Ignorance is Bliss? The Effect of Explanations on Perceptions of Voice Assistants

WILLIAM SEYMOUR, King's College London, UK
JOSE SUCH, King's College London, UK

Voice assistants offer a convenient and hands-free way of accessing computing in the home, but a key problem with speech as an interaction modality is how to scaffold accurate mental models of voice assistants, a task complicated by privacy and security concerns. We present the results of a survey of voice assistant users (n=1314) measuring trust, security, and privacy perceptions of voice assistants with varying levels of online functionality explained in different ways. We then asked participants to re-explain how these voice assistants worked, showing that while privacy explanations relieved privacy concerns, trust concerns were exacerbated by trust explanations. Participants' trust, privacy, and security perceptions also distinguished between first party online functionality from the voice assistant vendor and third party online functionality from other developers, and trust in vendors appeared to operate independently from device explanations. Our findings point to the use of analogies to guide users, targeting trust and privacy concerns, key improvements required from manufacturers, and implications for competition in the sector.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**; **Natural language interfaces**.

Additional Key Words and Phrases: voice assistants, AI assistants, explanations, mental models, trust

ACM Reference Format:
William Seymour and Jose Such. 2023. Ignorance is Bliss? The Effect of Explanations on Perceptions of Voice Assistants. In *CSCW '23, October 13–18, Minneapolis, MN, USA.* ACM, New York, NY, USA, 24 pages. https://doi.org/10.1145/xxxxxx

1 INTRODUCTION

Domestic voice assistants such as Alexa and Google Assistant continue to grow in popularity, offering easy access to information and automation in the home. But the use of speech as the primary mode of interaction for these AI-driven devices obscures their underlying configuration and mechanics in ways that might otherwise be afforded by the design of their interfaces (e.g. the presence of a setting in a menu affords its changing, but this visibility does not easily translate to a conversation). Human-computer speech is presently a lower bandwidth interaction modality, as devices lack the ability to vary volume, intonation, and cadence. This leads to incomplete mental models of how they work [1] and, combined with the always on and connected nature of voice assistants, has led to widespread privacy and security concerns about the extent to which they record and monetise what goes on around them [1, 33, 48, 80].

Recent work on people's perceptions of voice assistants reveals uncertainty about how voice assistants work, such as when devices are recording conversations, and what audio and transcripts are used for beyond the immediate actioning of requests [48, 57, 80]. Given that several major voice assistant manufacturers are global leaders in data-driven sectors such as advertising and e-commerce, the potential for conflicts of interest between profit and people's privacy are clear, and are reflective of wider concerns over the rise of surveillance capitalism [10, 93]. There is

CSCW '23, October 13–18, 2023, Minneapolis, MN, USA

an understanding that the main user-facing outcome of regulatory responses—privacy policies—inadequately address user concerns; often difficult to understand [39], contradictory [5], or simply left blank [25], privacy policies have become something that the majority of users blindly click through without reading.

The focus on 'privacy by design' in the EU General Data Protection Regulation (GDPR, Art. 25) is a welcome development, and is leading to a consistent set of building blocks to engineer GDPR principles into systems [34]. In particular, a number of design patterns and strategies are emerging towards this end [21], including those attempting to adequately inform users, provide actionable controls, and demonstrate compliance with stated policy goals [85]. However, some can be difficult to operationalise effectively when applied to complex intelligent systems like voice assistants [85] and this is further complicated by the highly inaccurate mental models that users currently have of them [1].

Related to uncertainties about how voice assistants work are concerns over the reliability of speech recognition. Studies of voice assistant logs suggest that a significant amount of voice assistant activations fail in some way. Sciuto et al. identify 26% of logged Alexa invocations as rapidly following another request, a characteristic feature of functional task failure, and 9.6% of remaining activations as unprocessable or misfires [79]. Dambanemuya and Diakopoulos similarly report a 30% rate of irrelevant or failed responses to news-related requests [20]. This makes it easy to believe media reports of voice assistants recording conversations and sending them to others[1], and leads to general misgivings about the maturity of voice recognition technology [38]. Examples of failed interactions in the literature suggest that an improved understanding of the underlying technology would help users recover from failed interactions [73]. Given this, finding ways to improve mental models of voice assistants and other smart home technology represents a major open challenge for the HCI community.

As such, the paper addresses the following research questions:

RQ1 How does the presence of first and third party entities in a voice assistant affect users' trust, privacy, and security perceptions of these devices?

RQ2 How does the provision of trust, privacy, and security information given in an explanation of a voice assistant affect users' trust, privacy, and security perceptions?

And in so doing makes the following contributions:

- Shows that voice assistants with first and third party entities are perceived as more concerning than offline assistants, with finer distinctions often made between the inclusion of first and third party entities.
- Shows that voice assistant explanations involving information about trust increase trust concerns, while those involving information about privacy reduce privacy concerns.
- Demonstrates that for some aspects of trust, users' existing choice of device is more important than functionality or explanation in determining concerns about new devices
- Finds that the 'incident anxiety' trust factor previously observed in other smart home devices also manifests in voice assistants.

More specifically, the paper presents the results of a study that sought to understand the effect of explanations on users' perceptions of voice assistants. We used a two level study design that described voice assistants with differing types of online connectivity using combinations of trust, privacy, and security information. We then tested to see how these factors impacted people's perceptions of the assistants and asked participants to explain how they worked back to us in their own words to see which words and concepts were retained. The results of the study reveal nuances

[1]https://www.theguardian.com/technology/2018/may/24/amazon-alexa-recorded-conversation

around the introduction of third parties, suggesting security reservations around the use of third party skills over first party or local ones, privacy reservations around the use of any non-local skills, and different responses to different kinds of information. Privacy explanations were found to ameliorate privacy concerns, whereas trust explanations increased concerns over trustworthiness. We also identify functionality and risks that were both well and poorly understood by participants, such wake words and skill squatting respectively. Reflecting on this, we lay out recommendations around the targeting of different concerns through explanations, development of analogies to build mental models, and the unique relationship of trust and voice assistants.

2 BACKGROUND

2.1 Explainable Smart Technology and AI

When considering explanations of intelligent systems such as voice assistants, an obvious related area of research is the burgeoning body of work on fairness, accountability, and transparency issues with ML-driven classifiers and decision-making systems. Explanations have been identified as a key tool in making these systems more understandable, promoting their safe, ethical, and effective use.

General work by Hoffman et al. on the criteria for effective explanations suggests that they should (1) give a satisfying, detailed, and complete understanding of how the system works; (2) be actionable; and (3) indicate how reliable and trustworthy the system is [35]. There is a growing body of work on producing explanations for the actions of intelligent systems, addressing a wide range of problems communicating things such as the link between data and conclusions and uncertainty in those conclusions [61], or the reasons and justifications for a recommendation [63, 64]. Often used in the context of machine learning systems and decision aids (such as credit scoring algorithms), they are frequently driven by ethics and seek to both address information asymmetries between data processors and subjects, and serve as a human-readable check when decisions need to be audited. Common methods for achieving this include the approximation of local boundaries [76] and generation of counterfactuals [89].

When we consider the applicability of this work to the research questions of the present study several differences from the general explainable AI literature become clear. Whereas the goal of the latter is mostly to contextualise the output of a model to an expert or provide insights to end users (e.g. about what would have been required for a different decisions to be reached), voice assistant explanations as we envisage them in this paper are about providing mental scaffolding to help people develop accurate mental models of how their devices work. Unlike with automated systems offering bank loans, a major problem in everyday connected life is the *lack* of visibility of algorithms (or 'AI') in contemporary platforms, apps, and services. Even when people are aware of the learning processes that shape their interactions, understandings of how these systems work and the associated benefits and drawbacks is highly variable (e.g. as seen with social feeds [22, 29, 75]). Work on visualising actions (including data sharing) by autonomous vehicles [45], mobile apps [83, 87] and smart home devices [81] has been developing strategies for further understanding how people intuitively interpret the information available to them, as well as developing techniques for explaining why and how devices behave as they do. These include the use of data to support personal strategies, information about *why* actions were taken, and the business models that drive observed behaviour. These efforts have been enabled by technical endeavours to reverse-engineer network flows from devices [4, 36], as for many devices exact information about where and how frequently information is shared is unavailable.

Many of these solutions deliver some form of privacy notice, and work on the design of these notices adopts similar recommendations to Hoffman et al., suggesting that they be relevant, actionable, and understandable. Specific recommendations that notices (1) understand the system's

data practices and users; (2) are short and specific; (3) highlight unexpected practices; and (4) provide details on demand [78], provide insights that are potentially applicable to voice assistant explanations. Given that unactionable privacy revelations often leave users feeling trapped and helpless [38, 83], this looks to be a key part of any implementation. Attempts to adapt the above to actual smart devices usually follow a design-led approach; the fictional Polly smart kettle displays data flows on the side of the kettle, and presents an explicit trade-off between functionality and data collection [52], and other approaches have focused on representing ambiguity around data collection and sharing [16] or the wider smart home design space [11]. Unfortunately this often runs counter to the prevailing minimalist design language that prioritises sleek and unobtrusive form factors over those that have more affordances and relay a richer set of information (summed up by Pierce and DiSalvo as "appiness" [72]), and this phenomena is only exacerbated by the hands-free nature of voice assistants that further rules out visual cues as a means of reliably communicating information about a device.

2.2 Voice Assistant Privacy and Security: Perceptions and Realities

Concerns over privacy and security are frequently found amongst users of voice assistants. It is not uncommon, for example, to find folk theories describing hacking as a bigger threat to privacy than data collection through surveillance capitalism [96] or conversations in the vicinity of VAs being used to target advertising [80] (an activity perceived much more negatively than data collection by companies for other purposes [77]). The knowledge that one is being recorded can lead to negative consequences, with prior work showing the damaging long term psychological effects of ubiquitous surveillance in the home [69].

But perceptions of privacy and security issues are often complicated by incomplete mental models of voice assistants, particularly when devices occupy spaces shared with cohabitant users [38] or non-user 'bystanders' [38, 58, 92]. This lack of understanding leads to a general lack of awareness about the risks involved [1, 48], leading to people falling back on non-technical coping strategies in an attempt to protect themselves [1, 81]. This is not helped by the fact that research has also overly focused the risks of human-voice assistant interaction at the expense of other parts of the voice assistant ecosystem [27], in turn skewing media coverage of privacy and security risks to users' everyday interactions.

Research on voice assistant 'skills' or 'actions', which extend the capabilities of voice assistants via conversational software developed by third parties (e.g. a Spotify skill to play music from a Spotify account), and the ecosystems they form highlights security and privacy problems with skill vetting, permissions, 'squatting', and data handling [25, 26, 46, 51]. More generally, people without technical backgrounds are less likely to understand the organisations and infrastructure that their devices are integrated with, including the risks to privacy and security that arise as a result [41]. Prior work also shows that knowledge of where data is being stored, processed, and shared by assistants is at best localised to users' particular brand of assistant [1]. This lack of knowledge leads to the avoidance of features that are perceived to be more risky, such as online shopping using a voice assistant [1, 48].

The small amount of systems security research that does exist in this area also demonstrates the dangers involved when connecting voice assistants to security-critical devices such as locks and garage doors; proof-of-concept attacks that utilise lasers [86] or high frequency sound [95] to manipulate voice assistant devices. These attacks are not typically seen in the wild, but further give the impression that voice assistants were not designed to provide the level of security required for the tasks they can be asked to do. Several versions of privacy and security labels have been proposed in order to communicate potential risks to consumers and allow for the comparison of factors such as availability of security updates between devices [28, 43]. There has also been legal

debate over the way that voice assistants often store audio recordings of users in the cloud where they can be accessed by law enforcement, bypassing existing protections on activities undertaken in the home (such as the Third Amendment in the US) [24].

2.3 Trust and Voice Assistants as Social Actors

Unlike traditional visual interfaces where there are established design patterns for communicating options and error states, the design language of the speech interfaces used in voice assistants is less mature [65], with ongoing research into fundamental aspects such as accessibility [84], progresivity [31], and conversational repair [13]. Alongside routine mis-interpretation and accidental activations, outbreaks of creepy laughter [49] and wake words being triggered by advertisements [9] have achieved high media visibility and furthered perceptions that users cannot trust these devices to work consistently and that they are potentially unsettling.

Gathering user perceptions about the reliability of voice assistants poses an interesting problem. On the one hand, conventional measures of reliability in AI/intelligent systems (e.g. [55], closely related to trust [12]) can be used to measure beliefs around task fulfilment, consistency over repeated interactions, and response accuracy. On the other hand, some studies have also highlighted the tendency for people to have satisfactory interactions with voice assistants even when they fail to complete the tasks given to them [53], possibly related to situations where social interactions *are* the main intended outcome of an interaction (as in [74]).

This is part of a larger tendency of people to anthropomorphise computers, particularly those that use speech. Early work by Nass et al. showed that people subconsciously treat computers as social actors, and do this whilst being fully aware that they are interacting computers rather than with people [67]. At the same time, voice assistants are often specifically *designed* to be anthropomorphised by users, leading researchers to investigate the extent to which they are (often implicitly) gendered and personified [3, 74]. Prior work has shown that the use of these mechanisms is associated with increased information disclosure, mindless enactment of social scripts, and more positive responses to information sharing [54, 62, 77].

Relatedly, work on partner models in speech interfaces shows the significant role they play in interaction. These models are generally composed of perceptions of a system's competence and dependability, it's human-likeness, and cognitive flexibility [23]. While people generally assume greater knowledge when interacting with voice interfaces than human interlocutors, they use assumptions about what other *people* know when estimating the knowledge of artificial partners [17]. It is worth noting, however, that greater human likeness in VAs also increases expectations of performance and can therefore reinforce the limitations of current solutions.[18]

These phenomena each add an additional layer of complexity to how people conceptualise voice assistants, blending rational and social ways of thinking; drawing studies of both adult and younger voice assistant users show a wide variety of depictions including humans, machines, and extraterrestrial satellites [50, 91]. These variations hint at deeper differences in people's mental models, such as the extent to which assistants are perceived to be shared between or unique to different users (e.g. whether there is one communal Alexa, or millions of individual instantiations) [50].

2.4 Designing the Voice Assistants of the Future

Looking to the future, where might voice assistant explanations fit in relation to other developments? Works envisaging voice assistants as virtual butlers [71], councillors [42], and music coaches [90], or as part of smart homes imbued with different 'personalities' [60] present intriguing variations to the social mechanisms typically seen with voice assistants used in the home. Developing techniques in verification and monitoring for ML systems [7, 19, 37] also make it increasingly possible that

future generations of voice assistants and smart devices could have provable privacy and security properties like those seen in cryptographic algorithms or mission-critical software (e.g. the Signal protocol [15] and the SEL4 microkernel [44]). We believe that explanations are well positioned to ease both of these transitions, for the same reasons that they are applicable to the assistants of today. Easing the transition towards assistants that are capable of being more proactive and engaging users in more of a discussion will require the ability develop trust in them and an understanding of how and why they might take certain actions.

2.5 Summary

In seeking to better understand our interactions with voice assistants and adjacent smart devices, researchers have uncovered a number of factors that contribute to a widespread lack of understanding about many aspects of these devices, particularly concerning privacy, security, and trust. These often involve misconceptions around the technical capabilities of devices, such as when they are recording, that are exacerbated by the 'appiness' [72] of sleek but deliberately low-bandwidth interfaces and the more socially-oriented cognitive processes that are engaged when interacting with devices via speech.

Current answers to these problems revolve around visual means—utilising visualisations and printed labels—which are less suitable for use with voice assistants. Prior work on explanations in adjacent contexts is extremely promising, but fails to address a number of phenomena that are specific to voice assistants, such as misfires and anthropomorphism. Therefore, we now explore responses to a number of different vignettes that describe functionality specific to voice assistants and could feasibly be delivered piecemeal via speech. By studying responses to these descriptions, and how well people are able to explain them in their own words, we hope to discover which approaches are the most effective and the effect that they have on people's perceptions of voice assistant technology.

3 METHODS

In order to explore the research questions above we constructed a survey of voice assistant users exploring perceptions of the voice assistant ecosystem and responses to explanations. The survey consisted of three main parts: the first presented users with a hypothetical voice assistant, an explanation of how it worked, and assessed their perceptions of its trustworthiness, privacy, and security; the second asked participants to re-explain the workings of this assistant in their own words; and the final part of the survey gathered baseline levels of privacy and security concern. Following Pavlou and Fygenson we adopted a definition of trust as "the belief that an entity will act cooperatively to fulfil clients' expectations without exploiting their vulnerabilities" [70]. This includes trust in devices, vendors, and other entities that are explained in more detail below. The survey was implemented using Qualtrics and we recruited via the Prolific Academic platform[2], compensating participants at an average rate of £8.70 an hour. Participants were all 18+ years of age, resident in the UK, and owned a voice assistant. All parts of the study were approved by our institution's IRB. Survey questions, anonymised results, and the qualitative codebook are provided as supplemental material and archived at https://osf.io/z7f5n.

3.1 Designing the Vignette Explanations

To investigate changes in perceptions towards assistants with different functionality explained in different ways, we utilised a two level (3x8) study design visualised in Table 1. We generated a set of 24 vignettes describing three fictional voice assistants with differing numbers of external entities,

[2]https://prolific.co

An assistant with the following connectivity:		
Offline functionality only (0E)	Online functionality with the first-party vendor/entity (1E)	Online functionality with third party entities (3E)
Explained using one of the following combinations of details:		
Functionality (F)	Functionality & Trust (T)	Functionality & Privacy (P)
Functionality & Security (S)	Functionality, Trust & Privacy (TP)	Functionality, Trust & Security (TS)
Functionality, Privacy & Security (PS)	Functionality, Trust, Privacy & Security (TPS)	-

Table 1. Overview of the 3x8 study design used to generate vignettes.

each explained in eight different ways. Entities were represented through the skills available to the assistant: offline/ on-device skills only (0E), first party online skills (1E), or third party online skills (3E). Explanations were generated from content snippets containing: functional descriptions of how a voice assistant would respond to a command (F); information about whether users could trust the assistant to function reliably and consistently, including ways that it might violate the interaction context [78] (T); information about data privacy (P); and information about the security of the device (S). All vignettes included a basic functional description of the assistant, with every combination of trust, privacy, and security represented for a total of 8 explanation combinations (hereafter F denotes an explanation containing *only* the functional description). We refer to these vignettes in the remainder of the paper using the shorthand given above in brackets, e.g. the vignette for third party entities with functional, trust, privacy, and security information would be 3E/TPS. An annotated example of this vignette is given Figure 1.

Conversational design for contemporary voice assistants commonly employs a model of *intents*, *utterances*, and *slots*. An intent represents an action that the user can perform (e.g. playing music), which can contain variables called slots (e.g. an artist or song title), and are invoked through utterances (e.g. "Play something by Taylor Swift"). We represented this approach in the vignettes by saying that the assistant software matched key words in a request to determine both the command to run and any variables stating how it should be executed. In order to minimise the potential for priming participants towards anthropomorphism we gave the voice assistant an invocation that was not a human name ("hey assistant"), and did not directly include words said by the assistant in the vignette. Prior work has suggested that the use of data-based framings of smart home devices can encourage creativity when imagining uses for data and devices [14], so we were specific about the kinds of information and processing that were taking place through the assistant.

3.2 Measuring Trust Perceptions

After presenting participants with a vignette we asked about their perceptions of its trustworthiness, including trust in its privacy and security, using eight questions from Cannizzaro et al. (originally Q8–Q15, reported here as Q1–Q8). The list of questions is given in Figure 2. While these all pertained to trust, they were divided into general trust in the device and its manufacturer (expressed as a combination of competence/performing tasks reliably, benevolence of vendors in respecting users' interests, and vendor's integrity [59]), trust in the privacy of the device, and trust in the security of the device. To aid readability we refer to these three dimensions as trust, privacy, and security in the remainder of the paper. These questions were answered using a seven point Likert scale from strongly agree to strongly disagree, with 'smart home devices' replaced with 'voice assistant'. While both the privacy and security questions make use of the word security, the former describe outcomes

Functionality	Morgan uses the assistant through a smart speaker—"Hey Assistant, ask Pact Coffee to order me another bag of coffee". The device keeps a recording of the last five seconds of what was said and checks to see if this contains the words "Hey Assistant".
Trust	Sometimes other things Morgan says sounds similar to this. This can lead to the assistant accidentally recording conversations and trying to interpret them as commands.
Security	The assistant can be configured to tell people apart based on their voices, but Morgan hasn't set this up yet. Sometimes Morgan's son tries to buy things with it.
Privacy	The recording of Morgan saying the request is sent over the internet to the assistant's provider and are stored on the provider's servers and are used to personalise skills and improve speech recognition algorithms.
Trust	The assistant will stop working if it can't connect to the Internet.
Functionality	An algorithm on the developer's server creates a transcript of Morgan's speech, and tries to match key words to a list of possible commands. Once it matches the words order and Pact Coffee, it forwards the request on to the store where Morgan normally buys coffee along with details about where Morgan lives so that the coffee can be delivered, and their card details. This gets arranged into sentences and sent back to the device. It is then turned into speech using Morgan's preferred voice and read out loud.
Privacy	A copy of the interaction is saved in the voice assistant's log.
Trust	Sometimes the assistant mishears and opens the wrong skill, or accidentally orders something
Privacy	causing Morgan's address and card details to be sent to the wrong company.
Functionality	The coffee company uses Morgan's address and card details to generate the order and sends confirmation of this back to the assistant provider.
Privacy	Each company has a different policy about what data it collects from Morgan and how long it is stored for.
Security	Some skills might deliberately be named to share an invocation with another skill (I.e. they are both called 'BBC'), or attempt to collect personal information without asking for the proper permissions. This can be used for fraud or to compromise Amazon or other online accounts.

Fig. 1. Treatment 3E/TPS: voice assistant with first and third party skills, explanation includes functionality, trust, privacy, and security elements. Annotated to show which elements of the explanation correspond with the four information types.

affecting only data and the latter ones where this leads to follow-on impacts such as fraud (users were not given the question classifications). The same distinction applies to the privacy and security components of the vignette explanations. While no validated questionnaire exists for these concepts in this context, the questions were determined to have content (logical) validity by the original authors [10]. In order to remove unnecessary variables we did not include information about the vendor of the fictional voice assistant in the vignettes and the trust component of the vignettes therefore mainly related to trust as competence, a limitation described further in Section 5.6.

(1) I would fully trust the voice assistant not to fail, and to function as I expect it to (general trust, reverse coded)
(2) Knowing that the voice assistant allowed companies or organisations to collect data about how I used it, and hence about my domestic habits, would restrict me from owning/using it (general trust)
(3) I would trust the manufacturer not to use data produced by the voice assistant for any purpose without my explicit consent (general trust, reverse coded)
(4) I think the likelihood of the security of voice assistants being compromised and resulting in a privacy/data breach is high (trust in privacy)
(5) I think the impact of the security of voice assistants being compromised and resulting in a privacy/data breach is low (trust in privacy, reverse coded)
(6) The Facebook user data sharing controversy makes me less willing to own/ use voice assistants (privacy)
(7) I think the likelihood of the security of voice assistants being compromised and resulting in an incident (e.g. burglary, fraud) is high (trust in security)
(8) I think the impact of the security of voice assistants being compromised and resulting in an incident (e.g. burglary, fraud) is low (trust in security, reverse coded)

Fig. 2. Survey questions adapted from Cannizzaro et al. [10].

3.3 User-Written Explanations

We also wanted to understand how easily participants were able to absorb and retain the concepts described in the different explanations. Participants were asked to re-explain how the assistant worked in their own words, as if they were describing it to a friend or family member (minimum of 150 characters/approx. 35 words). In order to balance the likelihood of participants copying the vignette explanation with the burden of unexpectedly being asked to remember it, we asked them to give their answer on the next page of the survey but allowed them to go back and re-read the explanation if they wished.

3.4 Gathering Baseline Perceptions and Attitudes

To avoid priming the survey closed by asking participants whether they felt that data shared with the assistant was more private or secure than that stored on their own personal computer, followed by standard questions about participants' general privacy and security practices and attitudes (an approach taken in prior work, such as [66]). To assess privacy attitudes we used the 10 questions from the core of the Internet Users' Information Privacy Concerns inventory (IUIPC) [56] that measure attitudes towards control, awareness, and collection in the context of online privacy. For security attitudes we used all six questions from the SA-6 inventory of security attitudes [30]. IUIPC was developed and validated based on data from a total of 742 interviews, and SA-6 was developed and validated based on a total of 687 survey responses.

3.5 Analysis

Before opening the survey, a priori power analysis revealed that the minimum number of participants required to detect a 'small' effect of size 0.2 across 3x8 groups at the $\alpha = 0.05$ level would be 693 (29 per group). We therefore ensured that the number of collected responses would exceed this level.

Calculations were carried out using Python, and a record of the pre-processing and statistical tests used is provided in the form of a Python notebook in the supplementary materials and OSF

repository. Given the presence of an underlying factor in the smart home survey that originally used Q1–8 from Cannizzaro et al. ('incident anxiety'), we were interested in testing for the presence of similar factor(s). This would indicate correlation between question responses that could then be taken into account. The suitability of the data for exploratory factor analysis (EFA) was verified via Bartlett's Test of Sphericity and the Kaiser-Meyer-Olkin Test (KMO). The significance of the first (3369.53, p=0.0) indicates that the data has a sufficient amount of intercorrelation for factor analysis, and the "meritorious" [40] KMO value (0.84) that the results may be useful in understanding the data via factor analysis.

We used ANOVA tests to identify the presence of significant effects on responses to the eight inventory questions followed with Tukey HSD tests, the latter of which corrects for Type I errors associated with testing multiple hypotheses simultaneously. These were used to identify the specific pairs of functionality or explanation levels that had significant differences in response means. Answers to baseline perception questions on privacy and security were summed to give an overall score for comparison. Pearson's ρ was used to calculate linear correlations. Answers to questions using a Likert scale were coded from 1 to 7 such that greater scores indicate greater concern, with questions reverse coded as appropriate. Mean differences between experiment conditions are reported using arrow notation to show the groups and direction of increase (e.g. $\Delta\bar{x}$ of 0.5 for 0E $\rightarrow$ 1E means that the average mean for 1E was 0.5 higher than for 0E).

For the participant-provided re-explanations, a sample of 400 responses was chosen to achieve good coverage of the data given the shorter nature of the responses, and contained a balance of the 24 different study conditions. One researcher analysed 200 of these responses to create a thematic codebook following Braun and Clarke [8]. The codebook was independently applied to the same sample by a second researcher, with a calculated agreement of $\kappa = 0.86$. While there is no agreed upon threshold for sufficiency when interpreting Cohen's Kappa, this value is large enough to suggest a reasonably descriptive and objective codebook [32, 47]. The remaining 200 responses in the sample were then divided between the researchers and coded using the codebook. Of particular interest during the qualitative analysis were the words and concepts used by participants to describe their understanding of voice assistants, drawing out the ideas that were most commonly internalised and the analogies used to understand this complex technology.

4 RESULTS

4.1 Demographics and Baseline Perceptions

A total of 1314 survey responses were received, with 24 responses rejected through the Prolific platform and excluded from the analysis. Reasons for this included incomplete answers to the survey questions, revoked consent, empty re-explanations, and verbatim repetition of the vignette explanation in participants' re-explanations.

This left 1290 full responses for analysis. Participants were aged between 18 and 83 (mean=35.4, median=34, σ=12.0), with 49.8% of participants identifying as women. A full breakdown of participant demographics is given in Table 2. As expected, baseline perceptions had an effect on the responses of participants to Q1–8. Participants that had higher general privacy concerns were more likely to show higher privacy and security concerns about the hypothetical assistants in Q4, 6, and 7 ($0.10 < \rho < 0.14$), but showed lower trust concerns in Q2 ($\rho = 0.13$). Greater baseline security attitudes resulted in greater concern about trust, privacy, and security for all questions except Q2, where there was a slight negative effect ($0.18 < \rho < 0.34$ and $\rho = -0.37$ respectively). In this respect, the focus of Q2 on the companies and organisations associated with the voice assistant may have changed the way that participants answered that particular question, tapping into negative perceptions of the major companies operating in this space. Age was weakly correlated with greater

Age	18–28	32.6%	(421)
	29-38	30.8%	(397)
	39-48	19.1%	(247)
	49-58	11.4%	(147)
	59+	4.6%	(59)
Gender Id.	Men	50.1%	(647)
	Women	49.8%	(643)
	Other/None	0.1%	(1)
Assistant	Amazon Alexa	55.7%	(719)
	Google Assistant	26.3%	(339)
	Siri	16.0%	(207)
	Other	2.0%	(26)
Computer Use at Work	$>\frac{2}{3}$ of time	59.3%	(766)
	$\frac{1}{3}$ to $\frac{2}{3}$ of time	13.6%	(176)
	$<\frac{1}{3}$ of time	13.2%	(170)
	None or N/A	13.9%	(179)

Table 2. Demographic information for survey participants.

trust concerns in Q1 ($\rho = 0.07$), and on average women rated the likelihood and impact of privacy and security events as higher than men in Q4, 5, 7, and 8 ($0.26 < \Delta\bar{x} < 0.38$). Owners of devices using Siri were much more likely to have privacy and security concerns than those using Alexa or the Google Assistant for Q6 and Q7 ($0.39 < \Delta\bar{x} < 0.52$). On the other hand, Alexa users were more likely to express trust concerns in Q2 than those using Siri ($\Delta\bar{x} = 0.37$). All results reported above were significant at the $p > 0.05$ level.

4.2 Exploratory Factor Analysis

After the Bartlett and KMO tests described in the methodology suggested that some patterns in the survey data may be caused by underlying factor(s), we conducted exploratory factor analysis. Applying Velicer's Minimum Average Partial method [88] to Questions 1–8 suggested the retention of a single factor with a proportional variance of 39%. Comparing to Cannizzaro et al. we found a similar distribution of eigenvalues, with only one eigenvalue significantly above 1 (the original method used for factor retention). In order to simplify the resulting factor and to allow easy comparison with prior work we retained the three questions loading at or above 0.7 (Q4, Q7, and Q8) and computed a weighted factor score for each participant. Unlike in prior work, this factor was not found to significantly correlate with trust in competence: the variables present in the underlying factor are the same as [10] with the exception of Q1, which was not included based on analysis of the current data set. These questions are centred around the likelihood of an adverse event and resulting physical risk ('incident anxiety'), and the similarity of questions and context suggest that a closely related phenomena exists with VAs. Factor loadings are given in Table 3. The results of ANOVA tests showed that there were significant interactions between incident anxiety and explanation content ($p = 0.0027$), online functionality ($p < 0.001$), and the interaction between the two ($p < 0.014$). Follow up Tukey HSD tests showed significant mean differences between all levels of online functionality, as well as between privacy explanations and those with security or security and reliability. These are included at the bottom of Table 4. These findings suggest that the introduction of online functionality and the absence of privacy explanations both contribute to a general sense of incident anxiety in users.

Question	Factor Loading
Q1 Trust in competence	0.51
Q2 Trust in benevolence	-0.57
Q3 Trust in integrity	0.51
Q4 Likelihood of a privacy breach	**0.75**
Q5 Impact of a privacy breach	0.62
Q6 Impact of controversy	0.50
Q7 Likelihood of incident	**0.78**
Q8 Impact of incident	**0.70**

Table 3. Factor loadings for the eight trust questions, with items in bold retained for calculating participant factor scores. Note when comparing with [10] that values have been inverted such that positive values indicate greater concern.

4.3 Differences Between Explanation Groups

In order to further explore the survey data we subsequently conducted a series of two way ANOVA tests on the functionality and explanation factors, which unsurprisingly revealed that both explained a significant amount of variance between the survey treatments for all questions in the inventory except Q2. There were no significant interactions between the effects of functionality and explanation for the individual questions except at a very small scale when considering the impact of security breaches (Q8, ω^2=0.009, p=0.025). Follow-up Tukey tests revealed a number of significant differences between the means of individual functionality and explanation content treatments. These are shown in Table 4. We briefly describe these results here to the extent that they extend and supplement the factor analysis.

The introduction of external entities was unsurprisingly considered to come at the cost of trust, privacy, and security. Significant mean differences were observed between offline and online assistants (i.e. 0E → 1E or 3E) in nine of the ten questions, in line with previous literature (e.g. [96]). What was more interesting were different ways in which this manifested: for security each additional level of external entities produced a significant increase in concerns (0E → 1E → 3E). For privacy the addition of any external entity produced an increase in concerns, with third party entities causing more than first party ones but not with enough of a distinction to show significant differences between first and third parties. This shows that people can and do distinguish between the number of external parties involved, with differing granularities of concern depending on what is at stake. This is true whether or not participants were aware of the existence of third party software on their own voice assistants or other devices, with the responses to Q10 largely suggesting the latter.

Vignette explanations that only contained privacy information were often associated with lower levels of concern than those that did not contain privacy information (i.e. the removal of privacy explanations often increased concerns about trust, privacy, *and* security). Another key signal in the results was that for Q1 (which asked about trust in competence), the addition of trust information to almost any explanation without it led to significantly higher trust concerns. This effect was present even when both explanations contained privacy information, suggesting that it was a separate effect from the one described above.

The two questions gauging trust in benevolence and integrity (Q2 and Q3) were only associated with one significant effect between them (0E → 3E). Given the focus of these items on device vendors, this indicates that the extent to which participants trust a device's manufacturer is not affected by the way that its use is explained, and only slightly by the presence of online capability. Notably, by distinguishing only between 0E and 3E, participants seem to trust online first party

Question	Conditions	Mean Difference
Q1—Trust in competence (general trust)	0E → 1E	0.50**
	0E → 3E	0.51**
	F → TS	0.53*
	P → T	0.81**
	P → TP	0.84**
	P → TPS	0.67**
	P → TS	0.85**
	P → S	0.66**
	PS → TP	0.53*
	PS → TS	0.54*
Q2—Trust in benevolence (general trust)	None	-
Q3—Trust in integrity (general trust)	0E → 3E	0.33*
Q4—Likelihood of a privacy breach (privacy)	0E → 1E	0.53**
	0E → 3E	0.61**
	P → T	0.66**
	P → TS	0.69**
	P → S	0.61**
Q5—Impact of a privacy breach (privacy)	0E → 1E	0.32*
	0E → 3E	0.48**
	P → T	0.58*
	P → TS	0.53*
Q6—Impact of controversy (privacy)	0E → 3E	0.33*
Q7—Likelihood of incident (security)	0E → 1E	0.56**
	0E → 3E	0.83**
	1E → 3E	0.27*
	P → TS	0.60*
Q8—Impact of incident (security)	0E → 1E	0.53**
	0E → 3E	0.82**
	1E → 3E	0.29*
Q9—Privacy Relative to Own Device	0E → 3E	0.36**
	1E → 3E	0.28*
Q10—Security Relative to Own Device	0E → 3E	0.39**
	1E → 3E	0.27*
	P → TS	0.50*
Underlying factor—"incident anxiety"	0E → 1E	1.21**
	0E → 3E	1.67**
	1E → 3E	0.47*
	P → TS	1.30**
	P → S	1.15*

Table 4. Significant mean differences between survey treatments. Higher means indicate greater concern. A * indicates an adjusted $p \leq 0.05$ and ** an adjusted $p \leq 0.005$. Descriptions of the questions are taken from [10]. Tests that did not meet the 0.05 significance threshold are not shown.

services as much as they trust the device itself, with the addition of external entities (e.g. third-party skills providers) triggering a drop in trust.

4.4 User-Written Re-explanations

Most participants were able to explain the concept of using wake words to initiate interactions with the voice assistant. While this was often linked with the assistant beginning to 'listen' for commands, a significant number of responses explicitly linked the activation/listening of the assistant with the *recording* of speech for either wake word detection or the interpretation of the commands. The proportion of responses coded for wake words and recording did not significantly vary between connectivity levels or explanation content, except for vignette explanations covering all content types (*RPS*) which were consistently higher.

The most common way that participants referred to the assistant was simply as 'the assistant', followed by demonstrative pronouns (e.g. 'it'), and personal pronouns (e.g. 'she'). Referring to the assistant as if it were a machine, computer or robot was much less common. Some participants included a description of how the assistant matched key words in their requests to possible commands and variables, but this more mechanical way of thinking did not seem to bring with it a noticeable shift in the way that participants did (or did not) anthropomorphise the device. In re-explanations that gave less detailed descriptions about the assistant, it was common for responses to say that the assistant "works out what command you've [said]" [P169] without reference to how this was done. In other cases, the concept of an algorithm was used as a catch-all term to describe the processing performed by the assistant ("algorithms figure out what is being asked" [P260]). This suggests that while participants had a functional understanding of what was happening, they lacked the detailed background knowledge required to enrich the explanation.

Others used their own voice assistant as a point of comparison, often subtly widening the scope of their re-explanation to talk about the entire class of devices: "To activate the voice assistant you must speak clearly, use the voice assistants name to activate and give a command or question" [P91]. Another strategy employed was to use other systems they were familiar with to encapsulate areas of functionality ("ask it questions like you would to a search engine like Google" [P25]). Despite this variation over the specifics of how the assistant interpreted requests, nearly every participant identified at least one aspect of the vignettes in their own explanations and very few made incorrect statements about the assistant.

A range of concerns arising from the content of the vignettes was present in participants' re-explanations. People often described how inaccuracies in the voice recognition process might lead to errors when the voice assistant 'misheard' an instruction, performing the wrong action or the right action with incorrect parameters. Participants also described the possibility that the assistant would be unable to "distinguish between what is appropriate for recording and what isn't" [P360] and start recording when it had not been asked to do anything. Given the similarities in their operation, the effect of reliability explanations on perceptions was often clear: "I'm not the best at explaining, but [I] would say it's a similar idea to an Alexa but not as trusting. And that I would be worried because of interference that it couldn't work properly." [P201]. Despite the fact that all participants owned a voice assistant themselves, several expressed that they would be hesitant to use the assistant described to them, often citing the apparent insecurity or unreliability of the assistant: "I wouldn't be comfortable using this personally." [P81].

In a smaller amount of these cases the errors of the assistant were directly linked to privacy concerns, such as sending personal data to the wrong external party. Those receiving privacy content as part of the offline vignette (0E), often instead mentioned the privacy *benefits* of an assistant that was not connected to the internet. While speaker recognition appeared in some re-explanations from participants who had seen security information, only a small number voiced

concerns over the potential for misuse without it. This was the only reason given for concerns over fraud or theft via the assistant—despite being described to a sixth of the participants, no user-given re-explanation mentioned 'skill squatting' or skills collecting data without permission.

5 DISCUSSION: CRAFTING EXPLANATIONS FOR VOICE ASSISTANTS

5.1 The Presence of First and Third Party Entities

The results suggest that online functionality is considered to be less trustworthy, private, and secure. On the surface this is unsurprising given that online functionality introduces additional complexity to a system and broadens its attack surface. But the fact that participants distinguished between first and third parties in privacy and security perceptions *was* unexpected given that many concerns identified in the re-explanations focused on local actions (e.g. inappropriate recording and speaker recognition), and the focus on user-device interactions in the security and privacy literature [27]. In explaining it we might turn to wider perceptions about connectedness and smartness, where the presence of app stores clearly demarcates the line between the involvement of first and third party entities. Prior work has shown that users identify app stores as potential risks [80], and it is possible they applied the same logic in the present study; local actions would therefore be concerning because data from them is expected to be shared. This would, however, conflict with other work which found that users often did not mention third parties when describing VA interactions involving third party skills [1]. Given the significant differences found *between* first and third party entities and the fact that this was largely absent from the re-explanations, it seems probable that users do distinguish between first and third party entities even if the associated mental models are not sufficiently detailed to explain why. Overall the nature of online functionality had less of an impact than the way that the assistant was explained, with mean response differences between connectivity conditions almost always lower than between explanation components.

5.2 The Provision of Trust, Privacy, and Security Information in Explanations

The results show a stark contrast in the way that the presence of privacy and trust information in explanations affect people's subsequent perceptions of a voice assistant; privacy explanations alleviated privacy concerns, while trust explanations in Q1 aggravated trust concerns. This is also reflected in the large number of trust concerns relative to privacy concerns in re-explanations and the small proportion of concerns over recording which were linked to privacy concerns. In line with previous findings about privacy revelations delivered in isolation [83] (e.g. without an associated increase in understanding or control mechanisms [81]), we had expected that providing participants with privacy explanations would have resulted in them feeling less reassured in the their privacy. It is possible that this was due to the fact that the privacy explanations given were less problematic than participants expected, which would be in line with the misplaced user threat assessments described by the background literature. An alternative explanation is that evidence that privacy risks had been considered was itself enough to assuage concerns. Another important thing to note is that failures of trust in competence are often associated with immediate negative outcomes (such as the assistant not working if disconnected from the internet), whereas failures of privacy are often much less immediately tangible, with longer term effects and the potential for leaked data to be exploited in the future (e.g. information sent to the wrong company then being sold or stolen).

Unlike the response to trust explanations seen in Q1, answers to the other two questions about trust (Questions 2 and 3) did not appear to be significantly impacted by trust explanations at all. Instead, the greatest influence on trust concerns was the brand of assistant that participants owned, followed by the addition of third party functionality. This highlights the unusual relationship of

trust when it comes to voice assistants. Conventionally, trust in a manufacturer is related to but distinct from trust in a product—trustworthy manufacturers are more likely to make trustworthy products. However, the complete control that voice assistant manufacturers have over the devices they sell and associated ecosystems means that trust in a voice assistant and trust in its manufacturer should rationally be the same, or a very similar, thing; one cannot trust Alexa without also trusting Amazon, and vice versa. This is also supported by the fact that for the more vendor–focused Q2 and Q3 the significant difference came with the addition of third party functionality, with no such difference between offline and first party functionality.

These differences between users of different devices aligns with findings by Abdi et al. that voice assistant users are unlikely to have developed mental models beyond their own voice assistant [1], suggesting that they may be projecting concerns over their own device onto the study vignettes. Interpreting these results also raises a question of causality—do people choose products like Siri over alternatives *because* they had greater general concerns over privacy and security, or do their concerns later align with public perceptions (e.g. valuing privacy more highly because Siri is marketed as being privacy preserving)? That product providers have the greatest effect on trust echoes the results of other recent work on anthropomorphism in voice assistants [82]. This raises interesting questions about competition and diversity in the sector, suggesting that new entrants distinguishing themselves with functionality (like the Mycroft assistant being "private and open"[3]) are much less likely to gain traction and become trusted than a voice assistant made by an established brand that people already trust. An exception to this might be moving functionality offline, something that Apple has announced will begin happening with Siri as of iOS 15[4].

As described above, there was an underlying incident anxiety factor similar to [10], centred around the likelihood of privacy and security breaches, as well as the impact of the latter. Interestingly, when comparing the results to Cannizzaro et al. trust in competence (Q1) did not emerge as a significant component in the factor. Further work is needed in order to determine the extent to which this is a result of the shift in focus from smart homes to voice assistants, different participant samples, and/or changes in smart homes between the times that the studies were carried out.

5.3 Targeting Privacy and Trust Concerns

At a high level, the participant explanations show a general trend whereby people were able to interpret and understand common behaviours and drawbacks associated with voice assistants, such as the use of wake words and the recording of speech. Problems with devices failing to understand user requests and failing to distinguish requests made by different users were also relatively well described by participants. At the same time, despite the blurred boundaries between participants' own devices and the hypothetical study device, explanation components were almost exclusively given when they were also present in the vignette shown to participants. One interpretation would be that this knowledge is usually backgrounded and/or acquired unconsciously but can be explained on demand: people may not think about these concepts day-to-day, but the vignettes gave them the mental scaffolding to express what they already knew. We hypothesise that where participants were unfamiliar with a concept presented to them in a vignette they were likely to omit it from their subsequent re-explanation, hence the lack of more complex security threats in responses.

At the same time, analysis of participant's perceptions shows that privacy concerns are mitigated by privacy explanations whereas trust/competence explanations exacerbate trust concerns. So if trust explanations require the most improvement from the baseline presented in this study, what is

[3]https://mycroft.ai/

[4]"Siri adds on-device speech recognition, so the audio of your requests is processed on your iPhone or iPad by default. And on-device processing also means Siri can perform many tasks without an internet connection." (https://www.apple.com/uk/ios/ios-15-preview/)

to be done? The incidence rate of key words in the re-explanations suggests that more can be done to improve users' mental models of how voice assistants process speech, but this mainly affects conversational repair. As with privacy concerns where prior work has shown the ineffectiveness of and feelings of dejected acceptance caused by warnings where no remedial action can be taken [83], designers should be wary of the same trap here; there is little users can do to act on feelings of trust or distrust towards current voice assistants.

As a result, we posit that the competence concerns identified by the study results represent important priorities for voice assistant *manufacturers* to address via software and platform updates. In the interim, we see the use of multimodal interaction as a key tool to mitigate the effects of these concerns. Encouraging the use of design elements such as visual feedback and smartphone confirmation dialogues gives users the ability to verify the correct operation of their assistant in ways that feel familiar and comfortable, while still allowing for convenient hands-free operation most of the time. This approach would also address more widespread problems around low user confidence when shopping via voice assistants [1]. Allowing users to balance connectivity with functionality is another option, and will likely ease concerns even if not ultimately used by the majority of users [52].

For privacy concerns, the way forward seems clear. If outlining the extent of privacy risks when using voice assistants reduces the concerns that they have, then providing these explanations is worthwhile. Further work is required to determine exactly why this is the case, as discussed above a possible reason is that people may also be more accustomed to seeing privacy warnings/agreeing to privacy policies which then have little noticeable impact on them, and privacy explanations alone may therefore not be appropriate for increasing knowledge and understanding (see [81]). This points to the adoption of different approaches depending on how well concepts are understood. Quick snippets delivered just-in-time would serve to remind users of important considerations before taking an action (e.g. when linking accounts), whereas more in-depth explanations supported by displays and other devices might be more appropriate for concepts that users are less likely to have an intuitive grasp of (and would therefore be unlikely to acquire from a quick analogy).

5.4 Guiding Users with Analogies

In their explanations, participants readily compared the hypothetical voice assistant in the vignette to their own devices and other digital technologies present in everyday life. A clear next step would be the development of analogies or short explanations that increase understanding about other aspects of how voice assistants work (such as using 'building blocks' to represent the parsing of requests), which could be delivered when the device fails to understand the user's intent as a form of conversational grounding [13]. This would also be appropriate when targeting poorly understood features or risks, such as the potential for 'skill squatting' [46]. Given the presence of anthropomorphism and potential uses of social reasoning present in the results, there may be a case for social analogies as well as technical ones (e.g. comparing microtransactions within skills/actions to buying a used car in terms of the social mechanics involved).

The prioritisation of explanation concepts is also likely to shift over time. The results of the study suggest that participants, likely primed by experiences with their own voice assistants, saw failings of speech recognition as the greatest concern. But as speech recognition improves and conversation models adapt to enable more natural exchanges (e.g. [6]), helping people better understand the functionality offered by voice assistants will become of greater importance; once the technology becomes sufficiently accurate "[end-users] do not expect to have that level of insight because they have no practical need of it" [68].

Indeed, these techniques will become increasingly important as we begin to make assistants that guide people through more complex requests that might include follow-on questions that narrow

the scope of an initial enquiry (e.g. styling an assistant as a 'coach' [90] or similar). Here these kinds of short, contextual snippets can not only convey the limitations of the assistant, but also the assumptions and boundaries of the dialectic process itself.

5.5 The Future of Voice Assistant Explanations

Based on our findings, we ask ourselves the extent to which explanations for voice assistants should focus on reassuring users versus describing the drawbacks of the technology. Going forward we see two main applications for explanations in voice assistants. The first, as demonstrated in the vignettes, is to convey the benefits and drawbacks of contemporary technologies, highlighting relevant information at the point of use that can be further supplemented by other media (e.g. information cards on a display). Following work by Schaub et al. [78], these explanations should be short and specific, delivered in-context during an interaction. They should also present users with meaningful opportunities to change their actions and/or remedy potential problems. For example, when explaining the potential privacy implications of sending information to a third party skill, offline processing could be offered as an alternative, with the privacy-reliability exchange briefly outlined. We present the further development of these explanations as an open challenge to the research community, and look forward to making our own contributions in future work.

The second is to convey technical mechanisms put in place to protect users. Given public perceptions and media reports, convincing users as to the efficacy of a protection mechanism is likely to be a challenging task, requiring the use of explanations subject to the caveats given above. For instance, while Amazon has a mechanism to allow users to delete audio recordings associated with previous voice commands, there have been public outcries reported in the media as many people are not aware that this is possible or that recordings are stored[5]. These types of explanations will become increasingly important as the mechanisms to protect users in voice assistants become more sophisticated. One example would be the introduction of controls over information flows across the voice assistant ecosystem (including to third parties as in our 3E condition), that could come with strong default configurations [2] and/or learn user preferences over time [94]. For these controls it would be crucial to explain in an accessible way and at the right time what may happen to users' data. A second key example is the introduction of provable privacy and security properties to future voice assistants, constraining the behaviour of their underlying AI/ML models. Research on the verification and dynamic monitoring of such models is currently a hot topic, with promising results to date [7, 19, 37].

5.6 Limitations

The greatest limitation of the study lies in the viewpoints and devices of its participants. To reduce complexity respondents were all resident in the UK, and as a result used a small set of devices. In future experiments we aim to further explore the ways in which the perceptions reported here might differ across the globe. Another artefact of the survey design was that some participants were exposed to more concepts than others, and it is possible (though unlikely) that a saturation point existed after which new concepts were not retained. This would manifest in the results as a lower incidence of certain codes amongst groups with longer explanations (e.g. 3E/TPS), but this did not appear to be the case during analysis.

Methodologically, there are also many possible ways to deliver explanations, of which the format presented here is but one. Different approaches to the challenges we discuss will yield different results; the goal of this paper was not to determine the best approach to voice assistant explanations,

[5]e.g. "How to listen to hidden Alexa recordings of your conversations – and then delete them" (https://www.thesun.co.uk/tech/10987517/listen-alexa-recordings-how-amazon-delete/)

but rather to determine their general utility and inform their future refinement. By focusing on trust as competence when designing the vignette explanations we were not able to manipulate trust as benevolence or integrity when answering the research questions, and the results suggest that participants subsequently relied on experiences with the vendors of their own devices when answering those questions. The lack of an existing validated questionnaire for the concepts being explored, while mitigated through careful re-use of existing questions, is unfortunate and remains an opportunity for future work.

6 CONCLUSION

Voice assistants occupy an unfortunate meeting point between sleek smart home gadgetry and lower bandwidth voice interfaces. This can make understanding how they work difficult, and makes it harder to assess the benefits and drawbacks of their use. By studying changes in peoples' perceptions of different voice assistants described in various ways, we show the vastly different responses to different kinds of explanations.

Combining this with participants' own re-explanations gave a unique opportunity to see the functionality—and concerns—that were reflected back, highlighting gaps in people's understanding and opening up a discussion on issues that explanations are apt to mitigate, as well as those they are not. The findings suggest that the introduction of online functionality and the absence of privacy explanations both contribute to a general sense of incident anxiety, and that users differentiate and respond to relatively subtle combinations of first and third party entities. The contrasting reactions to privacy and reliability explanations is an important one that can be used to shape the development of future voice assistants and associated explanations, and shows the importance of work by developers towards decreasing misfires and misinterpretations.

Reflecting on the unique role that trust plays in respect to voice assistants, we argue that trust in a voice assistant is logically equivalent to trust in its manufacturer, and lay out the challenges faced by new entrants to the voice assistant market. While the dominance of trust in manufacturers as a predictor of trust in voice assistants may appear to be bad news given the track records of the companies operating in this space, positive responses to offline functionality may yet prove to be good news for everyone (and particularly the privacy concerned).

ACKNOWLEDGMENTS

We would like to thank Carlota Vazquez Gonzalez for her assistance with data curation. This work is part of the EPSRC-funded Secure AI Assistants project (grant EP/T026723/1).

REFERENCES

[1] Noura Abdi, Kopo Ramokapane, and Jose 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] Noura Abdi, Xiao Zhan, Kopo M Ramokapane, and Jose Such. 2021. Privacy Norms for Smart Home Personal Assistants. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[3] Gavin Abercrombie, Amanda Cercas Curry, Mugdha Pandya, and Verena Rieser. 2021. Alexa, Google, Siri: What are Your Pronouns? Gender and Anthropomorphism in the Design and Perception of Conversational Assistants. arXiv:2106.02578 [cs.AI]

[4] Abbas Acar, Hossein Fereidooni, Tigist Abera, Amit Kumar Sikder, Markus Miettinen, Hidayet Aksu, Mauro Conti, Ahmad-Reza Sadeghi, and Selcuk Uluagac. 2020. Peek-a-Boo: I See Your Smart Home Activities, Even Encrypted!. In *Proceedings of the 13th ACM Conference on Security and Privacy in Wireless and Mobile Networks* (Linz, Austria) *(WiSec '20)*. Association for Computing Machinery, New York, NY, USA, 207–218. https://doi.org/10.1145/3395351.3399421

[5] Benjamin Andow, Samin Yaseer Mahmud, Wenyu Wang, Justin Whitaker, William Enck, Bradley Reaves, Kapil Singh, and Tao Xie. 2019. Policylint: investigating internal privacy policy contradictions on Google play. In *28th {USENIX} Security Symposium ({USENIX} Security 19)*. 585–602.

[6] Benett Axtell and Cosmin Munteanu. 2021. Tea, Earl Grey, Hot: Designing Speech Interactions from the Imagined Ideal of Star Trek. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 249, 14 pages. https://doi.org/10.1145/3411764.3445640

[7] Ben Batten, Panagiotis Kouvaros, Alessio Lomuscio, and Yang Zheng. 2021. Efficient neural network verification via layer-based semidefinite relaxations and linear cuts. In *30th International Joint Conference on Artificial Intelligence (IJCAI-21), accepted.*

[8] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. (2012).

[9] Matt Burgess. 2017. *Google stops 'What is the Whopper burger?' ad triggering Google Home.* https://www.wired.co.uk/article/google-home-burger-king-ad

[10] Sara Cannizzaro, Rob Procter, Sinong Ma, and Carsten Maple. 2020. Trust in the smart home: Findings from a nationally representative survey in the UK. *PLOS ONE* 15, 5 (05 2020), 1–30. https://doi.org/10.1371/journal.pone.0231615

[11] Yi-Shyuan Chiang, Ruei-Che Chang, Yi-Lin Chuang, Shih-Ya Chou, Hao-Ping Lee, I-Ju Lin, Jian-Hua Jiang Chen, and Yung-Ju Chang. 2020. Exploring the design space of user-system communication for smart-home routine assistants. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[12] 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.* Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376551

[13] Janghee Cho and Emilee Rader. 2020. The Role of Conversational Grounding in Supporting Symbiosis Between People and Digital Assistants. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1, Article 033 (May 2020), 28 pages. https://doi.org/10.1145/3392838

[14] Meghan Clark, Mark W. Newman, and Prabal Dutta. 2017. Devices and Data and Agents, Oh My: How Smart Home Abstractions Prime End-User Mental Models. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 44 (Sept. 2017), 26 pages. https://doi.org/10.1145/3132031

[15] Katriel Cohn-Gordon, Cas Cremers, Benjamin Dowling, Luke Garratt, and Douglas Stebila. 2020. A formal security analysis of the signal messaging protocol. *Journal of Cryptology* 33, 4 (2020), 1914–1983.

[16] Paul Coulton and Joseph Galen Lindley. 2019. More-than human centred design: Considering other things. *The Design Journal* 22, 4 (2019), 463–481.

[17] Benjamin R Cowan, Holly P Branigan, Habiba Begum, Lucy McKenna, and Eva Szekely. 2017. They Know as Much as We Do: Knowledge Estimation and Partner Modelling of Artificial Partners.. In *CogSci*.

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

[19] Natalia Criado and Jose Such. 2016. Selective norm monitoring. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)*. AAAI Press, 208–214.

[20] Henry K Dambanemuya and Nicholas Diakopoulos. 2020. "Alexa, what is going on with the impeachment?" Evaluating smart speakers for news quality. In *Proc. Computation+ Journalism Symposium*.

[21] G. Danezis, J. Domingo-Ferrer, M. Hansen, J. Hoepman, D. Metayer, R. Tirtea, and S. Schiffner. 2014. Privacy and Data Protection by Design-from policy to engineering. *ENISA* (2014).

[22] Michael A. DeVito, Darren Gergle, and Jeremy Birnholtz. 2017. *"Algorithms Ruin Everything": #RIPTwitter, Folk Theories, and Resistance to Algorithmic Change in Social Media.* Association for Computing Machinery, New York, NY, USA, 3163–3174. https://doi.org/10.1145/3025453.3025659

[23] Philip R Doyle, Leigh Clark, and Benjamin R. Cowan. 2021. *What Do We See in Them? Identifying Dimensions of Partner Models for Speech Interfaces Using a Psycholexical Approach.* Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445206

[24] Anna Dunin-Underwood. 2020. Alexa, can you keep a secret? Applicability of the third-party doctrine to information collected in the home by virtual assistants. *Information & Communications Technology Law* 29, 1 (2020), 101–119. https://doi.org/10.1080/13600834.2020.1676956

[25] Jide Edu, Xavier Ferrer-Aran, Jose Such, and Guillermo Suarez-Tangi. 2021. SkillVet: Automated Traceability Analysis of Amazon Alexa Skills. *IEEE Transactions on Dependable and Secure Computing (TDSC)* (2021).

[26] Jide Edu, Xavier Ferrer Aran, Jose Such, and Guillermo Suarez-Tangil. 2022. *Measuring Alexa Skill Privacy Practices across Three Years.* ACM, 670–680.

[27] Jide Edu, Jose Such, and Guillermo Suarez-Tangil. 2020. Smart Home Personal Assistants: A Security and Privacy Review. *ACM Comput. Surv.* 53, 6, Article 116 (Dec. 2020), 36 pages. https://doi.org/10.1145/3412383

[28] Pardis Emami-Naeini, Henry Dixon, Yuvraj Agarwal, and Lorrie Faith Cranor. 2019. *Exploring How Privacy and Security Factor into IoT Device Purchase Behavior.* Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300764

[29] Motahhare Eslami, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton, and Alex Kirlik. 2016. First I" like" it, then I hide it: Folk Theories of Social Feeds. In *Proceedings of the 2016 cHI conference on human factors in computing systems*. 2371–2382.

[30] 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)*.

[31] Joel E. Fischer, Stuart Reeves, Martin Porcheron, and Rein Ove Sikveland. 2019. Progressivity for Voice Interface Design. In *Proceedings of the 1st International Conference on Conversational User Interfaces* (Dublin, Ireland) *(CUI '19)*. Association for Computing Machinery, New York, NY, USA, Article 26, 8 pages. https://doi.org/10.1145/3342775.3342788

[32] Joseph L Fleiss, Bruce Levin, and Myunghee Cho Paik. 2013. *Statistical methods for rates and proportions.* john wiley & sons.

[33] 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* (Montreal QC, Canada) *(CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3170427.3188448

[34] Seda Gürses and Jose M Del Alamo. 2016. Privacy engineering: Shaping an emerging field of research and practice. *IEEE Security & Privacy* 14, 2 (2016), 40–46.

[35] Robert R Hoffman, Shane T Mueller, Gary Klein, and Jordan Litman. 2018. Metrics for explainable AI: Challenges and prospects. *arXiv preprint arXiv:1812.04608* (2018).

[36] Danny Yuxing Huang, Noah Apthorpe, Frank Li, Gunes Acar, and Nick Feamster. 2020. IoT Inspector: Crowdsourcing Labeled Network Traffic from Smart Home Devices at Scale. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 2, Article 46 (June 2020), 21 pages. https://doi.org/10.1145/3397333

[37] Xiaowei Huang, Marta Kwiatkowska, Sen Wang, and Min Wu. 2017. Safety verification of deep neural networks. In *International conference on computer aided verification*. Springer, 3–29.

[38] Yue Huang, Borke Obada-Obieh, and Konstantin (Kosta) Beznosov. 2020. *Amazon vs. My Brother: How Users of Shared Smart Speakers Perceive and Cope with Privacy Risks.* Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376529

[39] Carlos Jensen and Colin Potts. 2004. *Privacy Policies as Decision-Making Tools: An Evaluation of Online Privacy Notices.* Association for Computing Machinery, New York, NY, USA, 471–478. https://doi.org/10.1145/985692.985752

[40] Henry F Kaiser and John Rice. 1974. Little jiffy, mark IV. *Educational and psychological measurement* 34, 1 (1974), 111–117.

[41] Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. 2015. "my data just goes everywhere:" user mental models of the internet and implications for privacy and security. In *Eleventh Symposium On Usable Privacy and Security ({SOUPS} 2015)*. 39–52.

[42] Thomas Kannampallil, Joshua M Smyth, Steve Jones, Philip RO Payne, and Jun Ma. 2020. Cognitive plausibility in voice-based AI health counselors. *NPJ digital medicine* 3, 1 (2020), 1–4.

[43] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. 2009. A "Nutrition Label" for Privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security* (Mountain View, California, USA) *(SOUPS '09)*. Association for Computing Machinery, New York, NY, USA, Article 4, 12 pages. https://doi.org/10.1145/1572532.1572538

[44] Gerwin Klein, Kevin Elphinstone, Gernot Heiser, June Andronick, David Cock, Philip Derrin, Dhammika Elkaduwe, Kai Engelhardt, Rafal Kolanski, Michael Norrish, et al. 2009. seL4: Formal verification of an OS kernel. In *Proceedings of the ACM SIGOPS 22nd symposium on Operating systems principles*. 207–220.

[45] Jeamin Koo, Jungsuk Kwac, Wendy Ju, Martin Steinert, Larry Leifer, and Clifford Nass. 2015. Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *International Journal on Interactive Design and Manufacturing (IJIDeM)* 9, 4 (2015), 269–275.

[46] Deepak Kumar, Riccardo Paccagnella, Paul Murley, Eric Hennenfent, Joshua Mason, Adam Bates, and Michael Bailey. 2018. Skill Squatting Attacks on Amazon Alexa. In *27th USENIX Security Symposium (USENIX Security 18)*. USENIX Association, Baltimore, MD, 33–47. https://www.usenix.org/conference/usenixsecurity18/presentation/kumar

[47] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.

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

[49] Dave Lee. 2018. *Amazon promises fix for creepy Alexa laugh.* https://www.bbc.co.uk/news/technology-43325230

[50] Sunok Lee, Sungbae Kim, and Sangsu Lee. 2019. "What Does Your Agent Look like?": A Drawing Study to Understand Users' Perceived Persona of Conversational Agent. In *Extended Abstracts of the 2019 CHI Conference on Human Factors*

in Computing Systems (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312796

[51] Christopher Lentzsch, Sheel Jayesh Shah, Benjamin Andow, Martin Degeling, Anupam Das, and William Enck. 2021. Hey Alexa, is this Skill Safe?: Taking a Closer Look at the Alexa Skill Ecosystem. In *28th Annual Network and Distributed System Security Symposium (NDSS 2021). The Internet Society*.

[52] Joseph Lindley, Paul Coulton, and Rachel Cooper. 2017. Why the internet of things needs object orientated ontology. *The Design Journal* 20, sup1 (2017), S2846–S2857.

[53] Irene Lopatovska, Katrina Rink, Ian Knight, Kieran Raines, Kevin Cosenza, Harriet Williams, Perachya Sorsche, David Hirsch, Qi Li, and Adrianna Martinez. 2018. Talk to me: Exploring user interactions with the Amazon Alexa. *Journal of Librarianship and Information Science* (2018).

[54] Irene Lopatovska and Harriet Williams. 2018. Personification of the Amazon Alexa: BFF or a Mindless Companion. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval* (New Brunswick, NJ, USA) *(CHIIR '18)*. Association for Computing Machinery, New York, NY, USA, 265–268. https://doi.org/10.1145/3176349.3176868

[55] Maria Madsen and Shirley Gregor. 2000. Measuring human-computer trust. In *11th australasian conference on information systems*, Vol. 53. Citeseer, 6–8.

[56] 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.

[57] Nathan Malkin, Julia Bernd, Maritza Johnson, and Serge Egelman. 2018. "What Can't Data Be Used For?" Privacy Expectations about Smart TVs in the US. In *Proceedings of the 3rd European Workshop on Usable Security (EuroUSEC), London, UK*.

[58] Karola Marky, Alexandra Voit, Alina Stöver, Kai Kunze, Svenja Schröder, and Max Mühlhäuser. 2020. "I Don't Know How to Protect Myself": Understanding Privacy Perceptions Resulting from the Presence of Bystanders in Smart Environments *(NordiCHI '20)*. Association for Computing Machinery, New York, NY, USA, Article 4, 11 pages. https://doi.org/10.1145/3419249.3420164

[59] Roger C Mayer, James H Davis, and F David Schoorman. 1995. An integrative model of organizational trust. *Academy of management review* 20, 3 (1995), 709–734.

[60] Sarah Mennicken, Oliver Zihler, Frida Juldaschewa, Veronika Molnar, David Aggeler, and Elaine May Huang. 2016. "It's like Living with a Friendly Stranger": Perceptions of Personality Traits in a Smart Home. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) *(UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 120–131. https://doi.org/10.1145/2971648.2971757

[61] Brent Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. 2016. The ethics of algorithms: Mapping the debate. *Big Data & Society* 3, 2 (2016), 2053951716679679.

[62] Youngme Moon. 2000. Intimate Exchanges: Using Computers to Elicit Self-Disclosure from Consumers. *Journal of Consumer Research* 26, 4 (03 2000), 323–339. https://doi.org/10.1086/209566

[63] Francesca Mosca and Jose Such. 2021. ELVIRA: An explainable agent for value and utility-driven multiuser privacy. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. 916–924.

[64] Francesca Mosca and Jose Such. 2022. An explainable assistant for multiuser privacy. *Autonomous Agents and Multi-Agent Systems (JAAMAS)* 36, 1 (2022), 1–45.

[65] Christine Murad and Cosmin Munteanu. 2020. *Designing Voice Interfaces: Back to the (Curriculum) Basics*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376522

[66] Pardis Emami Naeini, Sruti Bhagavatula, Hana Habib, Martin Degeling, Lujo Bauer, Lorrie Faith Cranor, and Norman Sadeh. 2017. Privacy expectations and preferences in an {IoT} world. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*. 399–412.

[67] Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. 1994. Computers Are Social Actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, Massachusetts, USA) *(CHI '94)*. Association for Computing Machinery, New York, NY, USA, 72–78. https://doi.org/10.1145/191666.191703

[68] Tommy Nilsson, Andy Crabtree, Joel Fischer, and Boriana Koleva. 2019. Breaching the future: understanding human challenges of autonomous systems for the home. *Personal and Ubiquitous Computing* 23, 2 (2019), 287–307.

[69] Antti Oulasvirta, Aurora Pihlajamaa, Jukka Perkiö, Debarshi Ray, Taneli Vähäkangas, Tero Hasu, Niklas Vainio, and Petri Myllymäki. 2012. Long-Term Effects of Ubiquitous Surveillance in the Home. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (Pittsburgh, Pennsylvania) *(UbiComp '12)*. Association for Computing Machinery, New York, NY, USA, 41–50. https://doi.org/10.1145/2370216.2370224

[70] Paul A Pavlou and Mendel Fygenson. 2006. Understanding and predicting electronic commerce adoption: An extension of the theory of planned behavior. *MIS quarterly* (2006), 115–143.

[71] Sabine Payr. 2013. Virtual butlers and real people: Styles and practices in long-term use of a companion. In *Your Virtual Butler*. Springer, 134–178.

[72] James Pierce and Carl DiSalvo. 2018. *Addressing Network Anxieties with Alternative Design Metaphors.* Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174123

[73] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. *Voice Interfaces in Everyday Life.* Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3174214

[74] Alisha Pradhan, Leah Findlater, and Amanda Lazar. 2019. "Phantom Friend" or "Just a Box with Information": Personification and Ontological Categorization of Smart Speaker-Based Voice Assistants by Older Adults. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 214 (Nov. 2019), 21 pages. https://doi.org/10.1145/3359316

[75] Jennifer Pybus and M Coté. 2021. Did you give permission? Datafication in the mobile ecosystem. *Information, Communication & Society* (2021), 1–19.

[76] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. Why Should I Trust You?: Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1135–1144. https://doi.org/10.1145/2939672.2939778

[77] Shruti Sannon, Brett Stoll, Dominic DiFranzo, Malte F. Jung, and Natalya N. Bazarova. 2020. "I Just Shared Your Responses": Extending Communication Privacy Management Theory to Interactions with Conversational Agents. *Proc. ACM Hum.-Comput. Interact.* 4, GROUP, Article 08 (Jan. 2020), 18 pages. https://doi.org/10.1145/3375188

[78] Florian Schaub, Rebecca Balebako, and Lorrie Faith Cranor. 2017. Designing effective privacy notices and controls. *IEEE Internet Computing* (2017).

[79] 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* (Hong Kong, China) *(DIS '18)*. Association for Computing Machinery, New York, NY, USA, 857–868. https://doi.org/10.1145/3196709.3196772

[80] William Seymour, Reuben Binns, Petr Slovak, Max Van Kleek, and Nigel Shadbolt. 2020. Strangers in the Room: Unpacking Perceptions of 'Smartness' and Related Ethical Concerns in the Home. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 841–854. https://doi.org/10.1145/3357236.3395501

[81] William Seymour, Martin J. Kraemer, Reuben Binns, and Max Van Kleek. 2020. Informing the Design of Privacy-Empowering Tools for the Connected Home. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376264

[82] William Seymour and Max Van Kleek. 2021. Exploring Interactions Between Trust, Anthropomorphism, and Relationship Development in Voice Assistants. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 371 (Oct. 2021), 16 pages. https://doi.org/10.1145/3479515

[83] Irina Shklovski, Scott D. Mainwaring, Halla Hrund Skúladóttir, and Höskuldur Borgthorsson. 2014. Leakiness and Creepiness in App Space: Perceptions of Privacy and Mobile App Use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 2347–2356. https://doi.org/10.1145/2556288.2557421

[84] Kevin M. Storer, Tejinder K. Judge, and Stacy M. Branham. 2020. *"All in the Same Boat": Tradeoffs of Voice Assistant Ownership for Mixed-Visual-Ability Families.* Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376225

[85] Jose Such. 2017. Privacy and autonomous systems. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*. 4761–4767.

[86] Takeshi Sugawara, Benjamin Cyr, Sara Rampazzi, Daniel Genkin, and Kevin Fu. 2020. Light commands: laser-based audio injection attacks on voice-controllable systems. In *29th {USENIX} Security Symposium ({USENIX} Security 20)*. 2631–2648.

[87] Max Van Kleek, Reuben Binns, Jun Zhao, Adam Slack, Sauyon Lee, Dean Ottewell, and Nigel Shadbolt. 2018. *X-Ray Refine: Supporting the Exploration and Refinement of Information Exposure Resulting from Smartphone Apps.* Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173967

[88] Wayne F Velicer. 1976. Determining the number of components from the matrix of partial correlations. *Psychometrika* 41, 3 (1976), 321–327.

[89] Sandra Wachter, Brent Mittelstadt, and Chris Russell. 2017. Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harv. JL & Tech.* 31 (2017), 841.

[90] Jordan Wirfs-Brock, Sarah Mennicken, and Jennifer Thom. 2020. *Giving Voice to Silent Data: Designing with Personal Music Listening History.* Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3313831.3376493

[91] Ying Xu and Mark Warschauer. 2020. What Are You Talking To?: Understanding Children's Perceptions of Conversational Agents. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376416

[92] Yaxing Yao, Justin Reed Basdeo, Oriana Rosata Mcdonough, and Yang Wang. 2019. Privacy Perceptions and Designs of Bystanders in Smart Homes. 3, CSCW, Article 59 (Nov. 2019), 24 pages. https://doi.org/10.1145/3359161

[93] 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.

[94] Xiao Zhan, Stefan Sarkadi, Natalia Criado, and Jose Such. 2022. A Model for Governing Information Sharing in Smart Assistants. In *AAAI/ACM Conference on AI, Ethics, and Society (AIES)*.

[95] Guoming Zhang, Chen Yan, Xiaoyu Ji, Tianchen Zhang, Taimin Zhang, and Wenyuan Xu. 2017. *DolphinAttack: Inaudible Voice Commands*. Association for Computing Machinery, New York, NY, USA, 103–117. https://doi.org/10.1145/3133956.3134052

[96] Verena Zimmermann, Paul Gerber, Karola Marky, Leon Böck, and Florian Kirchbuchner. 2019. Assessing users' privacy and security concerns of smart home technologies. *i-com* 18, 3 (2019), 197–216.