6: Non-user recipients Regression Model. For categorical variables, the reference categories (not in the table) are Datatype=[Weather]; Recipient=[Law Enforcement Agencies]; Transmission principle=[Purpose, review and delete the data]; SPA use=[Yes]; Employment=[Not in paid work];Student status=[Yes];Gender=[Female];Technology use=[Daily].

		B	S.E	Wald	df	Sig.			B	S.E	Wald	df	Sig.
Datatype	[Email]	-1.040	.033	988.614	1	.000	**Transmission Principles**	[No Purpose/Condition]	-1.193	.022	2881.772	1	.000
	[Banking]	-1.392	.034	1681.685	1	.000		[Purpose, no condition]	-2.337	.028	7201.837	1	.000
	[Healthcare]	-.689	.033	447.000	1	.000		[Purpose, notified]	-.232	.021	123.821	1	.000
	[Door Locker]	-.956	.033	844.029	1	.000		[Purpose, anonymous]	-.044	.021	4.383	1	.036
	[Camera]	-.666	.033	412.822	1	.000		[Purpose, confidential]	-.153	.021	54.894	1	.000
	[Call Assistant]	-1.264	.034	1421.084	1	.000		[Purpose, stored]	-.544	.021	664.274	1	.000
	[Video Call]	-1.393	.034	1685.514	1	.000	**Gender**	[Male]	.113	.012	81.776	1	.000
	[Location]	-.798	.033	580.696	1	.000	**Student**	[No]	.058	.015	14.269	1	.000
	[Voice Recordings]	-1.058	.033	1015.699	1	.000	**Technology Use**	[Not at all]	-.062	.016	14.641	1	.000
	[To-do-lists]	-.806	.030	735.882	1	.000		[Weekly]	.026	.017	2.445	1	.118
	[Sleep Hours]	-.819	.033	620.210	1	.000	**Employment**	[Full-time]	-.136	.025	28.834	1	.000
	[Playlists]	-.326	.032	100.432	1	.000		[Part-time]	-.022	.026	.695	1	.405
	[Thermostat]	-.683	.033	432.743	1	.000		[Rather not say]	-.065	.028	5.356	1	.021
	[Shopping]	-.732	.033	501.188	1	.000		[Unemployed]	.033	.026	1.626	1	.202
Recipient	[Assistant Provider]	.004	.017	.044	1	.835	**Continuous Variables**	IUIPC Collection	-.286	.006	2088.526	1	.000
	[Relevant Skill]	.304	.017	302.671	1	.000		IUIPC Control	-.067	.007	86.212	1	.000
	[Non-relevant Skill]	-1.308	.019	4700.128	1	.000		SA_6	.172	.008	432.944	1	.000
	[Advertising Agencies]	-1.765	.020	7546.064	1	.000		Age	-.008	.001	161.405	1	.000
SPA use	[No]	.106	.012	73.750	1	.000		Education_level	-.097	.005	404.073	1	.000
Constant		2.970	.077	1504.052	1	.000							

Table 7: Non-User Model: Step Summary

	Improvement			Model			Correct	Variable
Step	**Chi square**	**df**	**Sig.**	**Chi-square**	**df**	**Sig.**	**Class %**	
1	17256.388	4	.000	17256.388	4	.000	65.9%	IN: Recipient
2	13531.986	6	.000	30788.374	10	.000	71.2%	IN: Transmission_principle
3	3588.701	14	.000	34377.075	24	.000	72.1%	IN: Datatype
4	2951.785	1	.000	37328.860	25	.000	72.8%	IN: IUIPC_collection
5	743.186	1	.000	38072.045	26	.000	73.1%	IN: Education_level
6	294.017	1	.000	38366.062	27	.000	73.1%	IN: Age
7	367.405	1	.000	38733.466	28	.000	73.1%	IN: SA_6
8	102.998	4	.000	38836.465	32	.000	73.1%	IN: Employment
9	93.650	1	.000	38930.115	33	.000	73.2%	IN: IUIPC_control
10	77.254	1	.000	39007.369	34	.000	73.2%	IN: SPA_use
11	74.630	1	.000	39081.999	35	.000	73.2%	IN: Gender
12	23.714	2	.000	39105.712	37	.000	73.2%	IN: Technology_use
13	14.264	1	.000	39119.977	38	.000	73.2%	IN: Student_status

coefficient than others like Playlists, which means scenarios containing them are less acceptable when compared to scenarios with Weather than the scenarios with Playlists.

Regarding recipients, coefficients are all positive, e.g., showing more acceptability w.r.t. the reference category (Visitors in general). However, notice that neighbors do not significantly differ from them. A general trend suggests that the closer the type of relationship, the more acceptable the recipient is w.r.t. the reference category.

Finally, as already stated, non-contextual, personal variables play a much less important role than contextual variables. However, privacy concerns (both the collection and control dimensions of IUIPC) and age also play a role in acceptability, with higher privacy concerns and age leading to less acceptability. While other personal variables are significant, they do have considerably smaller coefficients.

4.2.2 Non-user recipients Model. The results for the regression model for scenarios with non-user recipients are summarized in Table 6 and Table 7. Regarding the quality of the model, we can see in Table 7 that the model correctly classifies 73.2% of the cases, which is a significant improvement over the null model (without any explanatory variables), confirmed with an Omnibus Tests of Model Coefficients being highly significant ($\chi^2 = 38495.635, p = 0.000$). As with the previous model, we also performed a random 70/30 split for building the model and predicting, giving a similar 73% accuracy.

Regarding the influence of each variable, we can see in Table 7 the order in which the model included them at each step. The only variable left out of the model in this case was IUIPC_awareness. Based on the χ^2 quantification of improvement, we can see that Recipient seems to be, as for the case of user recipients, the most influential variable. Differently from the user recipients model though, recipient is followed by the transmission principles. This may be because non-user recipients are data processors, conditions under which the information flows happen to take more relevance, and in particular, as also shown in Section 4.1 and confirmed by looking at the coefficients of transmission principles with no conditions in Table 6 which suggest much less acceptability than the reference category (being able to review and delete the data). After them, Data attribute and IUIPC_collection contribute to improving the model similarly as with the user recipients model. The rest of the variables also contribute, but the improvement is much less important, even marginal in some cases (SPA_use, Gender, etc.). Therefore, as for user recipients, the Contextual Integrity parameters and IUIPC play the most critical role. The differences here are mainly that transmission principles seem to take a more prominent role than the actual data attribute and that Age seems to influence less acceptability than for user recipients.

Regarding the influence of particular categories within variables, Table 6 shows their coefficients and significance levels, among others. Similarly to the user recipients model, all data types seem less acceptable than the reference category (Weather). Regarding recipients, we see no significant difference between SPA providers and the reference category, which is law enforcement agencies, but non-relevant skills and advertising agencies appear less acceptable than the reference category. In stark contrast, relevant skills appear significantly more acceptable than the reference category. This further supports the clearly higher average acceptability shown when considering relevant skills. As we will discuss later (Section 5), this points to a crucial distinction between relevant and non-relevant skills, with implications for SPA providers and

skill providers. Regarding transmission principles, all seem less acceptable than giving the purpose and allowing to review and deleted the data collected (reference category), except giving the purpose but sharing the data anonymously, which does not show significant differences concerning the reference category. Finally, as in the user recipients model, the collection dimension of IUIPC seems the most remarkable among the personal variables.

4.3 General Privacy Norms

The regression models in the previous section are beneficial to understand better the role that the Contextual Integrity parameters and personal characteristics play in the acceptability of information flows. However, their predictive power was not as accurate as to be used to determine whether a particular information flow should be allowed or not in practice. Therefore, we were also very interested in exploring the possibility of having a smaller, more general subset of privacy norms that would hold across scenarios (RQ3). Nissenbaum posits in her Contextual Integrity theory [53, 54] that one may not generalize norms based on privacy concerns or indexes, but that privacy views *may* differ because of the context, and provided empirical evidence towards this [47]. This, however, does not mean that there may not be contexts leading to similar privacy views. Our aim was precisely to find which contexts in the SPA ecosystem are similar (share some attributes) and have the same acceptability. To this aim, we implemented data mining techniques, particularly association rule mining, to elicit a subset of more abstract and general privacy norms in the SPA ecosystem.

Association rule mining is a rule-based method for discovering interesting relations between variables in datasets [69]. Its purpose is to discover frequent rules that exist in a dataset. In other words, association rule mining is used for knowledge discovery, so it is an unsupervised machine learning method. The type of rules we are interested in are:

Contexual Integrity Parameters → Acceptable/Unacceptable

We used the well-known Apriori algorithm [3] for both user and non-user recipients to mine the association rules. The algorithm typically uses two inputs to restrict the rules it will mine, minimum *Support*, and minimum *Confidence* [13]. As an unsupervised technique, it is not easy to set these parameters systematically or based on a rule of thumb [69], and their choice of them depends on the application. *Support* represents the frequency of an itemset in the dataset, which ensures the left-hand side of the rule to have at least a minimum support. However, note that we generated the scenarios in our case, so the frequency of the variables in the left-hand side of a rule is very similar, so we set this value low (0.03) to consider all potential cases of interest. *Confidence* is used to measure the frequency of a specific rule in the whole dataset. In our case, it would relate to the frequency in which a particular subset of parameters created scenarios that were acceptable/unacceptable regardless of the other parameters. Therefore, *Confidence* is more critical in our case, and we set it to be at least 0.66 (out of a maximum of 1, which would mean every time a particular subset is observed, the scenario is always perceived as either acceptable or unacceptable). That is, by setting it to 0.66, it is similar to the notion of a qualified majority or super-majority of at least two-thirds (66%). To measure the reliability of the confidence of the mined rules, we used the *Lift* of each rule [13].

After using the Apriori algorithm, we applied a further filtering step to the association rules mined. This was done because the Apriori algorithm does not consider the semantic relationship between the values of some of our parameters. For instance, the transmission principle where having a purpose is contrary to not having it. Therefore, two rules with the same data type but one with purpose and the other without in the left-hand side, and the same acceptability, may be generalized to just the data type and the acceptability.

4.3.1 Rules Mined. Figure 3 and Figure 4 show the set of association rules mined with user recipients and non-user recipients, respectively. We show the confidence for each rule (conf.) and also its lift to measure the reliability of the confidence of the mined rule, so that if the lift is > 1, the left-hand side (antecedent) and right-hand side (consequent) of a rule are dependent on one another, and makes those rules potentially useful for predicting the consequent in future data sets [13], which is the case in all the rules mined. Note that the set of rules mined do not cover all cases, but they are a set of general rules (just knowing a recipient or a data type is enough in most cases) that could very much be considered a baseline or suitable defaults for the scenarios they represent.

Regarding the association rules mined for user recipients, as shown in Figure 3, we can see that the rules mined consider recipients, data type, and transmission principles. Most of the rules (nine) are very general in the sense that they suggest acceptability/unacceptability with just the value of one of the CI parameters. The rest have either recipient or data type, together with purpose/no purpose. Information flows received by the most distant type of relationships, including neighbors and visitors in general show are not acceptable, while information flows received by other much closer relationships like partner are acceptable, and by close friends are only unacceptable if there is no purpose for the flows. For other flows, the recipient seems to matter less, and it is the type of data, with data related to email, banking, voice recordings and the like making the flows not acceptable.

Regarding the association rules mined for the specific case of non-user recipients, as shown in Figure 4, there are only six rules mined, which are less than for user recipients, but the rules mined are very general, clearly linking unacceptability to two (non-relevant skills and advertising agencies) out of the five types of recipients. This seems to suggest that the unacceptability of flows regarding these two types of recipients suggested in previous sections is actually rather generalized. The rules also seem to suggest that no purpose and no condition is not acceptable regardless of the actual non-user recipient and the data type. Finally, three data types are considered unacceptable regardless of recipients and transmission principles (call assistant, video call, and banking).

5 DISCUSSION

We now discuss the main implications of our results, and provide, where appropriate, recommendations for SPA and third-party skill providers based on the privacy norms studied.

Figure 3: Rules mined for user recipients.

R1	Recipient=Close Friends, Transmission Principle=Without Purpose → Unacceptable	(conf. 0.693)	(lift 1.123)
R2	Recipient=Partner → Acceptable	(conf. 0.706)	(lift 1.843)
R3	Datatype=Todo List, Transmission Principle=Without Purpose → Unacceptable	(conf. 0.638)	(lift 1.293)
R4	Recipient=House Keeper → Unacceptable	(conf. 0.721)	(lift 1.169)
R5	Datatype=Sleeping Hours → Unacceptable	(conf. 0.752)	(lift 1.218)
R6	Datatype=Call Assistant → Unacceptable	(conf. 0.753)	(lift 1.221)
R7	Datatype=Voice Recording → Unacceptable	(conf. 0.776)	(lift 1.258)
R8	Datatype=Email → Unacceptable	(conf. 0.857)	(lift 1.389)
R9	Datatype=Banking → Unacceptable	(conf. 0.870)	(lift 1.409)
R10	Recipient=Neighbors → Unacceptable	(conf. 0.875)	(lift 1.418)
R11	Recipient=Visitors in general → Unacceptable	(conf. 0.881)	(lift 1.428)

Figure 4: Rules mined for non-user recipients.

R1	Datatype=Call Assistant → Unacceptable	(conf. 0.719)	(lift 1.100)
R2	Datatype=Video Call → Unacceptable	(conf. 0.751)	(lift 1.149)
R3	Datatype=Banking → Unacceptable	(conf. 0.755)	(lift 1.155)
R4	Transmission Principle= Without Purpose/Condition → Unacceptable	(conf. 0.778)	(lift 1.190)
R5	Recipient=Non-relevant Skills → Unacceptable	(conf. 0.797)	(lift 1.219)
R6	Recipient=Advertising Agencies → Unacceptable	(conf. 0.859)	(lift 1.314)

5.1 The Recipient is Key (but not on its own)

From all the contextual and personal parameters explored in this study, the recipient of the data was the most decisive influencing factor to determine the acceptability of information flows in the SPA ecosystem, both with user and non-user recipients. This was clearly based on average acceptability, regression analysis, and the association rules mined.

When considering user recipients, our results suggest more trusted/close relationships leading to higher acceptability. Previous work already suggested that trust may play a role with those in the same household in the SPA context [36]. Our results seem to confirm this for ten relationship types. Furthermore, this is in line with other studies looking more in general at the smart home. For instance, [75] studied the perceptions of bystanders and owners in a smart home; they found that again perceived trust and social relationship are important factors in acceptability of information sharing across the smart home. An interesting future line of research would be to also consider how the norms may in turn influence the personal/social relationships [9].

Beyond who the recipient was, it was evident that acceptability also depends on other contextual aspects, such as the type of information being considered (data related to Banking and Email was anyway deemed unacceptable on average and in the rules mined) and whether there was a purpose for the recipients to have access to the data (e.g., the rules mined suggested close friends may be acceptable only if there is a specific purpose). In addition, while the results give evidence for general privacy norms that could, for instance, be used as potential default privacy settings, this does not mean that options to change them if need be should not be offered as well. For instance, there might be specific circumstances that make some, in principle, acceptable flows as unacceptable, e.g., it has been shown that smart devices in general may be used by close family or partners to spy on or abuse others [42].

When considering non-user recipients, the evidence was robust in terms of non-relevant skills (more specifically about skills in the next section) and advertising agencies. These recipients consistently led to very low acceptability across scenarios regardless of other Contextual Integrity parameters and personal characteristics of users. The evidence was less clear-cut regarding SPA providers and law enforcement agencies and differences between them (the differences between the two in the regression model were not significant). In that case, however, we could see some clear trends when the type of information and the transmission principles are considered. For instance, access to smart camera data through the SPA by law enforcement agencies was deemed acceptable on average with a purpose (investigating a crime), which may actually be influenced by cases previously in which SPAs were used as evidence by law enforcement, i.e., the police [56, 72] to investigate a crime. Also, most users seemed ok with the SPA provider having access to the voice recordings to improve its functionality and performance despite the well-known outcry reported in the media not long ago regarding this [15], crucially, in this case, acceptability seemed to increase if the purpose was coupled with conditions like being able to review and delete the data, even though users seem not to actually make use of this much in reality [45], or the data being anonymized.

5.2 Relevant vs Non-relevant Skills

For non-user recipients, our results (including average acceptability, regression analysis, and association rule mining) suggest that when the recipient of data is a skill, then information flows are acceptable or unacceptable depending on whether the skill providing functionality is associated with the data being requested. For example, for datatype "email" if the skill collecting the data is relevant, e.g., Myemail skill allows SPA users to have their emails read to them and send emails using their voice was deemed acceptable. However, if the same data type is to be collected by skill not relevant to it, e.g., the Spotify skill that allows SPA users to play music with their voice, this was deemed unacceptable. It is worth pointing out that, even for the relevant skills, some data types are still unacceptable (suggesting that perhaps users are not comfortable doing certain things like banking through SPAs). Also, transmission principles seem to play a vital role, with the cases in which either purpose or conditions were not given, then information flows also became

unacceptable. All of this has implications for both skill providers and SPA providers, given the large and ever-growing number of third-party skills in the SPA ecosystem [71].

For SPA providers, these results offer useful insights in terms of the relevance of personal data collected by skills both at the vetting stage and after a skill is allowed into the skill marketplace. At the vetting stage, the personal data asked by a skill could be checked against privacy norms and compared with the personal data asked by skills in the same category. Note, however, that this may be restricted to the skills that use personal data that can be requested using Amazon's API [7], e.g., home address, but skills have many other ways of collecting personal information. Skills can also perform account linking, which links the SPA user to their account with the skill provider (e.g., the smart bulb provider wanting control with Alexa). A more complicated case is when a skill may ask for personal data *during* a conversation with the user. This is not easy to assess, as skills run in a remote Internet location and may only be possible to interact with them as a black box. There are emerging tools to facilitate this kind of analysis, such as Skill-Explorer [33], but they are still very limited, as they seem unable to interact with all skills available in a marketplace, particularly those with similar names/invocation commands. Something which can even be more problematic, given that skill impersonation attacks are possible [78].

For skill providers, it highlights the importance of using privacy by design principles when developing the skills, e.g., data minimization [23], only collecting the minimum data to provide the functionality of the skill. It also shows the importance for skill providers to consider the acceptability of certain conditions under which they will treat the data provided. This should go in addition to merely creating a legally valid privacy policy as required by SPA providers like Amazon [6] for skills dealing with personal data, though some skills do not even attempt to have policies [4].

5.3 Default and Dynamic Privacy in SPA

We were able to mine some general rules, and we observed trends that could be used to inform suitable privacy defaults for information flows across the SPA ecosystem. For instance, there was strong support for some flows to be unacceptable based on the type of data (e.g. banking) or the recipient (e.g. non-relevant skills, as already detailed in the previous section). Therefore, SPA providers should not allow these information flows by default. Importantly, differences between current SPA users and those who are not using SPA yet were small (according to the regression model results in Section 4), suggesting that by including these defaults it may be possible to also alleviate the privacy concerns of those who are yet not using SPA.

Beyond privacy defaults, we also observed variability in privacy norms based on the specific, fully-specified (with all Contextual Integrity parameters) context and users' personal characteristics. Therefore, there is a great opportunity beyond suitable defaults for SPAs to support users to configure information flows depending on the context, similar to what has been done in other domains like smart phones and social media [58, 73]. However, previous research [1, 79] has shown that users are rather unlikely to configure themselves access control mechanisms in the smart home in general. Thus, it may not be realistic to expect users to configure any access control mechanisms that may be added for information flows in the SPA ecosystem to match their preferred privacy norms. Current SPAs are equipped with personalizing capabilities, which could be extended to adapt and personalize SPA's privacy settings using the general privacy norms as defaults but learning what specific privacy norms users would prefer in particular over time. In fact, SPA learning is sometimes perceived as beneficial [1]. Here, recent lessons learned in terms of assisting users to manage their privacy in the smart home in general, such as finding the right automation-intervention tradeoff [21], as well as the adequate use of AI for privacy preference learning in the smart home [12] and in other domains [52], seem crucial. While creating the mechanisms for this to be possible is challenging, we believe that efforts towards (partly) automatic privacy configuration tools for SPA will be more fruitful than putting more effort into creating more traditional, manual access control mechanisms to control information flows across the SPA ecosystem.

5.4 Limitations and Future Work

Although our study explored acceptability of information flows in the SPA ecosystem with 120 scenarios across 15 types of data, 15 types of recipients, and 7 types of transmission principles, not every single case could be studied. Still, some of the results we got seem to hold across scenarios, with even a set of rather general privacy norms obtained via association rule mining. Also, information may be owned by multiple users living in a shared home (family, housemates), as studied by [30, 36, 76]. As future work, we plan to explore how Contextual Integrity may apply to those multi-user cases, as well as the type of privacy norms that may emerge in those cases, and whether/how they may change with respect to the privacy norms observed in this study. Finally, when considering non-user recipients, we focused on third-party entities (such as skill providers) or SPA providers. We did not consider other people who may not be users (casual people getting in range of the smart speaker), even though we did consider as potential *users* people who may not be necessarily registered as such, e.g. visitors and neighbors.

6 CONCLUSION

We studied privacy norms in the SPA ecosystem, considering the acceptability of the information flows among the several entities involved in it, including SPA providers, third-party skills providers, users and other parties. To do this, we applied the Contextual Integrity theory, and elicited privacy norms through a large-scale study based on the Contextual Integrity parameters instantiated to the SPA ecosystem.

Beyond the concrete implications we discussed in the previous section that the norms we elicited have for the parties involved in the SPA ecosystem, such as SPA providers and skill providers, there are other broader implications to consider. For instance, we believe that policy makers and regulation bodies should pay more attention to and investigate the privacy norms in the SPA ecosystem we elicited in this study. This would help them formulate corresponding restrictions and rules to monitor the behavior of organizations in the SPA ecosystem, and discover any privacy violations in this domain to help protect the consumer.

REFERENCES

[1] Noura Abdi, Kopo M Ramokapane, and Jose M Such. 2019. More than smart speakers: security and privacy perceptions of smart home personal assistants. In *Fifteenth Symposium on Usable Privacy and Security ({SOUPS} 2019)*.

[2] Prolific Academic. [n.d.]. Prolific Participants Recruiting Platform. https://prolific.ac/

[3] Rakesh Agrawal, Ramakrishnan Srikant, et al. 1994. Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB*, Vol. 1215. 487–499.

[4] Abbdulaziz Alhadlaq, Jun Tang, Marwan Almaymoni, and Aleksandra Korolova. 2017. Privacy in the Amazon Alexa Skills Ecosystem. *HotPETS* (2017).

[5] Amazon. [n.d.]. Amazon Alexa Skill Store (US). https://www.amazon.com/alexa-skills/b?node=13727921011

[6] Amazon. 2019. Security Testing for an Alexa Skill. https://developer.amazon.com/docs/custom-skills/security-testing-for-an-alexa-skill.html. [Online; last accessed 03-July-2019].

[7] Amazon.com. 2019. 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. [Online; last accessed 16-December-2019].

[8] Tawfiq Ammari, Jofish Kaye, Janice Y Tsai, and Frank Bentley. 2019. Music, Search, and IoT: How People (Really) Use Voice Assistants. *ACM Transactions on Computer-Human Interaction (TOCHI)* 26, 3 (2019), 1–28.

[9] Noah Apthorpe, Arunesh Mathur, Pardis Emami-Naeini, Marshini Chetty, and Nick Feamster. 2019. You, Me, and IoT: How Internet-Connected Home Devices Affect Interpersonal Relationships. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*. 142–145.

[10] Noah Apthorpe, Yan Shvartzshnaider, Arunesh Mathur, Dillon Reisman, and Nick Feamster. 2018. Discovering smart home internet of things privacy norms using contextual integrity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (2018), 59.

[11] Noah Apthorpe, Sarah Varghese, and Nick Feamster. 2019. Evaluating the Contextual Integrity of Privacy Regulation: Parents' IoT Toy Privacy Norms Versus {COPPA}. In *28th {USENIX} Security Symposium ({USENIX} Security 19)*. 123–140.

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

[13] Roberto J Bayardo Jr and Rakesh Agrawal. 1999. Mining the most interesting rules. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*. 145–154.

[14] Sebastian Benthall, Seda Gürses, Helen Nissenbaum, et al. 2017. *Contextual integrity through the lens of computer science*. Now Publishers.

[15] Bloomberg. 2019. Amazon Workers Are Listening to What You Tell Alexa. https://www.bloomberg.com/news/articles/2019-04-10/is-anyone-listening-to-you-on-alexa-a-global-team-reviews-audio

[16] Christopher Champion, Ilesanmi Olade, Constantinos Papangelis, Haining Liang, and Charles Fleming. 2019. The Smart^2 Speaker Blocker: An Open-Source Privacy Filter for Connected Home Speakers. *arXiv preprint arXiv:1901.04879* (2019).

[17] Yuxin Chen, Huiying Li, Shan-Yuan Teng, Steven Nagels, Zhijing Li, Pedro Lopes, Ben Y Zhao, and Haitao Zheng. 2020. Wearable Microphone Jamming. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

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

[19] Eugene Cho, S Shyam Sundar, Saeed Abdullah, and Nasim Motalebi. 2020. Will Deleting History Make Alexa More Trustworthy? Effects of Privacy and Content Customization on User Experience of Smart Speakers. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[20] Hyunji Chung, Michaela Iorga, Jeffrey Voas, and Sangjin Lee. 2017. Alexa, can I trust you? *Computer* 50, 9 (2017), 100–104.

[21] Jessica Colnago, Yuanyuan Feng, Tharangini Palanivel, Sarah Pearman, Megan Ung, Alessandro Acquisti, Lorrie Faith Cranor, and Norman Sadeh. 2020. Informing the design of a personalized privacy assistant for the internet of things. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[22] Alice Coucke, Alaa Saade, Adrien Ball, Theodore Bluche, Alexandre Caulier, David Leroy, Clement Doumouro, Thibault Gisselbrecht, Francesco Caltagirone, Thibaut Lavril, Mael Primet, and Joseph Dureau. 2018. Snips Voice Platform: an embedded Spoken Language Understanding system for private-by-design voice interfaces. *CoRR* abs/1805.10190 (2018).

[23] G Danezis, J Domingo-Ferrer, M Hansen, JH Hoepman, D Le Metayer, R Tirtea, and S Schiffner. 2014. Privacy and Data Protection by Design-from policy to engineering. *TP-05-14-111-EN-N* (2014).

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

[25] Jide Edu, Jose M Such, and Guillermo Suarez-Tangil. 2020. Smart Home Personal Assistants: A Security and Privacy Review. *Comput. Surveys* (2020).

[26] Pardis Emami-Naeini, Henry Dixon, Yuvraj Agarwal, and Lorrie Faith Cranor. 2019. Exploring How Privacy and Security Factor into IoT Device Purchase Behavior. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 534.

[27] Cori Faklaris, Laura A Dabbish, and Jason I Hong. 2019. A self-report measure of end-user security attitudes (SA-6). In *Fifteenth Symposium on Usable Privacy and Security ({SOUPS} 2019)*.

[28] Forbes. 2018. How Alexa Is Changing The Future Of Advertising. https://www.forbes.com/sites/ilkerkoksal/2018/12/11/how-alexa-is-changing-the-future-of-advertising

[29] Nathaniel Fruchter and Ilaria Liccardi. 2018. Consumer Attitudes Towards Privacy and Security in Home Assistants. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, LBW050.

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

[31] Google. [n.d.]. Google Assistant Action Store. https://assistant.google.com/explore

[32] Rebecca Hofstein Grady, Rachel Leigh Greenspan, and Mingnan Liu. 2019. What Is the Best Size for Matrix-Style Questions in Online Surveys? *Social Science Computer Review* 37, 3 (2019), 435–445.

[33] Zhixiu Guo, Zijin Lin, Pan Li, and Kai Chen. 2020. SkillExplorer: Understanding the Behavior of Skills in Large Scale. In *29th USENIX Security Symposium (USENIX Security 20)*. 2649–2666.

[34] David J Hauser and Norbert Schwarz. 2016. Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior research methods* 48, 1 (2016), 400–407.

[35] Weijia He, Maximilian Golla, Roshni Padhi, Jordan Ofek, Markus Dürmuth, Earlence Fernandes, and Blase Ur. 2018. Rethinking access control and authentication for the home internet of things (iot). In *27th {USENIX} Security Symposium ({USENIX} Security 18)*. 255–272.

[36] Yue Huang, Borke Obada-Obieh, and Konstantin Beznosov. 2020. 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*. 1–13.

[37] Catherine Jackson and Angela Orebaugh. 2018. A study of security and privacy issues associated with the Amazon Echo. *International Journal of Internet of Things and Cyber-Assurance* 1, 1 (2018), 91–100.

[38] Yujin Kim, Jennifer Dykema, John Stevenson, Penny Black, and D Paul Moberg. 2019. Straightlining: overview of measurement, comparison of indicators, and effects in mail–web mixed-mode surveys. *Social Science Computer Review* 37, 2 (2019), 214–233.

[39] Young-Bum Kim, Dongchan Kim, Joo-Kyung Kim, and Ruhi Sarikaya. 2018. A Scalable Neural Shortlisting-Reranking Approach for Large-Scale Domain Classification in Natural Language Understanding. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 3 (Industry Papers)*. 16–24.

[40] Jason E King. 2008. Binary logistic regression. *Best practices in quantitative methods* (2008), 358–384.

[41] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. 2018. Alexa, Are You Listening?: Privacy Perceptions, Concerns and Privacy-seeking Behaviors with Smart Speakers. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 102.

[42] Roxanne Leitão. 2019. Anticipating Smart Home Security and Privacy Threats with Survivors of Intimate Partner Abuse. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 527–539.

[43] Mingnan Liu and Alexandru Cernat. 2018. Item-by-item versus matrix questions: A web survey experiment. *Social Science Computer Review* 36, 6 (2018), 690–706.

[44] Naresh K Malhotra, Sung S Kim, and James Agarwal. 2004. Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information systems research* 15, 4 (2004), 336–355.

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

[46] Sunil Manandhar, Kevin Moran, Kaushal Kafle, Ruhao Tang, Denys Poshyvanyk, and Adwait Nadkarni. 2020. Towards a natural perspective of smart homes for practical security and safety analyses. In *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE, 482–499.

[47] Kirsten Martin and Helen Nissenbaum. 2016. Measuring privacy: an empirical test using context to expose confounding variables. *Colum. Sci. & Tech. L. Rev.* 18 (2016), 176.

[48] Winter Mason and Siddharth Suri. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods* 44, 1 (2012), 1–23.

[49] Donald McMillan, Barry Brown, Ikkaku Kawaguchi, Razan Jaber, Jordi Solsona Belenguer, and Hideaki Kuzuoka. 2019. Designing with Gaze: Tama–a Gaze Activated Smart-Speaker. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–26.

[50] Marissa Merrill. 2020. An Uneasy Love Triangle Between Alexa, Your Personal Life, and Data Security: Exploring Privacy in the Digital New Age. *Mercer Law Review* 71, 2 (2020), 7.

[51] Abraham Mhaidli, Manikandan Kandadai Venkatesh, Yixin Zou, and Florian Schaub. 2020. Listen Only When Spoken To: Interpersonal Communication Cues as Smart Speaker Privacy Controls. *Proceedings on Privacy Enhancing Technologies* 2020, 2 (2020), 251–270.

[52] Gaurav Misra and Jose M Such. 2017. Pacman: Personal agent for access control in social media. *IEEE Internet Computing* 21, 6 (2017), 18–26.

[53] Helen Nissenbaum. 2004. Privacy as contextual integrity. *Wash. L. Rev.* 79 (2004), 119.

[54] Helen Nissenbaum. 2009. *Privacy in context: Technology, policy, and the integrity of social life.* Stanford University Press.

[55] Norbert Nthala and Ivan Flechais. 2018. Informal support networks: an investigation into home data security practices. In *Fourteenth Symposium on Usable Privacy and Security ({SOUPS} 2018)*. 63–82.

[56] Douglas A Orr and Laura Sanchez. 2018. Alexa, did you get that? Determining the evidentiary value of data stored by the Amazon® Echo. *Digital Investigation* 24 (2018), 72–78.

[57] Leonard J Paas and Meike Morren. 2018. Please do not answer if you are reading this: Respondent attention in online panels. *Marketing Letters* 29, 1 (2018), 13–21.

[58] Sameer Patil and Jennifer Lai. 2005. Who gets to know what when: configuring privacy permissions in an awareness application. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 101–110.

[59] Peter Peduzzi, John Concato, Elizabeth Kemper, Theodore R Holford, and Alvan R Feinstein. 1996. A simulation study of the number of events per variable in logistic regression analysis. *Journal of clinical epidemiology* 49, 12 (1996), 1373–1379.

[60] Eyal Peer, Joachim Vosgerau, and Alessandro Acquisti. 2014. Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior research methods* 46, 4 (2014), 1023–1031.

[61] Joss Roßmann, Tobias Gummer, and Henning Silber. 2018. Mitigating satisficing in cognitively demanding grid questions: evidence from two web-based experiments. *Journal of Survey Statistics and Methodology* 6, 3 (2018), 376–400.

[62] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I Hong. 2018. " Hey Alexa, What's Up?" A Mixed-Methods Studies of In-Home Conversational Agent Usage. In *Proceedings of the 2018 Designing Interactive Systems Conference*. 857–868.

[63] William Seymour. 2018. How loyal is your Alexa? Imagining a Respectful Smart Assistant. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–6.

[64] Yan Shvartzshnaider, Schrasing Tong, Thomas Wies, Paula Kift, Helen Nissenbaum, Lakshminarayanan Subramanian, and Prateek Mittal. 2016. Learning privacy expectations by crowdsourcing contextual informational norms. In *Fourth AAAI Conference on Human Computation and Crowdsourcing*.

[65] Strategy Analytics. 2020. Global Smart Speaker Vendor & OS Shipment and Installed Base Market Share by Region: Q4 2019.

[66] Jose M Such. 2017. Privacy and autonomous systems. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. 4761–4767.

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

[68] Madiha Tabassum, Jess Kropczynski, Pamela Wisniewski, and Heather Richter Lipford. 2020. Smart Home Beyond the Home: A Case for Community-Based Access Control. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[69] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. 2016. *Introduction to data mining*. Pearson.

[70] Abbas Tashakkori, Charles Teddlie, and Charles B Teddlie. 1998. *Mixed methodology: Combining qualitative and quantitative approaches*. Vol. 46. Sage.

[71] TechRepublic. 2020. Alexa Skills: Cheat Sheet. https://www.techrepublic.com/article/alexa-skills-cheat-sheet/

[72] The Independent. 2018. Amazon ordered to give alexa evidence in double murder case. https://www.independent.co.uk/life-style/gadgets-and-tech/news/amazon-echo-alexa-evidence-murder-case-a8633551.html

[73] Pamela J Wisniewski, Bart P Knijnenburg, and Heather Richter Lipford. 2017. Making privacy personal: Profiling social network users to inform privacy education and nudging. *International Journal of Human-Computer Studies* 98 (2017), 95–108.

[74] Minhua Wu, Sankaran Panchapagesan, Ming Sun, Jiacheng Gu, Ryan Thomas, Shiv Naga Prasad Vitaladevuni, Bjorn Hoffmeister, and Arindam Mandal. 2018. Monophone-based background modeling for two-stage on-device wake word detection. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 5494–5498.

[75] Yaxing Yao, Justin Reed Basdeo, Oriana Rosata Mcdonough, and Yang Wang. 2019. Privacy Perceptions and Designs of Bystanders in Smart Homes. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–24.

[76] Eric Zeng, Shrirang Mare, and Franziska Roesner. 2017. End user security and privacy concerns with smart homes. In *Thirteenth Symposium on Usable Privacy and Security ({SOUPS} 2017)*. 65–80.

[77] Eric Zeng and Franziska Roesner. 2019. 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)*. 159–176.

[78] Nan Zhang, Xianghang Mi, Xuan Feng, XiaoFeng Wang, Yuan Tian, and Feng Qian. 2019. Dangerous skills: Understanding and mitigating security risks of voice-controlled third-party functions on virtual personal assistant systems. In *2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, 1381–1396.

[79] Serena Zheng, Marshini Chetty, and Nick Feamster. 2018. User Perceptions of Privacy in Smart Homes. *arXiv preprint arXiv:1802.08182* (2018).