The Influence of Misconceptions, Manufacturer Trust, and User Status on Smart Speaker Security Behaviours

Karam Altabbaa (2773821)

Bachelor Thesis

Department of Psychology, University of Twente

Supervisors: Michelle Walterscheid & Nicole Huijts

Abstract

This study explores the influence of misconceptions and manufacturer trust on smart speaker security protective behaviours, comparing users and non-users and the direct effect of user status. While existing literature suggests that correcting user misconceptions can lead to improved security behaviours, findings from this study challenge that assumption. Using an online survey, participants were assessed on their misconceptions, trust, and security behaviours. Results showed no significant relationship between misconceptions and protective behaviours, nor differences in misconception levels between users and non-users. User status emerged as the strongest predictor of behaviour, with users engaging in fewer protective actions than non-users' intended behaviours by asking them to imagine that they were gifted a smart speaker before being asked about their intended behaviour. Additionally, trust was found to be multidimensional and have different effects: general trust negatively predicted user status, while reputational trust positively predicted account-related protective behaviours. These findings suggest that user security practices are shaped more by habitual interaction patterns and psychological framing of trust than by misconceptions.

Introduction

As technology continues to evolve around the internet, there has been a significant increase in the number of smart devices, also referred to as Internet of Things (IoT) devices. IoT refers to a network of devices that contain embedded technology to interact with their environment such as software and sensors allowing them to collect and act on information, and have the ability to communicate with the internet to send and receive information (IBM, 2025). An example of this is a smart home starting the heating of the home once it detects the user is going home, by tracking their phone's location. These devices range from smart household and entertainment devices to transportation systems and complex industrial machines (IBM, 2025). The number of IoT connections is expected to reach 20.1 billion in 2025, up from 18 billion in 2024, and is expected to double within the next eight years (Statista, 2024). Since these IoT devices store and share sensitive user data, there exist several privacy risks, which usually stem from improper user security practices, such as a lack of devices updates, and not changing passwords, in these cases users become vulnerable to cyber-attacks that put their sensitive data at risk. (Tawalbeh et al., 2020).

A common example of such devices are internet protocol (IP) surveillance cameras, which are usually protected with a changeable default password that usually remain unchanged, and are connected to the internet (Kaminsky, 2024). This has become so common that websites showcasing random IP cameras have propped up, allowing you access to a wide variety of IP cameras around the globe that have a default password or no password, some websites only recently changed their policy to exclude cameras that invade someone's personal life (Insecam, 2024).

The above example illustrates a fraction of the importance of regulating and securing IoT devices as they are all potentially vulnerable. Many experts consider the security of IoT devices questionable, since many of the vulnerabilities surrounding these devices are out of the users hand such as cloud infrastructure and improper data encryption (Pourrahmani et al., 2023). However, the simplest and most common reasons are that users are not taking sufficient proactive security measures to protect themselves against cyber-attacks, and likely are not willing to do so. These security measures are as simple as changing weak or default passwords, failure to update their device, or in some cases, insufficient physical security by placing devices in easily accessible areas, leaving them vulnerable to misuse by others (Pawar, 2024). In 2022 the number of IoT cyberattacks amounted to 112 million, almost quadruple the reported 32 million cases in 2018 (Statista, 2024). This heavy spike in incidents indicates an increase in the targeting of IoT devices, which presses the urgency for users to develop security protective behaviours and begin taking security protective measures to reduce the possibility of these attacks.

IoT devices not only collect personal data such as phone numbers, and names, but also can collect and store user activities, such as favourite television programs, daily diet, and general location (Yang et al., 2017). An increasing popular IoT device is smart speakers such as Amazon's Alexa and Google Home. These devices have seen a rapid increase in adoption in recent years, to the point that 20% of respondents in the Netherlands reported owning one in 2021, making them one of the most widely owned household smart devices (CBS, 2022). When it comes to smart speakers specifically, the breadth of possible recorded data is immense. Smart speakers are always listening for a ''wake-word" to activate and interact with users, which then begins a live-audio transmission over the internet, which can be also be intercepted by other parties (Lutz & Newlands, 2021). Furthermore, smart speakers record and store every user interaction even in cases when a similar sounding yet incorrect wake-word is used, such as Alexis for Alexa devices (Schnider et al., 2023), with some studies estimating that up to 10% of all recordings were unintentional (Malkin et al., 2019).

Thus, IoT devices can have severe implications when it comes to user privacy, especially in the context of data breaches. If leaked to the wrong hands, this personal data can lead to identity theft, blackmail (Vallabhaneni, 2024), targeted scams using the individuals voice to target said individual's friends and family through deepfake technology (Visser, 2024), as well as reputational damage, and financial losses for the company (Secarma, 2024). Therefore it is imperative that privacy is maintained.

This study focuses specifically on smart speakers in the home and aims to examine what predicts secure protective behaviour when using these devices, specifically, how misconceptions about smart speaker security and trust in manufacturers might reduce protective behaviours. It also explores whether these relationships differ between users and non-users, given that non-users may be future adopters and may approach security differently.

Misconceptions

Although this technology is being adopted at a rapid rate, many users and non-users hold misconceptions regarding the privacy of smart speakers, such as, 'smart speakers are not always listening' (Meng et al., 2021). Many common misconceptions exist when it comes to smart speakers. In this paper misconceptions will be defined as (Malaterre et al., 2023, p.718) "False ideas that are held that run contrary to what is commonly accepted as knowledge.". These misconceptions range from believing that companies do not store user recordings or store them temporarily (Malkin et al., 2019) to having general misconceptions regarding malicious devices and third-party apps (Zeng et al., 2017). The reasons as to why individuals may not take protective security measures may also lie in the misconceptions they have surrounding IoT devices. One empirical study, with a sample of over 12.000, found that individuals who hold misconceptions about digital security tend to engage in less secure behaviour compared to those without such misconceptions (Herbert et al., 2023). This effect of misconceptions on secure behaviour is also observed in password creation and maintenance (Mayer & Volkamer, 2018), thereby suggesting that correcting these misconceptions may lead to more secure behaviour and in turn, less susceptibility to cyber-attacks.

Furthermore, Lau et al (2018) suggested that in the context of smart speakers, many users have knowledge gaps and misconceptions regarding the privacy risks involved in the use of such devices, which translates into a lack of privacy concerns, and therefore fewer protective measures in securing their devices. On the other hand, another study surrounding smart speaker misconceptions and protective behaviour show no link between the variables (Scheuneman, 2025). Thus, improper security practices may stem from misconceptions regarding IoT devices, their vulnerabilities, and how they operate, record, store, and send data (Lau et al., 2018).

Trust

Another possible reason as to why individuals may not take protective security measures concerning their IoT devices can be attributed to their trust in the manufacturer of said product. The findings of Lau et al (2018) suggest that users have less concerns regarding their security the more they trust the manufacturer of their device, as they believe it is the company's duty to protect their privacy and device security. This is supported by Butavicus et al's (2020) study which found that cybersecurity behaviours are negatively correlated with trust in technological security solutions, as well as other findings suggesting that trust in company security solutions make employees less secure in their protective behaviours (Greulich et al,. 2024). Therefore, the assumption is that trust in the security solutions provided by the manufacturer would be correlated with lower smart speaker protective behaviour. That said, Hapke (2023) found a negative correlation between trust and smart speaker protective behaviour and Pottkamp (2024) found no effect, which could suggest that the relationship may be more nuanced.

User Status

User status, in this study, will be defined as having access to a smart speaker in the home, in common parlance, this can be understood as ownership. User status is likely a key variable in understanding how individuals engage with smart speaker security, offering insight into how direct engagement with the technology may influence protective security behaviours, misconceptions, and trust perceptions. Suh & Han's (2003) findings suggested that manufacturer trust is one of if not the major factor in final user adoption. This relationship between user status and trust is in supported by Lau et al's (2018) study, suggesting that users have greater manufacturer trust, which suggests that manufacturer trust significantly influences smart speaker adoption. As for the effect of user status on misconceptions, it has been observed that users tend to have less misconceptions than non-users, likely due to their familiarity with the devices, however users can and do have many misconceptions regarding smart speakers (Zeng et al., 2017).

However, Scheuenman (2025) found that misconceptions were equally prevalent between users and non-users. This suggests that misconceptions may persist regardless of user status, and warrants further investigation of the effect of user status on misconceptions. Finally, user status may be related to protective behaviours in the context of smart speakers. Multiple studies on the effect of user stauts on protective behaviours in the context of smart speakers suggest that non-users intend to perform more secure behaviours than those performed by users (Scheuneman (2025); Pottkamp (2024); Hapke (2023). This gap may be explained due to the difference between intended behaviour and performed behaviour, as non-users have not interacted with smart speakers and are therefore only hypothetically engaging in said behaviours.

Another explanation of this could be that users face an increase of security demands in their digital lives, potentially leading to security fatigue (Stanton et al., 2016). This phenomenon describes a state of frustration, and disengagement towards security behaviours, stemming from overwhelming requirements for vigilance and the perceived futility of constant effort. This can be especially relevant with always-on devices like smart speakers where the convenience of daily usage may slowly erode adherence to security measures.

Aim of this Paper

This research aims to bridge gaps in understanding the relationship between misconceptions, manufacturer trust, and security protective behaviours and the effect and interaction of user status on these variables. This will be examined through the use of an online survey wherein the participant is identified as a user or non-user, which allows for comparison regarding smart speaker misconceptions, security behaviour, and manufacturer trust, additionally giving insight on factors influencing smart speaker usage. Afterwards, their misconceptions regarding the privacy of smart speakers, security behaviours, and level of manufacturer trust are measured

The research questions that will be answered during this study are:

RQ 1: “Do individuals with more misconceptions perform less secure behaviours actions and how does this differ between users and non-users?”

RQ 2: “What role does manufacturer trust play in security protective behaviours regarding smart speakers?”

RQ 3: “Is there a difference in misconceptions between users and non-users?”

Based on the literature and these research questions the following hypotheses are formulated:

H1: “Misconceptions are negatively correlated with smart speaker security protective behaviours’’.

H2: “Users will perform less secure smart speaker security protective behaviours compared to the intended security protective behaviour of non-users’’.

H3: ‘’Individuals who score higher on manufacturer trust will perform less secure smart speaker protective behaviours”.

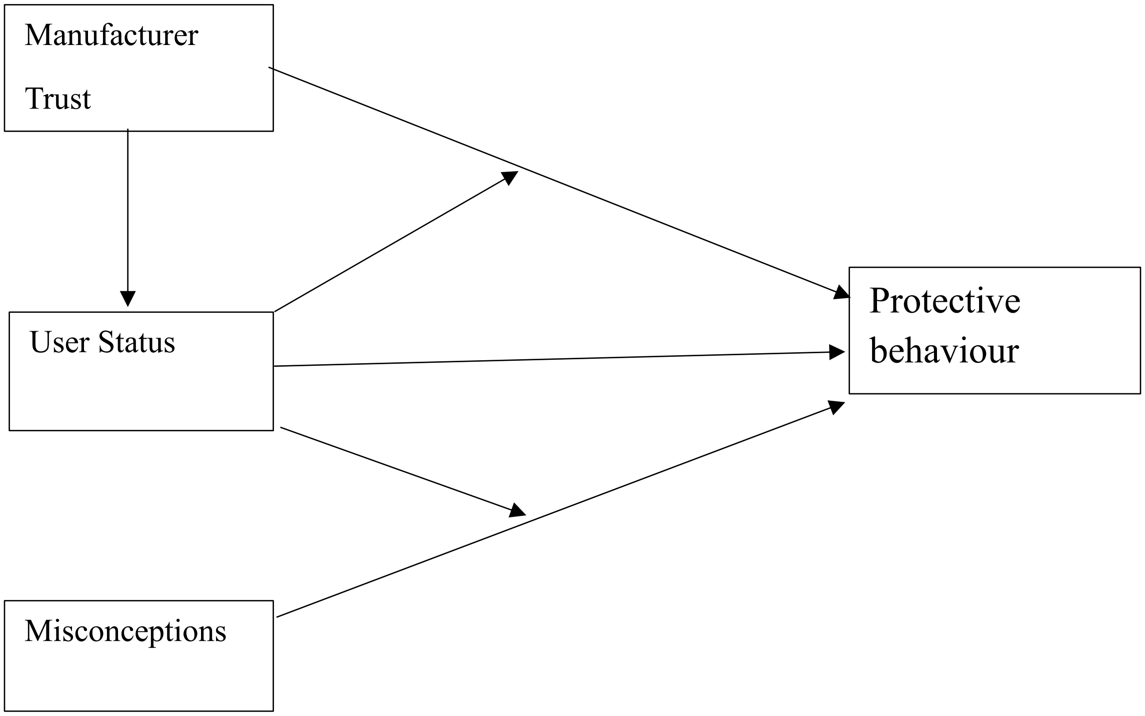
H4: “Individuals who score higher on manufacturer trust will be more likely to use smart speakers”.

H5: “Users of smart speakers have fewer misconceptions than non-users concerning smart speaker security’’.

This study also explores whether user status influences the effects of trust and misconceptions on protective behaviour. See Figure 1 for a representation of all the hypotheses

Figure 1

Proposed model of factors predicting protective behaviour



Methods

Participants

Participants were recruited through a combination of convenience, volunteer, network, and snowball sampling methods. Additionally, participants were recruited through the University of Twente's study recruitment website; SONA, and received 0.25 credits for participating. All participants were required to read the informed consent form in full (see Appendix A), upon agreeing they were shown the survey. Additionally, ethical approval was obtained from the Humanities & Social Sciences Ethical Committee of the University of Twente. The survey received 154 responses, of which only 101 remained due to failed attention checks (N= 30), not reading or agreeing to the informed consent form (N = 15), or not completing the survey (N = 8). Of the final 101 participants, 66 participants were male, 33 were female, one preferred not to say, and one participant identified themselves as non-binary. Participant age ranged from 18 to 62, M = 28.1 , SD = 12.2 . The sample was overall international, with the majority of participants being from Jordan (20), followed by the Netherlands (14) and Turkey (13), with 10 German participants. Below is a table with distribution of participant nationality, countries with three or less respondents are grouped by region, for the full table of participant nationalities, refer to Appendix I

Table 1

Participant nationality distribution

Nationality/Region	Frequency
Jordanian	20
Dutch	14
Turkish	13
German	10
Spain	7
Other	
European	20
North African	3
North American	2
East Asian	4
Middle Eastern	7
South Asian	1

Of the 101 participants, 28 were smart speaker users. Participants also ranged in education level, with 43 completing high school, two have completed vocational school, 43 holding a bachelor's degree, 12 holding a master's degree, and one holding a PHD.

Procedure

A cross-sectional, correlational survey design was used to examine the relationships between variables. The questionnaire was created and hosted on Qualtrics (see Appendices C–

H), and participants completed the survey via a Qualtrics link. First, participants were given information on the structure and purpose of this study, the right to withdraw, and information pertaining to the handling of their data such as, anonymity and confidentiality. They were then required to read and agree to the informed consent form of the study to proceed. Next, participants were asked to fill in their demographics (e.g. age, gender and level of education), followed by a question regarding smart speaker user status, and usage (if applicable). Based on smart speaker user status, participants received a smart speaker security protective behaviour scale for users or non-users. Additionally, non-users were asked to imagine that they received a smart speaker as a gift and to indicate the likelihood of engaging in the given behaviours. The questions were also reformulated for non-users, measuring their intended behaviour while using a smart speaker, as opposed to measuring users' smart speaker security behaviour. Next, participants completed the adapted manufacturer trust scale, and finally the smart speaker misconceptions scale. Additionally, all the questions were presented in the same sequence for both groups, except for the misconceptions scale, which presented four blocks in a random order for each participant. The question order within the blocks was also randomized. At the end of the questionnaire, participants were thanked for their participation. It took the participants approximately 13.5 minutes to complete the questionnaire.

Materials

The scales measuring trust, and protective behaviours were adapted from existing research and the misconception scale was provided by the research supervisor. All the scales were designed in the context of smart speakers. In addition, the survey had some control items regarding the user status and participants' demographic data, such as gender, education, age, and nationality.

Manufacturer Trust

To test for manufacturer trust, Hapke's (2023) scale was used, specifically, the scale measuring ''Trust in smart speaker companies''. It is measured with seven items on a 5--point Likert scale with options ranging from ''Strongly Disagree'' to ''Strongly Agree''. All the questions are also positively phrased, such as "Smart speaker companies care about protecting my data to maintain their positive brand image.". Therefore, a high score suggests high trust in smart speaker manufacturers. The scale demonstrated good internal consistency, with Cronbach's $\alpha = .86$. The mean score was $M = 2.49$, with a standard deviation of $SD = 0.72$, which is below the midpoint, indicating low baseline trust levels towards manufacturers (see Appendix G for the full scale).

An exploratory factor analysis (EFA) was conducted on the seven items measuring manufacturer trust, using maximum likelihood extraction with varimax rotation. The KMO value was 0.76, indicating adequate sampling adequacy. All individual items also showed acceptable KMO values, ranging from 0.63 to 0.88. Bartlett's Test of Sphericity was significant ($\chi^2(21) = p < .001$), and was therefore suitable for factor analysis (Bartlett, 1954). Parallel analysis was done and a scree plot was generated which suggested that manufacturer trust items measured two constructs, this was further confirmed based on the number of factors with an eigenvalue above one. The analysis revealed that all seven items loaded moderately to strongly on either factor, with standardized loadings ranging from 0.37 to .99. The exploratory factor analysis yielded two distinct factors, which were subsequently labelled General Trust and Reputational Trust based on their item content.

The General Trust factor encompassed items that reflected participants' broad confidence and trust in the smart speaker manufacturers, represented by items such as 'Smart speaker companies are trustworthy in handling the data the smart speaker collects about me.'. In contrast, the Reputational Trust factor consisted of items that emphasized manufacturers' motivations tied to their public image and reputation, such as 'Smart speaker companies care about protecting my data to maintain their positive brand image. Item 5 loaded on both factors (General trust = .37, Reputational trust = .66) , however it loaded highly on only one factor, and thus was assigned to the factor reputational trust. Communalities ranged from 0.29 to 1.00, suggesting low to high shared variance between items and the latent factor. The first factor accounted for 32% of the total variance and the second factor accounted for another 21%, bringing the total variance explained to 54%. Factor loadings can be found in the table below.

Table 2

Factor Loadings Manufacturer Trust

Item	General Trust	Reputational Trust
Smart speaker companies are trustworthy in handling the data the smart speaker collects about me.	.70	
I trust that smart speaker companies keep my best interests in mind when dealing with the information collected about me by the smart speaker.	.72	
Smart speaker companies are in general predictable and consistent regarding the usage of the information collected about me.	.54	
Smart speaker companies are careful with sharing my personal data with third parties.	.60	
Smart speaker companies are always honest with customers when it comes to using the information that they provide.	.54	
Smart speaker companies intend to protect my data well because they want to keep their market shares.	.37	.66
Smart speaker companies tare about protecting my data to maintain their positive brand image.		.99

The general trust (M= 2.26, SD = 0.76) and reputational trust factors (M= 3.04, SD = 1.0) both showed good internal consistency with Cronbach's α = .86 and .82 respectively.

Smart Speaker Security Protective Behaviours

Smart speaker security behaviour was measured with Pottkamp's (2024) questionnaire using the scale measuring 'Protective Behaviour'. It is measured with a 5 point Likert scale with options ranging from ''Never'' to ''Always'' for users, and ''Extremely Unlikely'' to ''Extremely Likely'' for non-users. All the questions are positively phrased, with a higher score suggesting more secure protective behaviours, with questions like; ''I unplug the smart speaker when I am not using it''. The question tenses were reformulated from past tense for users to future tense for non-users. The previously used example in the case of non-users would be; "I would unplug the smart speaker when I am not using it''. Since protective behaviour was rephrased for users and non-users, factor analysis should be done on both versions of the scale. However, the small sample size for users (N = 28) and non-users (N=73) does not meet the generally accepted threshold for conducting ''good'' factor analysis, which is a minimum of 100 (Kline, 2014). Therefore, exploratory factor analysis will be done on both groups together, and subsequent exploratory factor analyses will be conducted on non-users and users separately. The 21 item responses from the user and non-user scales were merged into a unified 21-item scale reflecting parallel constructs.

The overall Kaiser-Meyer-Olkin measure was excellent (KMO = .87), and Bartlett's test of sphericity was significant, $\chi^2(210) = p < .001$, providing support for its factorability. A parallel analysis suggested a three-factor solution, which was also supported based on the eigenvalue criteria. The three factors were extracted using maximum likelihood estimation with varimax rotation with a cutoff of 0.3.

The model explained 52% of the total variance across three factors. Only the item “I did/will not write down the password on a piece of paper or share it with house members or visitors” did not load on any of the factors, it was then removed and factor analysis was repeated. The KMO measure remained high at 0.87, and Bartlett’s test was significant, $\chi^2(190) =$ $p < .001$, confirming the data's factorability. Parallel analysis again supported a three-factor solution. The new model showed improved factor structure with cleaner loading patterns and maintained high factor score adequacy (regression score correlations ≥ .96). The cumulative variance explained was 54%. items showed many primary loadings, ranging from low to high, as well as two cross-loadings and can be found in Table 3.

Table 3

The items measuring protective behaviours of the combined groups and their factor loadings for a two-factor model.

Item	Physical Security Behaviours	Account Management	Password management
I will turn off/turned off the smart speaker when I am/was not using it	0.68		
I will unplug/unplugged the smart speaker when I am/was not using it	0.53		
I will unplug/unplugged the smart speaker when I am/was having serious/private conversations	0.82		
I will turn/turned off the smart speaker when I am/was having serious/private conversations	0.84		
I will mute/muted the smart speaker's microphone when I am/was not using it	0.73		
I will review/reviewed the privacy settings of my smart speaker in the provider's (e.g., Alexa or Google) account		0.78	
I will review/reviewed which applications/services have access to my smart speaker		0.97	
I (will) regularly spend time reviewing audio logs in the mobile app and delete those that I want to be deleted.		0.45	
I will restrict/restricted the amount of data that the device is allowed to collect through the smart speaker's settings		0.59	
I will delete/deleted my smart speaker recordings	0.37	0.41	
In the app, I will delete/deleted sensitive information that the smart speaker stored about me	0.33	0.44	
I will speak/spoke very quietly around the smart speaker, in case I don't/didn't want to be recorded	0.37		
I will moderate/moderated my language around the smart speaker so that it doesn't/didn't record private matters, even accidentally	0.45		
I will avoid/avoided sensitive/private conversations around the smart speaker	0.65		
When/if I have/had a visitor, I will inform/informed them that I have a smart speaker		0.94	
When/if I have/had a visitor, I will offer/offered to switch the smart speaker off		0.74	
I will place/placed the smart speaker so that it is isn't/wasn't positioned in areas where I (would) typically engage(d) in sensitive or private conversations.	0.52		
I (will) set a new and difficult password that I only use for my smart speaker.			0.32
I will change/changed the password again after some time			0.49
I will/did not place the smart speaker in a privacy sensitive room like my bedroom		0.43	

The items extracted from the protective behaviour scale loaded low to high on the extracted factors, and were relatively easy to group, although there were cross loadings. The two items that cross loaded were theoretically related to the second factor, which was also the factor they loaded higher on, and were subsequently assigned to that factor. The final item was not theoretically related to the factor, but was left in due to its loading. The first factor identified is physical security behaviours with a total of nine items (M = 2.56, SD = 1.08, α = .91) and revolved around actions that the individual would physically take in order to ensure their security, such as “I will mute/muted the smart speaker’s microphone when I am/was not using it”. The second factor was account management also with nine items (M = 2.94, SD = 1.19, α = .85) and revolved around actions individuals can perform on their accounts to increase their smart speaker security. This factor had questions such as “I will review/reviewed which applications/services have access to my smart speaker”. The third factor was identified as password management and had two items (M =2.25, SD = 1.11, α = .78), and as the name suggests revolved around proper password management, with questions such as “I (will) set a new and difficult password that I only use for my smart speaker.”. The separate factor analyses for the protective behaviour scale for users and non-users can be found in Appendix E). It seems that the factors of protective behaviours may be grouped differently for users and non-users, with non-users having three factors identified within the scale, and users only having two. Especially when considering that some items that loaded in the non-user and overall factor analysis, did not load for the user factor analysis. This may be due to the significant difference in sample size and the difference of measuring intended behaviour versus actual behaviour.

Smart Speaker Misconceptions

To test the participants’ misconceptions a 34 item scale was used that was provided by the research supervisor, this scale was assessed by four expert reviewers. Each question on the scale has 2 parts, the first part being a true or false question to test the participants knowledge. The second part being a 6 point Likert scale to measure the participants confidence in their answer, the choices range from ‘’Just Guessing” to ‘’Absolutely Sure” with no neutral option. Internal consistency for the misconception scale was adequate α = .77.. Exploratory factor analysis was done on the misconceptions scale, however, the factors extracted will not be used in analysis and can be found in Appendix K, respectively. Finally, further exploratory analyses to improve the scale's internal consistency did not yield substantial improvements; removing any single item resulted in a maximum Cronbach's alpha change of only .01 (e.g., from .77 to .78).

Data Analysis

The analyses were done using R-Studio (Version 2024.12.1). Initially, the data was cleaned, removing all incomplete responses and all responses with failed attention checks. Participant demographics were then analysed through descriptive statistics and frequencies. Afterwards, each one of the scales were analysed for internal consistency and reliability by using reliability analyses (Cronbach’s Alpha), descriptive statistics, and factor analysis. To determine the suitability of the items for factor analysis, both the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy and Bartlett’s Test of Sphericity were conducted.

Prior to analysing the data, assumptions of Linearity, Normality Multicollinearity, Independence, and Homoscedasticity were checked and were met. All continuous predictor variables were mean-centered prior to analysis to reduce multicollinearity and improve the interpretability of main effects. A Pearson's correlation coefficient and two linear regression models (with and without interaction effects between user status and the dependent variables) were used to examine the relationships between misconceptions and protective behaviours, and between manufacturer trust and protective behaviours. Independent samples t-tests were used to determine whether there were significant differences in misconception levels and security behaviours between users and non-users. Binary logistic regression was used, treating smart speaker user status (user vs. non-user) as the dependent variable and manufacturer trust as the predictor to assess the likelihood of owning when considering manufacturer trust.

For the regression model, the gender variable was recoded to ensure consistent numerical treatment in regression analysis. Participants who identified as non-binary (coded as 3) or preferred not to disclose their gender (coded as 4) were both recoded as 1.5, resulting in a continuous variable where 1 = male, 2 = female, and 1.5 represented all other gender responses.

Results

Table 4 summarizes the results of the Pearson correlation and will be used to support H1-H3. Table 5 summarizes the regression coefficients for the model without interaction effects across all three protective behaviour factors, and will be used to further test H1-H3. Table 6 summarizes the model with interaction effects.

Preliminary Analysis

Table 4

Pearson's Correlation between Misconceptions and Protective Behaviour with the Dependent Variables

	Misconceptions		Physical security Behaviours		Account Management		Password management	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Misconceptions			.18	.080	.16	.116	.10	.332
User Status	-.01	.931	**-.46**	**<.001**	**-.47**	**<.001**	**-.31**	**.002**
General Trust	**.21**	**.036**	.14	.172	.14	.174	.03	.788
Reputational trust	.10	.301	.10	.343	.17	.085	.07	.518
Physical security Behaviours	.18	.080			**.72**	**<.001**	**.58**	**<.001**
Account Management	.16	.116	**.72**	**<.001**			**.55**	**<.001**
Password management	.10	.332	**.58**	**<.001**	**.55**	**<.001**		

Note. All significant correlations are marked in bold.

The three protective behaviour factors were significantly inter-correlated, providing support for their conceptual relatedness while remaining distinct factors. Notably, in line with H2, user status was significantly and negatively correlated with all three protective behaviour factors; Physical Security Behaviours: (r = -.46, *p* < .001); Account Management: (r = -.47, *p* < .001); Password Management: (r = -.31, *p* = .002). These findings suggest that smart speaker owners, on average, engage in fewer protective behaviours across all three dimensions compared to non-owners. Furthermore, misconceptions showed a significant positive correlation with General Trust (r = .21, *p* = .036), indicating that individuals reporting higher levels of general trust in manufacturers also reported more misconceptions regarding smart speaker security. All other bivariate correlations were not statistically significant and do not provide support for any other hypotheses.

Regression Model

The direct effects models (Table 5) were statistically significant for Physical Security Behaviours F(6, 94) = 5.53, p < .001, R^2= .261, Account Management F(6, 94) = 6.17, p < .001, R^2 = .283, and Password Management F(6, 94) = 1.93, p = .039, R^2 = .143. This model will test H1 – H3.

Table 5

Regression Model Without Interaction Effects

	Physical Security Behaviours				Account Management				Password Management			
	B	SE	t	*p*	B	SE	t	*p*	B	SE	t	*p*
Misconceptions	0.04	.02	1.71	.091	0.40	.02	1.66	.999	0.02	.02	0.96	.338
User Status	**-1.13**	.22	-5.02	**<0.01**	**-1.32**	.24	-5.35	**<0.01**	**-0.81**	.25	-3.24	**.002**
General Trust	-0.02	.15	-0.11	.915	-0.09	.16	-0.60	.548	-0.13	.16	-0.79	.432
Reputational trust	0.11	.11	0.98	.331	**0.24**	.12	2.01	**.047**	0.12	.12	0.97	.333
Age	0.01	.01	1.29	.197	0.01	.01	0.72	.474	0.01	.01	1.21	.231
Gender	0.13	.22	0.62	.535	-0.08	.23	-0.34	.732	0.13	.24	0.56	.574

Note. All significant effects are marked in bold. Model Significance: Physical Security Behaviours F(6, 94) = 5.53, p < .001, R^2= .261 , Account Management F(6, 94) = 6.17, p < .001, R^2 = .283, Password Management F(6, 94) = 1.93, p = .039, R^2 = .143.

Hypothesis testing

H1 proposed that misconceptions were negatively correlated with smart speaker security protective behaviours. Consistent with the results of the correlation, multiple linear regression analyses (see Table 5) did not show misconceptions as a significant direct predictor for any factor of protective behaviour. Therefore, H1 was rejected across all three protective behaviour factors.

H2 assessed whether users perform less secure smart speaker security protective behaviours compared to the intended security protective behaviour of non-users". In the direct effects models, user status emerged as a significant negative predictor for all factors of protective behaviour. Therefore, H2 was accepted, with user status being a significant negative predictor.

H3 assessed if individuals who score higher on manufacturer trust will perform less secure smart speaker protective behaviours. In the direct effects model (see Table 5), general trust was not a significant predictor for any of the three protective behaviour factors in the direct effects model. Reputational trust was a significant positive predictor exclusively for Account Management in the direct effects model. Therefore, H3 was partially accepted, with reputational trust having a direct positive relationship with Account Management.

H4 assessed whether trust in smart speaker manufacturers positively predicted device user status. To test this, a binary logistic regression was conducted with both trust factors. The model revealed that general trust significantly predicted user status (B = -0.74, SE = 0.37, z = -1.98, *p* = .048). For every one-unit increase in `general trust`, the odds of being a smart speaker owner decreased by 51.7% (Odds Ratio = **0.48**, 95% CI = [0.22, 0.96]). This indicates that individuals with higher general trust were less likely to be smart speaker owners, which was contrary to the expected positive relationship. `Reputational trust` was not a significant predictor

(B = 0.30, SE = 0.26, z = 1.13, *p* = .260). Thus, H4 was partially accepted, being a significant negative predictor for user status.

H5 proposed that smart speaker users would have fewer misconceptions than non-users. A Welch's independent samples t-test was conducted to compare misconceptions scores between owners and non-owners. A Welch’s t-test was selected over a standard t-test due to the unequal sample sizes between the two groups. The results indicated no significant difference in misconceptions scores between non-owners (M = 9.09, SD = 4.87) and owners ((M = 9.00, SD = 4.37), (t(48.18) = 0.09, *p* = 0.928), 95% CI = [1.96, 2.15]. This suggests that user status is not associated with misconceptions, thus, H5 was rejected.

Table 6

Regression Model with Interaction Effects

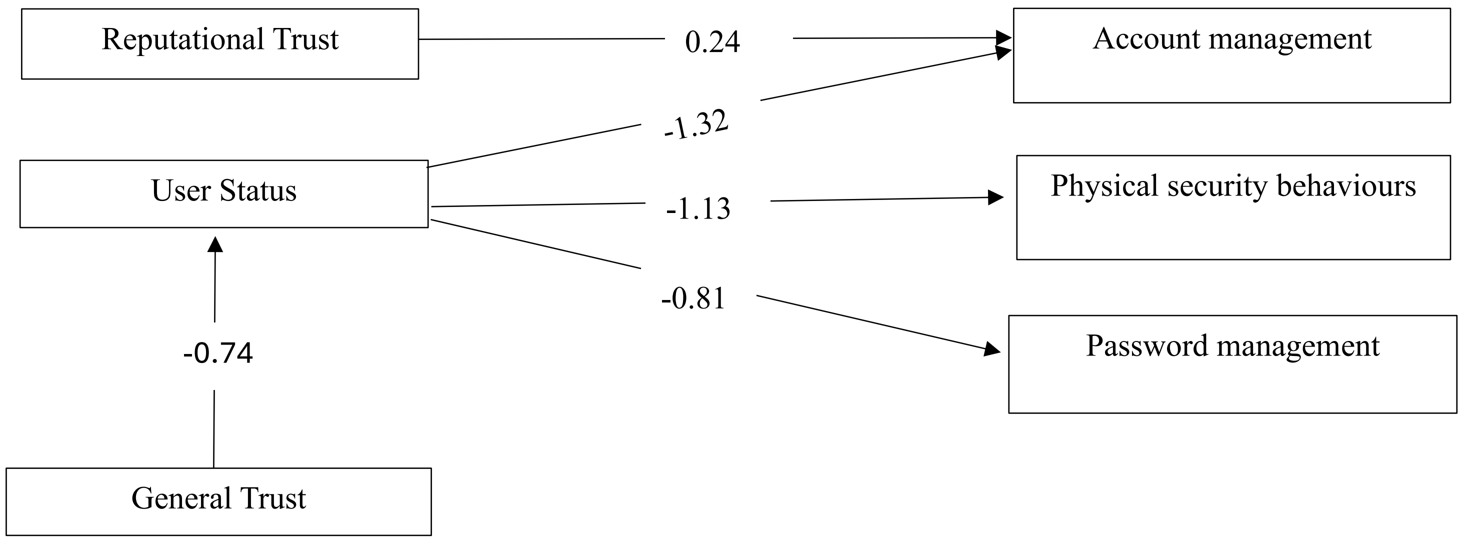
	Physical Security Behaviours				Account Management				Password Management			
	B	SE	t	p	B	SE	t	p	B	SE	t	p
Misconceptions	0.04	.02	1.74	.085	0.04	.02	1.71	.090	0.02	.02	0.95	.342
User Status	**-1.15**	.23	-4.65	**<.001**	**-1.40**	.25	-5.37	**<.001**	**-0.83**	.26	-3.19	**.002**
General Trust	-0.01	.15	-0.06	.949	-0.10	.16	-0.65	.519	-0.14	.17	-0.86	.394
Reputational trust	0.09	.11	0.87	.386	**0.24**	.12	2.07	**.041**	0.11	.12	0.98	.332
Age	0.01	.01	1.51	.136	0.01	.01	0.67	.508	0.01	.01	1.19	.239
Gender	0.18	.22	0.81	.418	-0.08	.24	-0.32	.747	0.13	.25	0.51	.613
Misconceptions * User Status	0.04	.05	0.72	.476	0.03	.06	0.53	.602	0.00	.06	0.04	.971
General Trust * User Statis	0.33	.36	0.91	.363	-0.39	.39	-1.01	.315	-0.09	.40	-0.22	.829
Reputational Trust * User Status	0.07	.25	0.29	.776	-0.09	.28	-0.33	.741	0.32	.28	1.12	.266

Note. All significant effects are marked in bold. Model significance: Physical Security Behaviours F(9, 91) = 3.97, *p* <.001, R^2 = .282), Account Management F(11, 89) = 4.30, *p* < .001, R^2 = .299), Password Management F(11, 89) = 1.68, *p* = .11 , R^2 = .142).

The model below displays the significant variables when it comes to predicting protective behaviour.

Figure 2

Full model of significant predictors with effect sizes



Exploratory Analyses

Interaction Effects Regression Model

The interaction effects of trust and misconceptions with user status were also explored, The full models with interaction terms (Table 6) were statistically significant for Physical Security Behaviours F(9, 91) = 3.97, p <.001, R2 = .282), Account Management F(11, 89) = 4.30, p < .001, R2 = .299). The full model for Password Management F(11, 89) = 1.68, p = .11 , R2 = .142) was not statistically significant. User status remained a consistent negative predictor of protective behaviour, and Reputational Trust was positively correlated with Account Management. No interaction effects were significant.

Discussion

This study aimed to investigate how misconceptions, trust in manufacturers and user status influence smart speaker security protective behaviours, and the possible moderating effect of user status on the effect of misconceptions and trust on the protective behaviours. Users actual behaviour frequency was measured, whereas non-users were asked to imagine that they received a smart speaker as a gift and indicate the likelihood of them engaging with the given behaviours, thereby measuring intended behaviour. Contrary to H1, misconceptions about smart speakers did not predict security protective behaviours. This finding challenges the assumption that misconceptions translate into less secure behaviours (Herbert et al., 2023). However, a difference emerged in H2, with non-users reporting significantly more protective intentions than users' actual behaviours, which was also observed in Hapke's (2023), Scheuneman's (2025), and Pottkamp's studies (2024) The result of H3 was different from what was expected, it was observed that reputational trust positively predicted protective behaviours, meaning individuals with higher reputational trust performed more secure protective behaviours.

The result of H4 was also an unexpected finding, trust did show a relationship with smart speaker user status but in the opposite direction of what was expected. It was observed that general trust actually negatively predicts user status of smart speakers. This suggests that, individuals with lower general trust are more likely to be smart speaker owners. Additionally, contrary to H5, smart speaker users and non-users showed no significant differences in the average number of misconceptions, suggesting that user status does not influence misconceptions, which is surprising given that owners are more acquainted with the technology.

On the model as a whole, It was observed that only user status is a consistent and significant negative predictor across all factors of the protective behaviour scale. Additionally, reputational trust positively predicted the account management factor of the protective behaviour scale.

Misconceptions and Protective Behaviour

The lack of association between misconceptions and protective behaviours is inconsistent with previous literature and theoretical expectations. Research by Herbert et al. (2023) suggested that misconceptions lead to less secure behaviour online, and other research indicated that it lead to less secure password protection (Mayer & Volkamer, 2018). The findings suggest this relationship may be different in the context of smart speaker protective behaviour specifically. There are several possible reasons as to why this discrepancy exists; one such reason is that the relationship between misconceptions and behaviour in security contexts may be moderated or mediated by other factors not measured in this study, and beyond the scope of this study. This aligns with broader theoretical frameworks such as the theory of planned behaviour, which suggests that knowledge alone is insufficient for behaviour change without accompanying behavioural skills, attitudes, and motivational factors (Etheridge et al., 2023).

Trust and User Status

The negative relationship between general trust and user status challenges the findings of Suh and Han (2003), who suggested that manufacturer trust is one of the major factors positively influencing final user adoption. The results also contradict Lau et al.'s (2018) findings that users have greater manufacturer trust, which suggested that manufacturer trust significantly influences smart speaker adoption. One possible explanation for this contradiction is that, over time, users may have been exposed to negative media coverage about security issues and their smart speaker manufacturer, which gradually eroded their trust in the brand (Di Domenico & Ding, 2023).

The identification of two trust dimensions, general trust and reputational trust, broadens the understanding of trust in the context of smart speakers. Interestingly, reputational trust was found to be a significant positive predictor of Account Management, indicating that individuals with higher reputational trust reported more secure account management behaviours. This finding has not been observed in current literature on the topic, to my knowledge. However, one explanation for this could be that reputational trust may also enhance protective behaviour because it reflects a more sceptical, calculated belief in corporate motivations. This mindset can foster a sense of conditional vigilance: users may think, 'I believe they'll protect my data as long as it benefits them—but I still need to check.' Consequently, these individuals may be more inclined to take relatively simple but effective security protective behaviours. General trust did not predict any protective behaviour factor, and no interactions were significant.

User versus Non-User Behaviour Patterns

The finding that non-users report higher protective intentions than users' actual behaviours raises important questions about the intention-behaviour gap in security contexts. This gap may reflect the difference between hypothetical decision-making and real-world behaviour when it comes to actual usage, with factors such as intention strength influencing the final outcome (Conner & Norman, 2022) . Specifically, for smart speaker users, the continuous engagement with these devices may lead to security fatigue (Stanton et al., 2016). As users become more familiar with smart speakers, the repeated demands for vigilance, privacy settings, and ongoing security protocols can become overwhelming, fostering disengagement from protective behaviours. This suggests that the discrepancy observed between users and non-users might not only be an intention-behaviour gap but also a consequence of security fatigue.

Comparison with Previous Findings

The current study's findings can be compared directly with both Hapke (2023) and Pottkamp's (2024) research, as all studies used the same scale for measuring protective behaviour and manufacturer trust. Across these studies, interesting patterns emerge regarding trust and protective behaviours. Hapke (2023) found a significant relationship between manufacturer trust and protective behaviours, with trust predicting less secure protective behaviour. On the other hand, this study found a significant relationship of reputational trust positively predicting protective behaviours, the opposite of Hapke’s (2023) finding. Pottkamp's (2024) study found trust significantly correlated with privacy risk perception but not with protective behaviour overall.

Cultural differences may explain these variations. This research’s diverse, mostly non-European sample juxtaposed against Hapke’s (2023) mostly German (80% German) and Pottkamp’s (2024) primarily German/European participants suggest that cultural differences may significantly influence how trust is understood and how it translates into behaviour.

A notable distinction also emerged in how trust was conceptualized. While all three studies used the same scale, both Pottkamp (2024) and Hapke (2023) found that the manufacturer trust scale was unidimensional, however, factor analysis on the same scale in this study revealed two distinct dimensions: general trust and reputational trust. This may be due to the difference in samples, comparing an international sample to a European, predominantly German sample, as well as the differences in gender, age, and education level. It is possible that the conceptualization of manufacturer trust differs across cultures, perhaps influenced by varying experiences with data privacy regulations or the perceived level of government surveillance in their respective countries, which could shape how individuals perceive and differentiate aspects of trust in manufacturers.

Although there are many differences between the studies, several key findings show consistency across the studies. The protective behaviour scale needs further refinement, as evidenced by all three studies independently finding factor structures different from the original. Cultural and sample composition consistently emerge as important considerations, highlighting the need for cross-cultural validation of findings. It is also worthy to mention that all studies had more non-users than users. Finally, user status consistently emerges as a crucial predictor of protective behaviours across all three studies, suggesting that users and non-users have fundamental differences in their behaviours. This may also reflect measurement inconsistencies, as protective behaviour was assessed as actual actions for users and as intended behaviour for non-users, potentially capturing different psychological processes. As a result, comparisons between groups may not reflect true behavioural differences, but rather differences in how behaviour is reported.

Strengths and limitations

Several strengths and limitations should be considered when interpreting these findings. One strength of this study is that misconceptions have not been studied in depth when it comes to smart speaker security behaviours. This study may serve as a stepping stone for future research into the effect of misconceptions on smart speaker security protective behaviours; and particularly on the relationship between trust, user status, and misconceptions.

Another strength is that the sample for this study was international, especially when compared to the Eurocentric and mainly German samples of previous studies. Thus the findings can be applied on diverse populations.

The most impactful limitation for this study was the sample size for smart speaker users (n = 28). This may have limited the study's predictive power to detect effects in user-specific analyses and when measuring group differences. This may also explain the differences in findings between this study and previous studies. The small sample size, combined with the disproportionate ratio of users to non-users, may have affected the interpretability of findings related to user status. Future research with larger user samples would strengthen confidence in these findings.

Additionally, the study employed a cross-sectional design, which limits causal inferences about the relationships observed. Longitudinal research following individuals as they adopt and use smart speakers could provide valuable insights into how protective behaviours evolve over time and what different factors influence security protective behaviours. Future longitudinal studies should be done on how an individual's intended behaviour and actual behaviour differ, before and after owning a smart speaker.

Lastly, the vast majority of respondents completed the survey in English, which is not their primary language. This fact may have affected response quality, possibly increasing the number of random selections when the question is vague or difficult to understand.

Future recommendations

These points culminate in several recommendations for future research. First, future studies should investigate variables that may moderate the relationship between knowledge or misconceptions and security protective behaviours. Prior research suggests that knowledge alone is often insufficient to drive behavioural change, particularly in digital security contexts (Brough & Martin, 2019). One such mediating factor is self-efficacy, or the belief in one's own ability to perform secure behaviours. Higher self-efficacy has been consistently linked to increased engagement in protective behaviours across various domains, including cybersecurity (Wald et al., 2024; Etheridge et al., 2023). Additionally, perceived barriers such as, time constraints, perceived complexity, or a lack of perceived effectiveness, have been shown to hinder security behaviour adoption, regardless of knowledge levels (De Kimpe et al., 2021). Users may also experience resignation when they feel overwhelmed by security demands, further weakening the knowledge–behaviour link (Haney, 2023; Stanton et al., 2016). Finally, practical constraints, such as poor device interface design or the absence of clear guidance from manufacturers, may prevent even well-informed users from acting securely (Lutz & Newlands, 2021). Thus, to fully understand why knowledge does not always translate into secure behaviour, future work should move beyond cognitive factors and examine motivational, contextual, and behavioural determinants in an integrated model.

Another recommendation would be development of more robust misconception scales regarding smart speakers, building on the lessons learned from the current study's psychometric challenges. Lastly, similar studies should be conducted across different cultural contexts, particularly comparing European and non-European samples given the differences observed in this study compared to previous studies.

Product Design and Policy

Since the variables investigated do not appear to directly influence protective behaviours, companies might focus on making security measures more default and automatic rather than relying on users to implement them voluntarily, in that way, any influencing variables would not affect the quality of security for the user. However, implementing these security measures may be costly for the companies and would likely not be done voluntarily. A solution to this may be to promote changes in legislation, that would require these security measures to be applied. This implies however, that users should not trust manufacturers to take protective measures on their behalf, which is problematic for the companies in question. Although, the finding that reputational trust influences protective behaviours may indicate that trust building with the consumer may be an effective solution to increasing protective behaviours. Perhaps more open communication between the user and the provider regarding these safety risks would increase reputational trust and promote more secure protective behaviours.

The significant difference between users' and non-users' security behaviours suggests that once individuals adopt smart speakers their security behaviours may decrease. This may be due to a sense of resignation to cyber security demands (Haney, 2023), security fatigue (Stanton et al., 2016), or perhaps that individuals' intended security behaviour does not translate to reality (Norberg et al., 2007). This indicates the need for design solutions that maintain security standards regardless of user behaviour, such as automatic security updates.

Given that misconceptions do not predict protective behaviours, interventions targeting smart speaker security should move beyond correcting misconceptions. Instead, interventions might focus on behavioural approaches that make protective actions more convenient and automatic, in order to address the practical barriers to taking protective actions.

Conclusion

This study explores how misconceptions, and manufacturer trust influence security protective behaviours in the context of smart speakers, as well as the effect of user status when it comes to influencing said behaviours. This study contributes to the overall understanding of smart speaker security by challenging several widely held assumptions about the relationships between misconceptions, trust, and security behaviours. Mainly, that misconceptions do not correlate with protective behaviours, as well as the finding that trust positively influences a factor of protective behaviours in that users with greater trust perform more protective security behaviours. The findings suggest that smart speaker security behaviour may be shaped by a narrower set of psychological factors than previously assumed, with user status emerging as the most consistent factor. While this simplifies the model, it also implies fewer clear intervention points for improving user security. As smart speakers continue to proliferate and integrate more deeply into our homes and daily lives, understanding these security dynamics becomes increasingly critical.

References

Brough, A. R., & Martin, K. D. (2019). Critical roles of knowledge and motivation in privacy research. *Current Opinion in Psychology*, *31*, 11–15. https://doi.org/10.1016/j.copsyc.2019.06.021

Butavicius, M., Parsons, K., Lillie, M., McCormac, A., Pattinson, M., & Calic, D. (2020). When believing in technology leads to poor cyber security: Development of a trust in technical controls scale. *Computers & Security*, *98*, 102020. https://doi.org/10.1016/j.cose.2020.102020

CBS. (2022, January 17). *How many people use the Internet of Things? - The Netherlands in numbers 2021*. How Many People Use the Internet of Things? - the Netherlands in Numbers 2021 | CBS. https://longreads.cbs.nl/the-netherlands-in-numbers-2021/how-many-people-use-the-internet-of-things/

Conner, M., & Norman, P. (2022). Understanding the intention-behavior gap: The role of intention strength. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.923464

De Kimpe, L., Walrave, M., Verdegem, P., & Ponnet, K. (2021). What we think we know about cybersecurity: an investigation of the relationship between perceived knowledge, internet trust, and protection motivation in a cybercrime context. *Behaviour and Information Technology*, *41*(8), 1796–1808. https://doi.org/10.1080/0144929x.2021.1905066

Di Domenico, G., & Ding, Y. (2023). Between brand attacks and broader narratives: How direct and indirect misinformation erode consumer trust. *Current Opinion in Psychology*, *54*, 101716. https://doi.org/10.1016/j.copsyc.2023.101716

Etheridge, J. C., Sinyard, R. D., & Brindle, M. E. (2023). Implementation research. In *Elsevier eBooks* (pp. 563–573). https://doi.org/10.1016/b978-0-323-90300-4.00043-4

Greulich, M., Lins, S., Pienta, D., Thatcher, J. B., & Sunyaev, A. (2024). Exploring contrasting effects of trust in organizational security practices and protective structures on employees' Security-Related Precaution taking. *Information Systems Research*, *35*(4), 1586–1608. https://doi.org/10.1287/isre.2021.0528

Haney, J. (2023, March). *Users are not stupid: Six cyber security pitfalls overturned*. NIST. https://www.nist.gov/publications/users-are-not-stupid-six-cyber-security-pitfalls-overturned

Hapke, J. (2023). *Examining Factors that Undermine Privacy Risk Perception and Protective Behaviour Concerning Smart Speakers - University of Twente Student Theses* [Bachelor Thesis, University of Twente]. https://essay.utwente.nl/95412/

Herbert, F., Becker, S., Schaewitz, L., Hielscher, J., Kowalewski, M., Sasse, A., Acar, Y., & Dürmuth, M. (2023). A World Full of Privacy and Security (Mis)conceptions? Findings of a Representative Survey in 12 Countries. *ACM*, 1–23. https://doi.org/10.1145/3544548.3581410

IBM. (2025). Internet of Things. *IBM*. https://www.ibm.com/think/topics/internet-of-things

Insecam. (2024). *FAQ on insecure cameras*. http://www.insecam.org/en/faq/

Kaminsky, S. (2024, August 12). Which devices on your network are most vulnerable? *Kaspersky*. https://www.kaspersky.com/blog/riskiest-it-and-iot-devices-in-organization/51958/#:~:text=VoIP%20devices%20and%20IP%20surveillance,used%20with%20default%2C%20insecure%20settings.

Kline, P. (2014). An easy guide to factor analysis. In *Routledge eBooks*. https://doi.org/10.4324/9781315788135

Lau, J., Zimmerman, B., & Schaub, F. (2018). Alexa, are you listening? *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 1–31. https://doi.org/10.1145/3274371

Lutz, C., & Newlands, G. (2021). Privacy and smart speakers: A multi-dimensional approach. *The Information Society*, *37*(3), 147–162. https://doi.org/10.1080/01972243.2021.1897914

Malaterre, C., Javaux, E. J., & López-García, P. (2023). Misconceptions in science. *Perspectives on Science*, *31*(6), 717–743. https://doi.org/10.1162/posc_a_00590

Malkin, N., Deatrick, J., Tong, A., Wijesekera, P., Egelman, S., & Wagner, D. (2019). Privacy attitudes of smart speaker users. *Proceedings on Privacy Enhancing Technologies*, *2019*(4), 250–271. https://doi.org/10.2478/popets-2019-0068

Mayer, P., & Volkamer, M. (2018). Addressing misconceptions about password security effectively. *Association for Computing Machinery*, 16–27. https://doi.org/10.1145/3167996.3167998

Meng, N., Keküllüoğlu, D., & Vaniea, K. (2021). Owning and sharing. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–29. https://doi.org/10.1145/3449119

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. *Journal of Consumer Affairs*, *41*(1), 100–126. https://doi.org/10.1111/j.1745-6606.2006.00070.x

Pawar, S. P. (2024, August 2). *The rise of IoT attacks: Endpoint protection via trending technologies*. Cybersecurity Exchange. https://www.eccouncil.org/cybersecurity-exchange/ethical-hacking/the-rise-of-iot-attacks-endpoint-protection-via-trending-technologies/#:~:text=Insufficient%20Physical%20Security%3A%20IoT%20devices,with%20the%20devices&apos;%20physical%20layer.

Pottkamp, C. M. (2024). *Assessing privacy risk perception and protective behaviours in relation to smart speakers : Expanding a previous model* [Bachelor Thesis, University of Twente]. https://essay.utwente.nl/100039/

Pourrahmani, H., Yavarinasab, A., Monazzah, A. M. H., & Van Herle, J. (2023). A review of the security vulnerabilities and countermeasures in the Internet of Things solutions: A bright future for the Blockchain. *Internet of Things*, *23*, 100888. https://doi.org/10.1016/j.iot.2023.100888

Scheuneman, L. (2022). *Exploring smart speaker privacy misconceptions and knowledge Gaps - University of Twente Student Theses* [Bachelor Thesis, University of Twente]. https://essay.utwente.nl/105001/

Secarma. (2024, July 31). *The growing threat of IoT cyber-attacks – What you need to know.* https://secarma.com/the-growing-threat-of-iot-cyber-attacks#:~:text=Cyber%20threats%20to%20IoT%20systems,enough%20to%20cause%20lasting%20damage.

Stanton, B., Theofanos, M. F., Prettyman, S. S., & Furman, S. (2016). Security fatigue. *IT Professional*, *18*(5), 26–32. https://doi.org/10.1109/mitp.2016.84

Statista. (2024, December 10). *Global annual number of IoT cyber attacks 2018-2022.* https://www.statista.com/statistics/1377569/worldwide-annual-internet-of-things-attacks/#statisticContainer

Suh, B., & Han, I. (2003). The impact of customer trust and perception of security control on the acceptance of electronic commerce. *International Journal of Electronic Commerce*, *7*(3), 135–161. https://doi.org/10.1080/10864415.2003.11044270

Tawalbeh, L., Muheidat, F., Tawalbeh, M., & Quwaider, M. (2020). IoT privacy and Security: challenges and solutions. *Applied Sciences*, *10*(12), 4102. https://doi.org/10.3390/app10124102

Vallabhaneni, R. (2024). Effects of Data Breaches on Internet of Things (IoT) Devices within the Proliferation of Daily-Life Integrated Devices. *Engineering and Technology Journal*, *09*(07). https://doi.org/10.47191/etj/v9i07.13

Visser, F. (2024, March 7). What is a deepfake – and how are they being used by scammers? *TBIJ*. https://www.thebureauinvestigates.com/stories/2024-03-07/what-is-a-deepfake-and-what-are-the-different-types

Wald, R., Piotrowski, J. T., Van Oosten, J. M., & Araujo, T. (2024). Who are the (Non-)Adopters of Smart Speakers? A Cross-Sectional Survey Study of Dutch Families. *Tijdschrift Voor Communicatiewetenschappen*, *52*(1), 4–28. https://doi.org/10.5117/tcw2023.x.001.wald

Yang, Y., Wu, L., Yin, G., Li, L., & Zhao, H. (2017). A survey on security and privacy issues in Internet-of-Things. *IEEE Internet of Things Journal*, *4*(5), 1250–1258. https://doi.org/10.1109/jiot.2017.2694844

Zeng, E., Mare, S., Roesner, F., & University of Washington. (2017). End User Security and Privacy Concerns with Smart Homes. In *Proceedings of the Thirteenth Symposium on Usable Privacy and Security*. USENIX Association. https://www.usenix.org/system/files/conference/soups2017/soups2017-zeng.pdf

Appendix A

Informed consent form

Exploring factors influencing smart speaker security protective behaviours in users and non-users

Study Investigators:

Researcher: Karam Altabbaa, University of Twente

Supervisors: Michelle Walterscheid, Nicole Huijts

Contact Information: karamahmadabdelelahaltabbaa@student.utwente.nl

Invitation to Participate

You are invited to participate in a research study that examines the misconceptions surrounding smart speakers (such as Amazon Alexa and Google Home) and the relationship between these misconceptions and security protective behaviors in both users and non-users. The study is conducted by Karam Altabbaa at the University of Twente under the supervision of Michelle Walterscheid and Nicole Huijts. Your participation is voluntary, and you may choose to withdraw at any time without any negative consequences and your data will not be used in the analysis. There are no known risks to participating in this study. Students recruited through SONA will be awarded with SONA points. Otherwise, there are no other direct benefits to you, however your responses will help improve our understanding of privacy issues with smart speakers, potentially leading to better security protections. This form provides important details about the study, if you have any questions, please contact the lead researcher at: karamahmadabdelelahaltabbaa@student.utwente.nl

Purpose of the Research Study

The purpose of this study is to understand how misconceptions about smart speakers influence user security behaviors online, and if there is a difference between users and non-users in that regard. The study will also examine the role of trust in technology providers and how it impacts security practices among smart speaker users and non-users.

What You Will Be Asked to Do

If you choose to participate, you will be asked to complete an online survey that will take approximately 15-20 minutes to complete. The survey will consist of:

- Questions regarding your demographics, and smart speaker ownership status.

- Questions that assess your knowledge and misconceptions about smart speaker privacy and security.

- Questions that measure your trust in smart speaker manufacturers and your security protective behaviors.

Who Can Take Part in the Research Study?

We are recruiting adults aged 18 and older who either use or do not use smart speakers. Participants must have access to the internet and be willing to complete the online survey.

Privacy and Confidentiality

- Your responses will be anonymized and stored in a secure database.
- No personally identifiable information will be linked to your responses.
- Data will be anonymized and retained for at least ten years for the purpose of research integrity and may be used for future research related to privacy and IoT security

Reporting of Results

Results from this study may be published in academic journals, presented at conferences, or included in policy briefs. All findings will be reported in a way that ensures individual participants cannot be identified. If you wish to receive a summary of the research findings, please email the lead researcher at karamahmadabdelelahaltabbaa@student.utwente.nl

Withdrawing from the Study

Participation is completely voluntary. You may withdraw at any time without providing a reason. If you choose to withdraw, your data will be removed from the study unless results have already been published.

Questions and Contact Information

If you have any questions about the study, please contact: Karam Altabbaa, University of Twente, karamahmadabdelelahaltabbaa@student.utwente.nl or Michelle Walterscheid, University of Twente, m.k.walterscheid@utwente.nl

If you have questions about your rights as a research participant, you may contact: The Humanities & Social Sciences Ethical Committee of the University of Twente at ethicscommittee-hss@utwente.nl.

Appendix B

Signature Page

Exploring factors influencing smart speaker security protective behaviours in users and non-users

Lead Researcher: Karam Altabbaa

Supervisors: Michelle Walterscheid, Nicole Huijts

Statement of Consent

By pressing agree on this page, I agree that:

- The study has been explained to me.

- I understand that personal identifiable information will not be collected

- I consent that the information I provide may be used for analysis and research, and may be published in scientific outlet in an anonymized form.

- I have the opportunity to ask questions.

- I understand that my participation is voluntary and that I may withdraw at any time without consequence.

- I agree to participate in the study.

If you would like to receive a summary of the research findings, please email the lead researcher at karamahmadabdelelahaltabbaa@student.utwente.nl

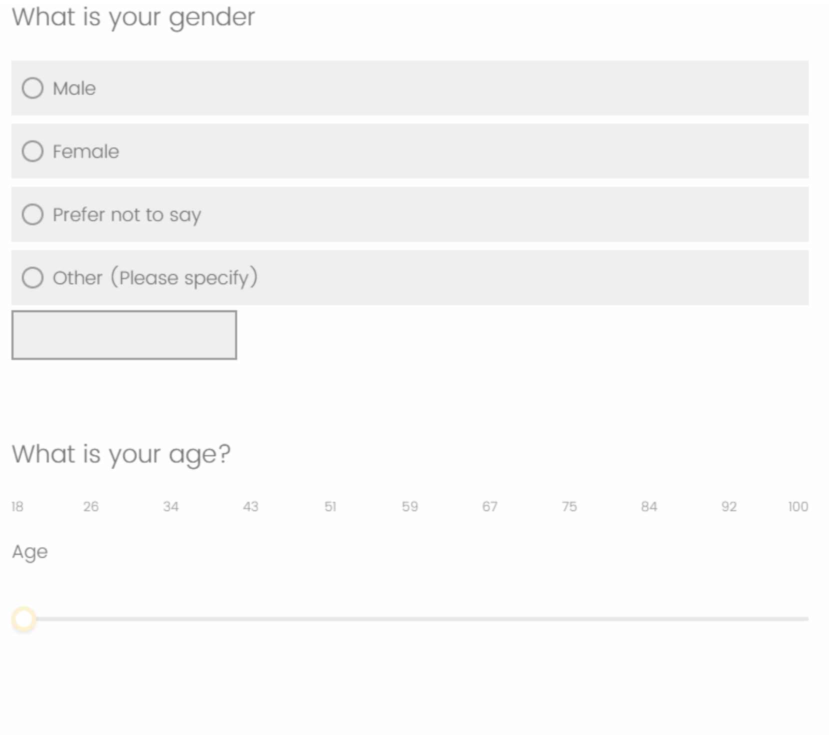
Thank you for your participation!

In order to continue with this survey, you have to agree with the aforementioned information and consent to participate in the study. Clicking "I agree and consent to participating in this study and confirm that I am over 18 years old" indicates that you have been informed about the nature and method of this research in a manner that is clear to you, you have been given the time to read the page, and that you voluntarily agree to participate in this study.

- ➢ I agree and consent to participating in this study and confirm that I am over 18 years old
- ➢ No, I do not agree to participate in this study

Appendix C

Demographic Questions



What is your gender

○ Male

○ Female

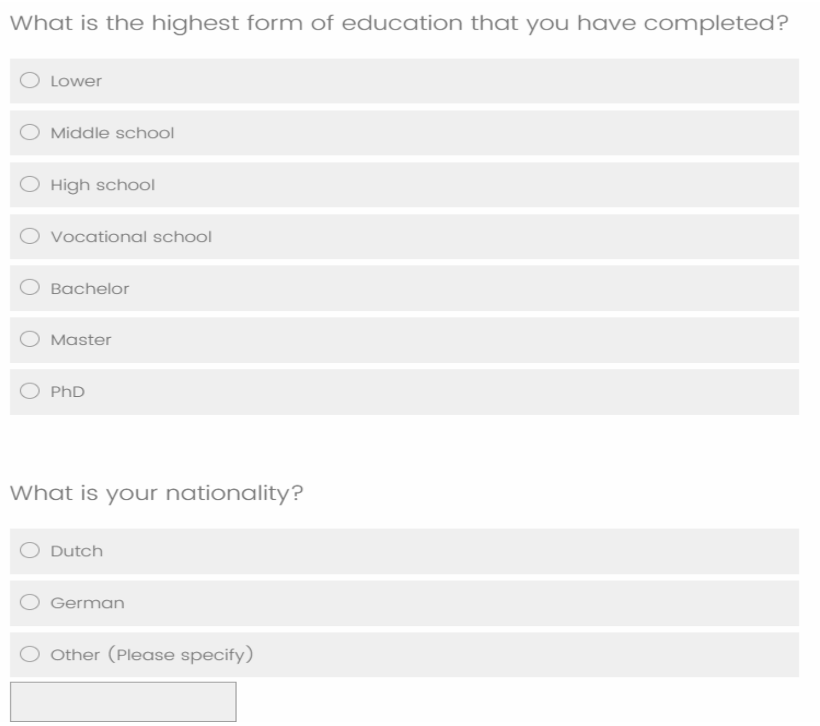
○ Prefer not to say

○ Other (Please specify)

What is your age?

18 26 34 43 51 59 67 75 84 92 100

Age



What is the highest form of education that you have completed?

○ Lower

○ Middle school

○ High school

○ Vocational school

○ Bachelor

○ Master

○ PhD

What is your nationality?

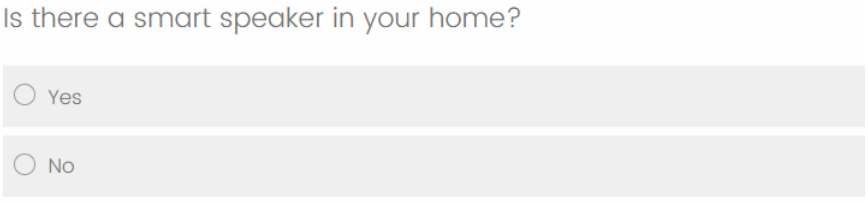
○ Dutch

○ German

○ Other (Please specify)

Appendix D

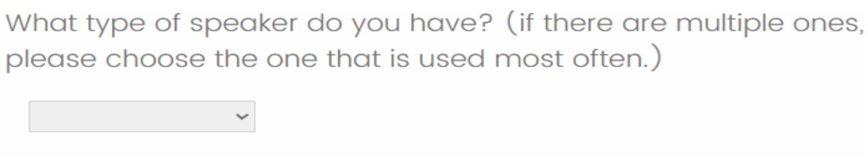
Smart Speaker control questions



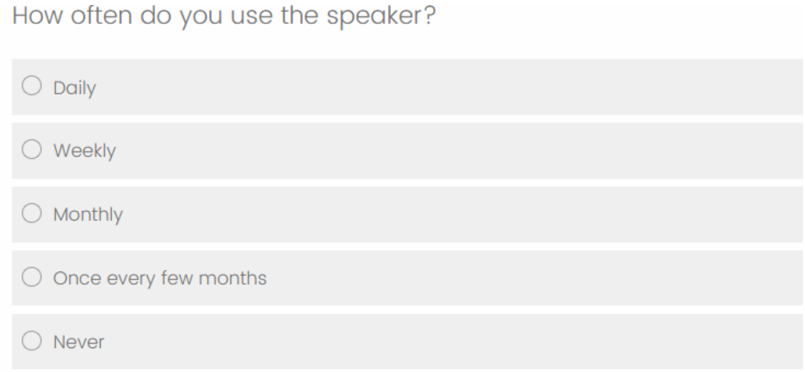
Is there a smart speaker in your home?

○ Yes

○ No



What type of speaker do you have? (if there are multiple ones, please choose the one that is used most often.)



How often do you use the speaker?

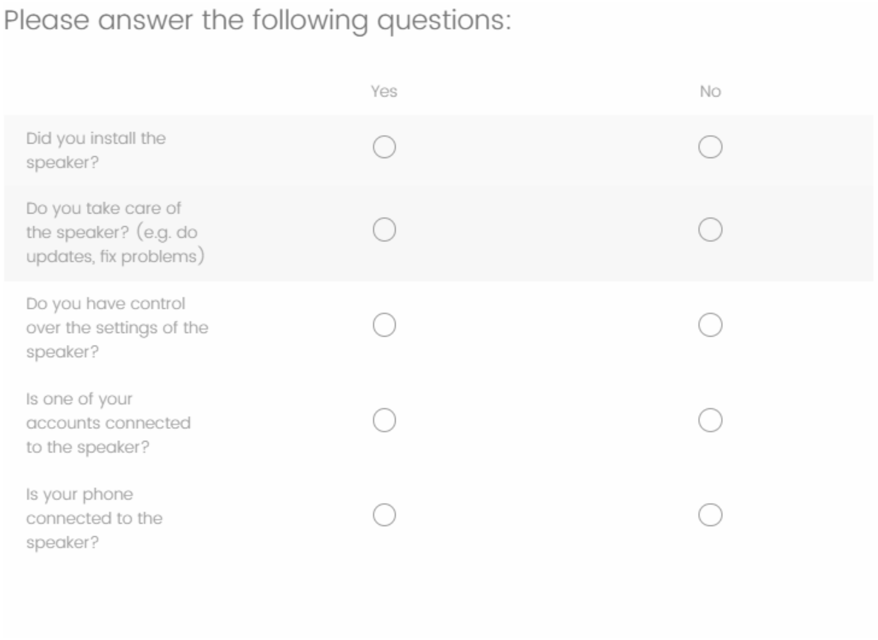
○ Daily

○ Weekly

○ Monthly

○ Once every few months

○ Never

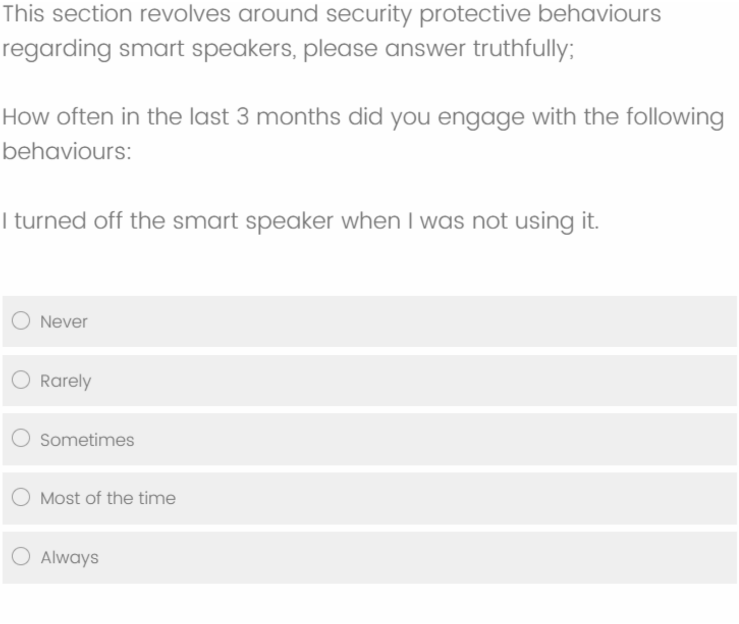


Please answer the following questions:

	Yes	No
Did you install the speaker?	○	○
Do you take care of the speaker? (e.g. do updates, fix problems)	○	○
Do you have control over the settings of the speaker?	○	○
Is one of your accounts connected to the speaker?	○	○
Is your phone connected to the speaker?	○	○

Appendix E

User protective behaviour questions



This section revolves around security protective behaviours regarding smart speakers, please answer truthfully;

How often in the last 3 months did you engage with the following behaviours:

I turned off the smart speaker when I was not using it.

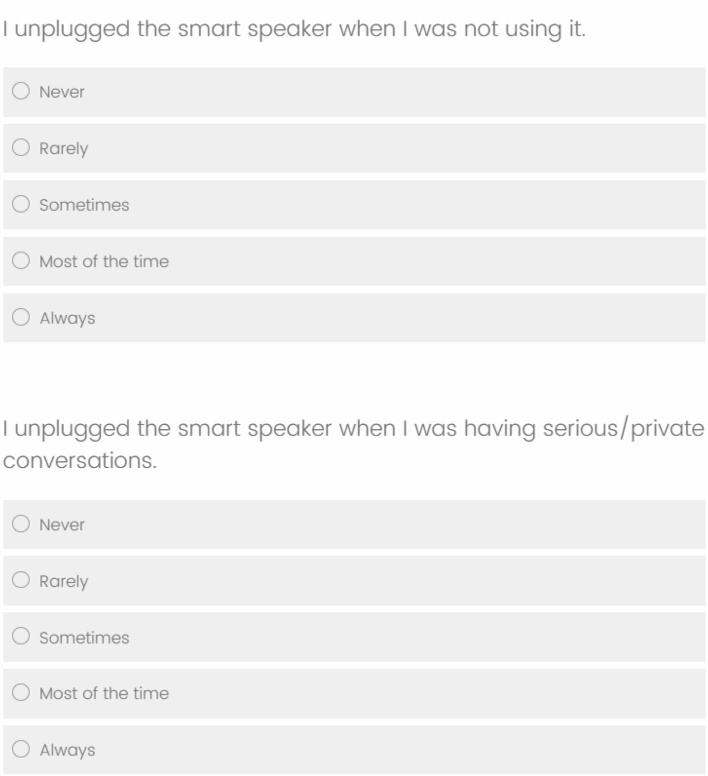
○ Never

○ Rarely

○ Sometimes

○ Most of the time

○ Always



I unplugged the smart speaker when I was not using it.

○ Never

○ Rarely

○ Sometimes

○ Most of the time

○ Always

I unplugged the smart speaker when I was having serious/private conversations.

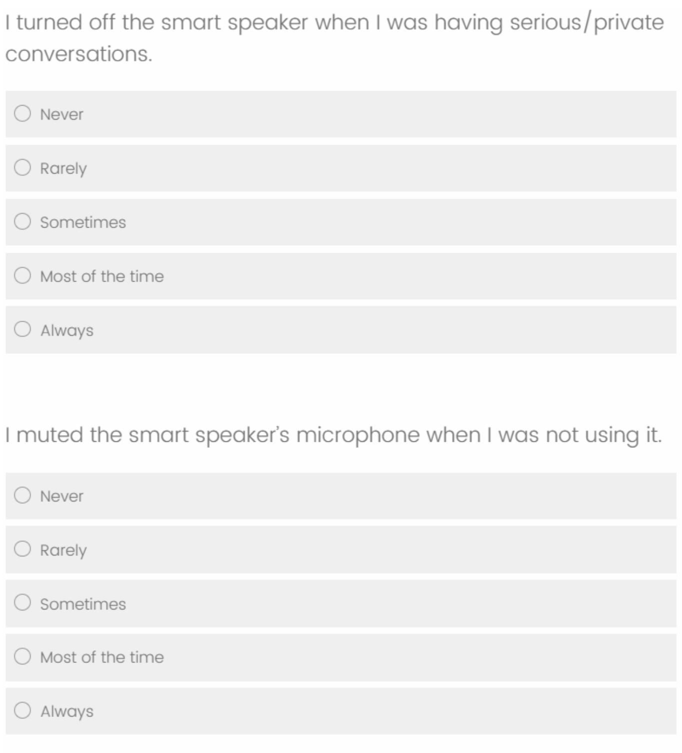
○ Never

○ Rarely

○ Sometimes

○ Most of the time

○ Always



I turned off the smart speaker when I was having serious/private conversations.

○ Never

○ Rarely

○ Sometimes

○ Most of the time

○ Always

I muted the smart speaker's microphone when I was not using it.

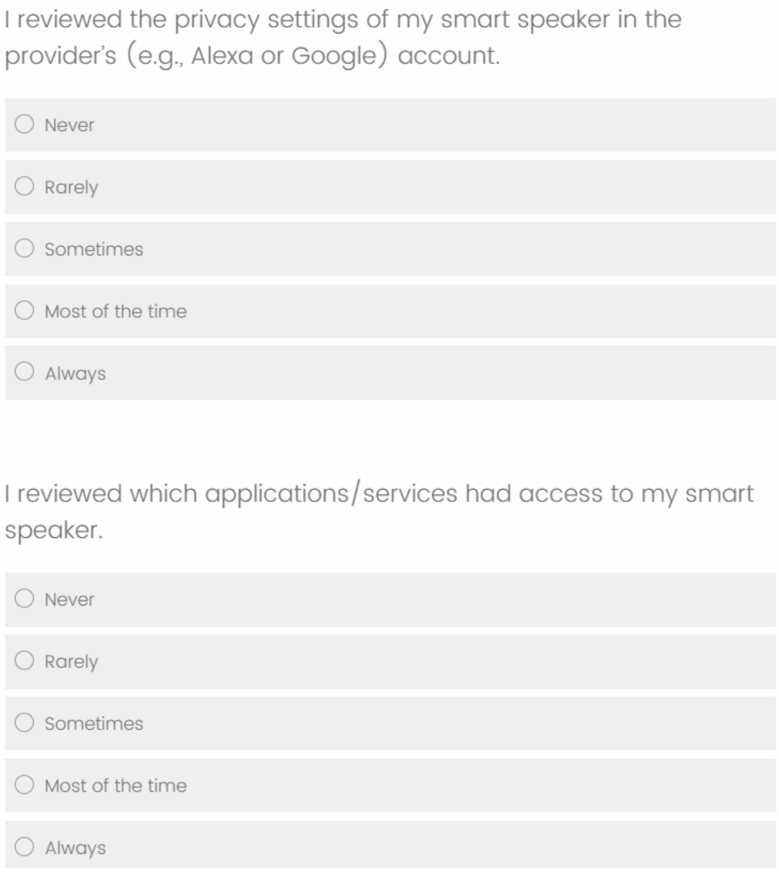
○ Never

○ Rarely

○ Sometimes

○ Most of the time

○ Always



I reviewed the privacy settings of my smart speaker in the provider's (e.g., Alexa or Google) account.

○ Never

○ Rarely

○ Sometimes

○ Most of the time

○ Always

I reviewed which applications/services had access to my smart speaker.

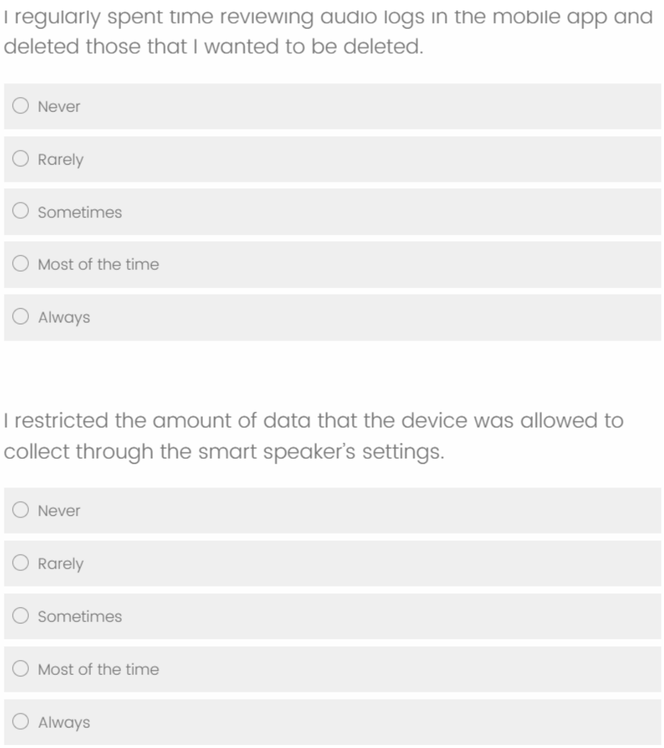
○ Never

○ Rarely

○ Sometimes

○ Most of the time

○ Always



I regularly spent time reviewing audio logs in the mobile app and deleted those that I wanted to be deleted.

○ Never

○ Rarely

○ Sometimes

○ Most of the time

○ Always

I restricted the amount of data that the device was allowed to collect through the smart speaker's settings.

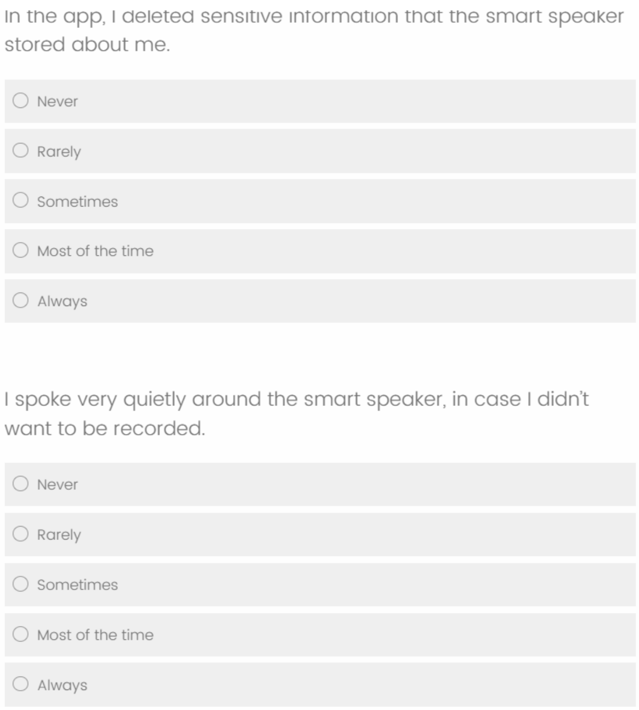
○ Never

○ Rarely

○ Sometimes

○ Most of the time

○ Always



In the app, I deleted sensitive information that the smart speaker stored about me.

○ Never

○ Rarely

○ Sometimes

○ Most of the time

○ Always

I spoke very quietly around the smart speaker, in case I didn't want to be recorded.

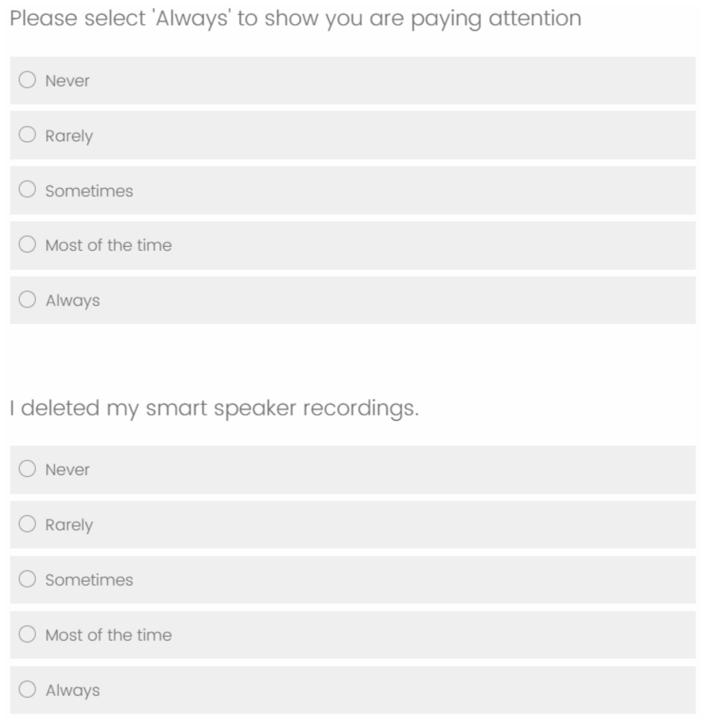
○ Never

○ Rarely

○ Sometimes

○ Most of the time

○ Always



Please select 'Always' to show you are paying attention

○ Never

○ Rarely

○ Sometimes

○ Most of the time

○ Always

I deleted my smart speaker recordings.

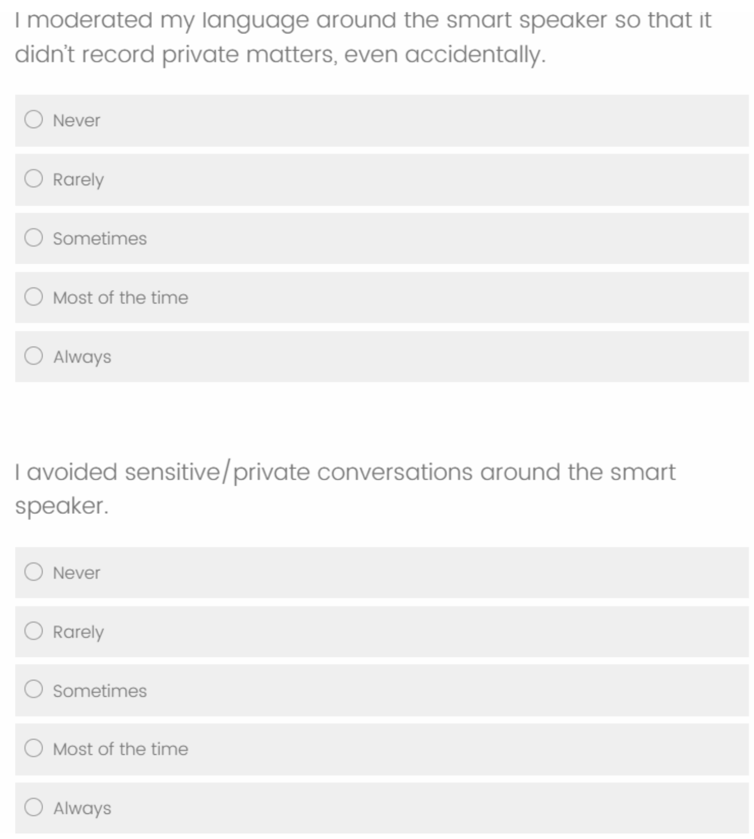
○ Never

○ Rarely

○ Sometimes

○ Most of the time

○ Always



I moderated my language around the smart speaker so that it didn't record private matters, even accidentally.

○ Never

○ Rarely

○ Sometimes

○ Most of the time

○ Always

I avoided sensitive/private conversations around the smart speaker.

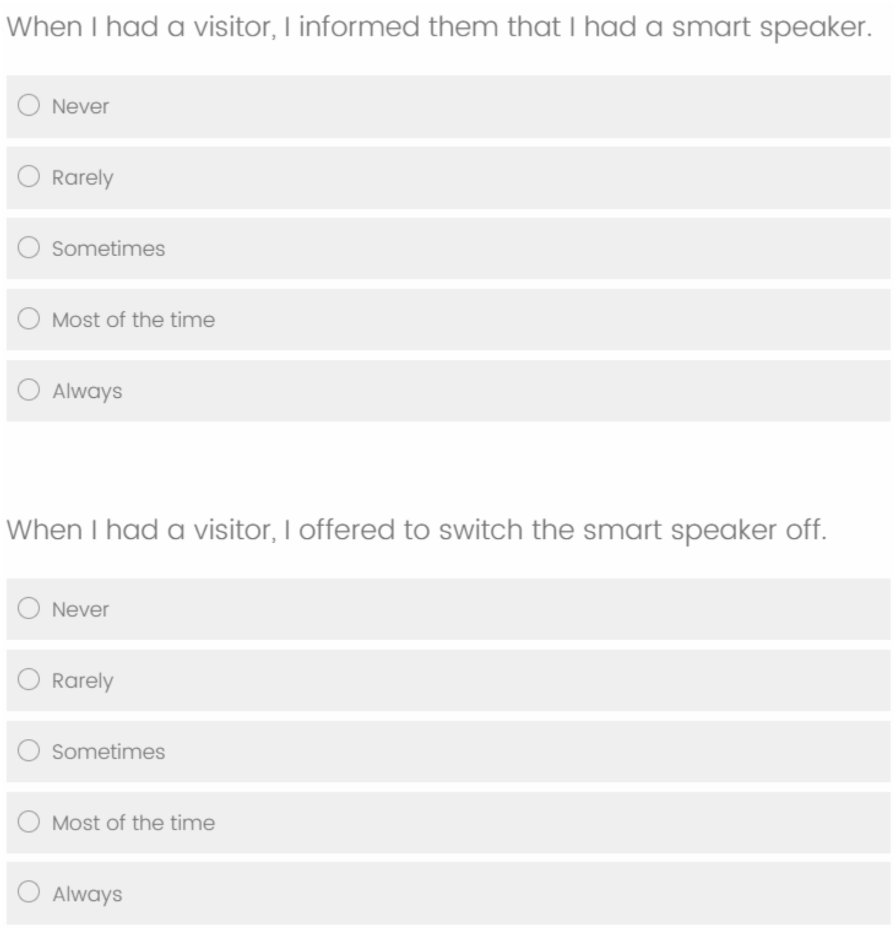
○ Never

○ Rarely

○ Sometimes

○ Most of the time

○ Always



When I had a visitor, I informed them that I had a smart speaker.

○ Never

○ Rarely

○ Sometimes

○ Most of the time

○ Always

When I had a visitor, I offered to switch the smart speaker off.

○ Never

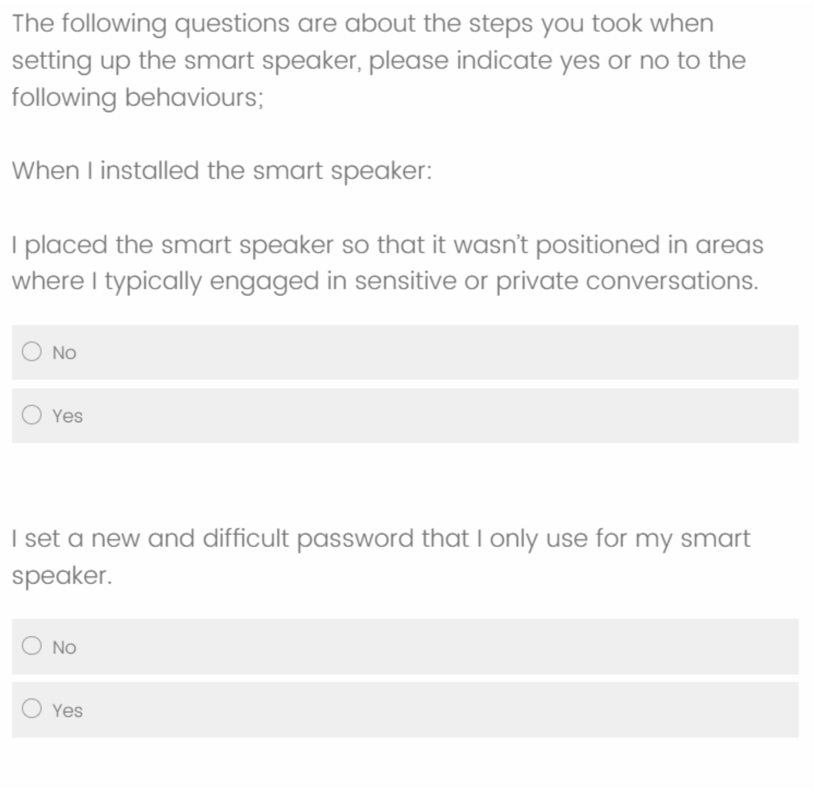
○ Rarely

○ Sometimes

○ Most of the time

○ Always

User smart speaker setup behaviour questions



The following questions are about the steps you took when setting up the smart speaker, please indicate yes or no to the following behaviours;

When I installed the smart speaker:

I placed the smart speaker so that it wasn't positioned in areas where I typically engaged in sensitive or private conversations.

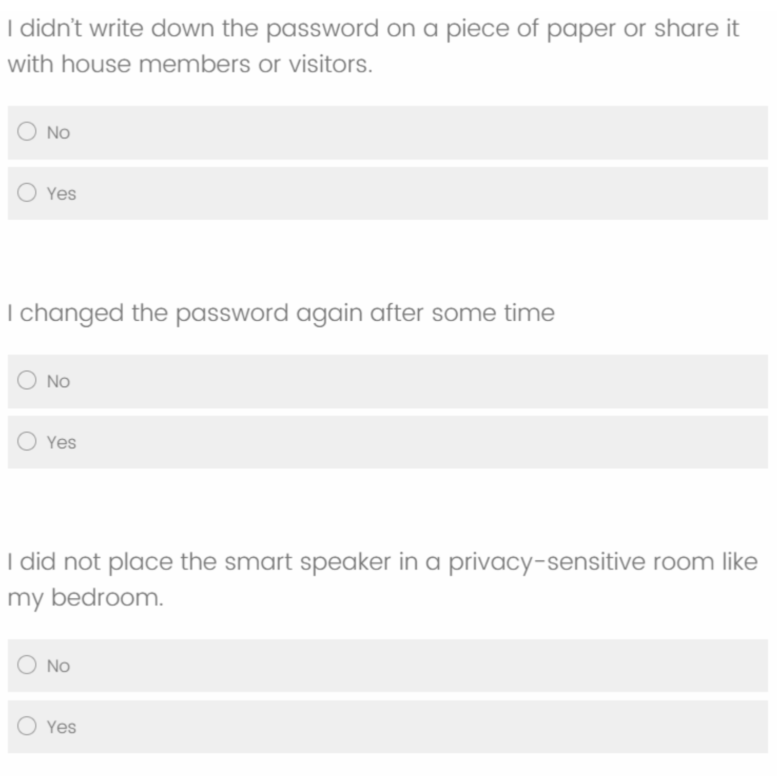
○ No

○ Yes

I set a new and difficult password that I only use for my smart speaker.

○ No

○ Yes



I didn't write down the password on a piece of paper or share it with house members or visitors.

○ No

○ Yes

I changed the password again after some time

○ No

○ Yes

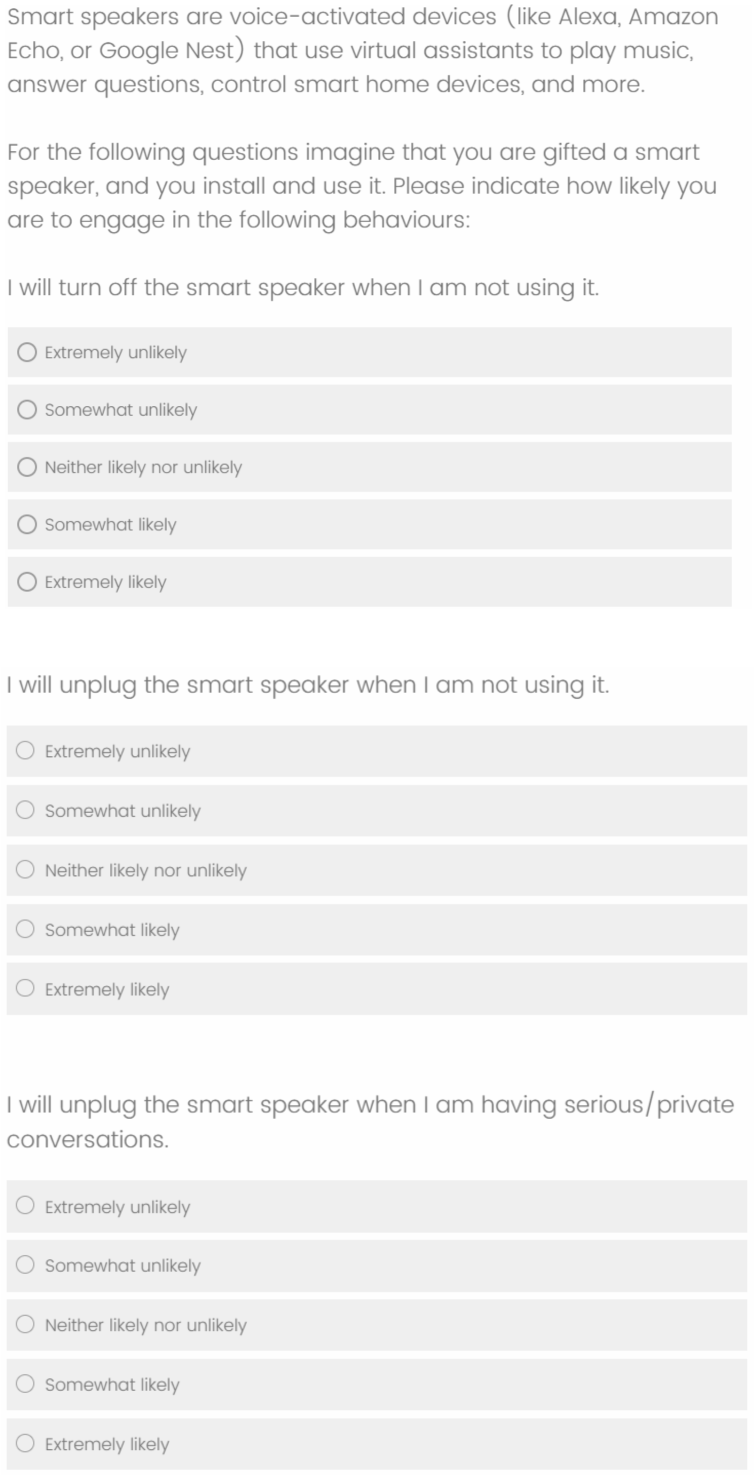
I did not place the smart speaker in a privacy-sensitive room like my bedroom.

○ No

○ Yes

Appendix F

Non-User protective behaviour scale



Smart speakers are voice-activated devices (like Alexa, Amazon Echo, or Google Nest) that use virtual assistants to play music, answer questions, control smart home devices, and more.

For the following questions imagine that you are gifted a smart speaker, and you install and use it. Please indicate how likely you are to engage in the following behaviours:

I will turn off the smart speaker when I am not using it.

○ Extremely unlikely

○ Somewhat unlikely

○ Neither likely nor unlikely

○ Somewhat likely

○ Extremely likely

I will unplug the smart speaker when I am not using it.

○ Extremely unlikely

○ Somewhat unlikely

○ Neither likely nor unlikely

○ Somewhat likely

○ Extremely likely

I will unplug the smart speaker when I am having serious/private conversations.

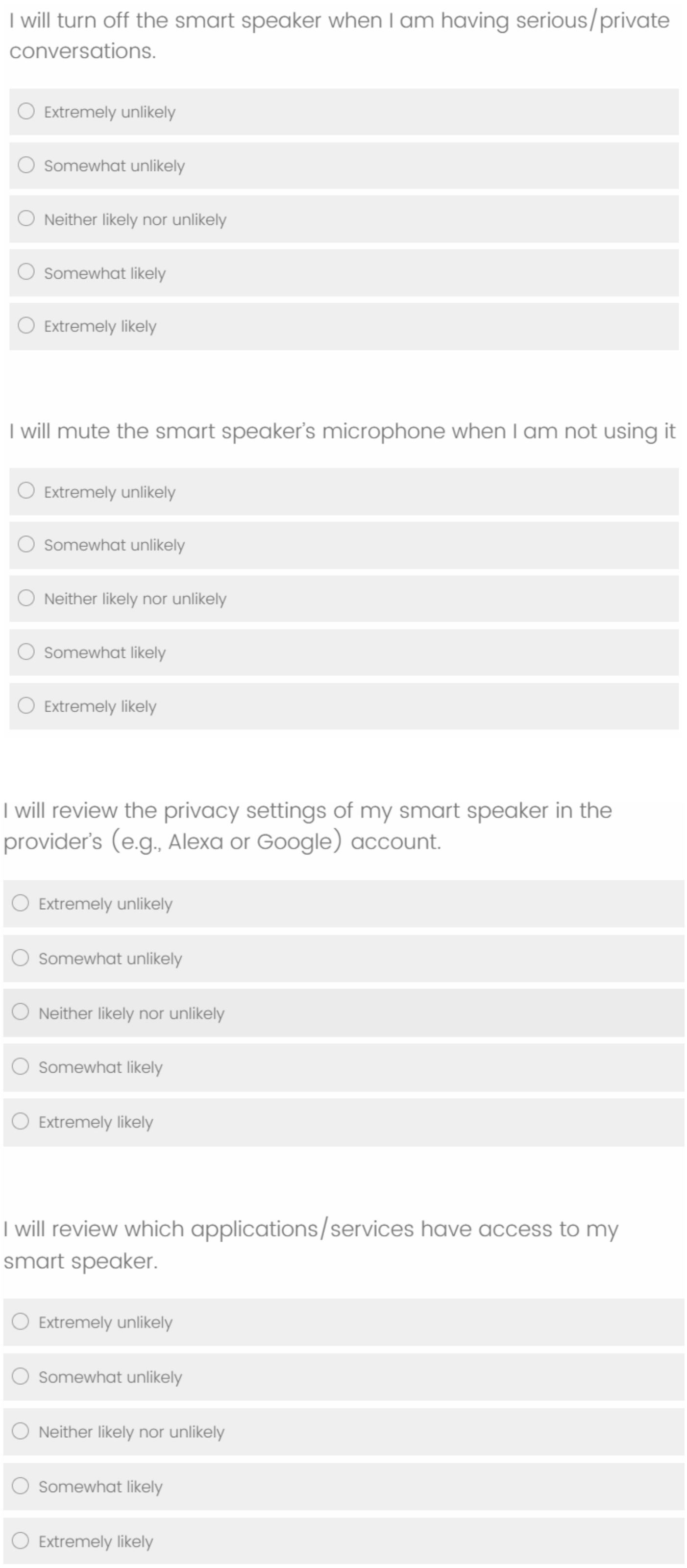
○ Extremely unlikely

○ Somewhat unlikely

○ Neither likely nor unlikely

○ Somewhat likely

○ Extremely likely



I will turn off the smart speaker when I am having serious/private conversations.

○ Extremely unlikely

○ Somewhat unlikely

○ Neither likely nor unlikely

○ Somewhat likely

○ Extremely likely

I will mute the smart speaker's microphone when I am not using it

○ Extremely unlikely

○ Somewhat unlikely

○ Neither likely nor unlikely

○ Somewhat likely

○ Extremely likely

I will review the privacy settings of my smart speaker in the provider's (e.g., Alexa or Google) account.

○ Extremely unlikely

○ Somewhat unlikely

○ Neither likely nor unlikely

○ Somewhat likely

○ Extremely likely

I will review which applications/services have access to my smart speaker.

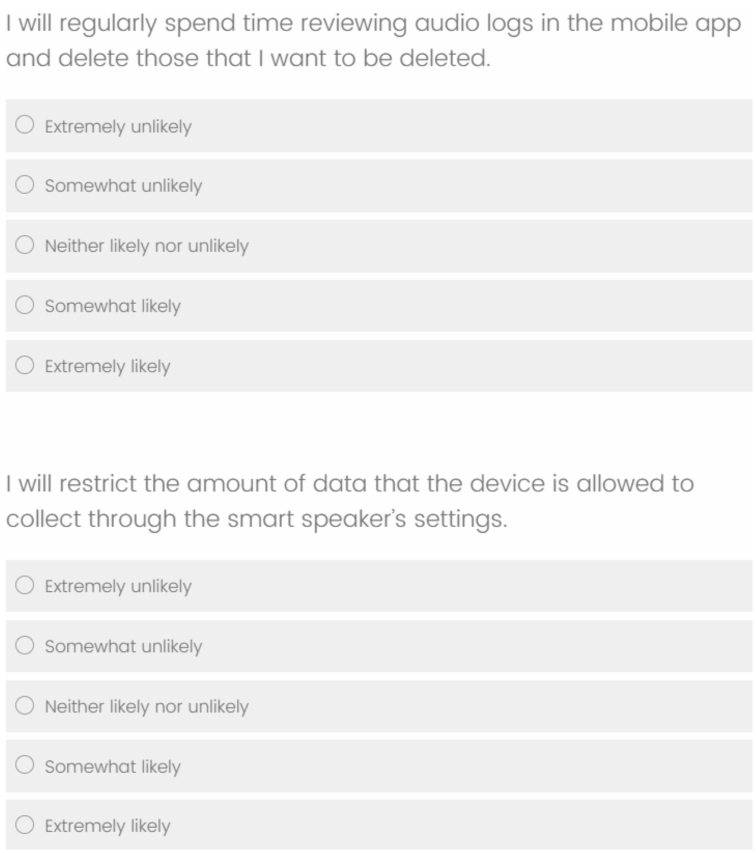
○ Extremely unlikely

○ Somewhat unlikely

○ Neither likely nor unlikely

○ Somewhat likely

○ Extremely likely



I will regularly spend time reviewing audio logs in the mobile app and delete those that I want to be deleted.

○ Extremely unlikely

○ Somewhat unlikely

○ Neither likely nor unlikely

○ Somewhat likely

○ Extremely likely

I will restrict the amount of data that the device is allowed to collect through the smart speaker's settings.

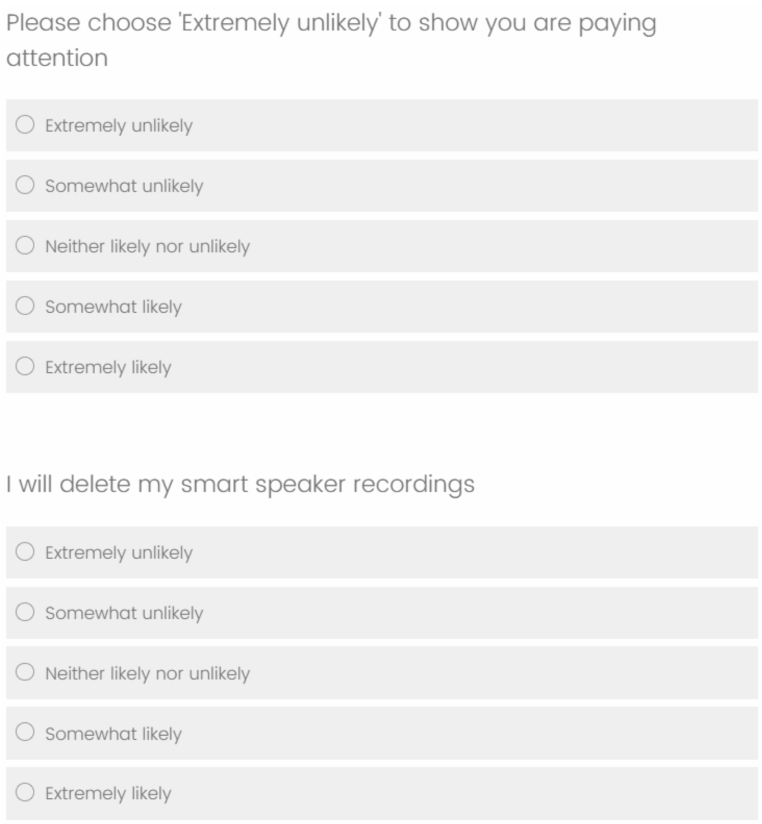
○ Extremely unlikely

○ Somewhat unlikely

○ Neither likely nor unlikely

○ Somewhat likely

○ Extremely likely



Please choose 'Extremely unlikely' to show you are paying attention

○ Extremely unlikely

○ Somewhat unlikely

○ Neither likely nor unlikely

○ Somewhat likely

○ Extremely likely

I will delete my smart speaker recordings

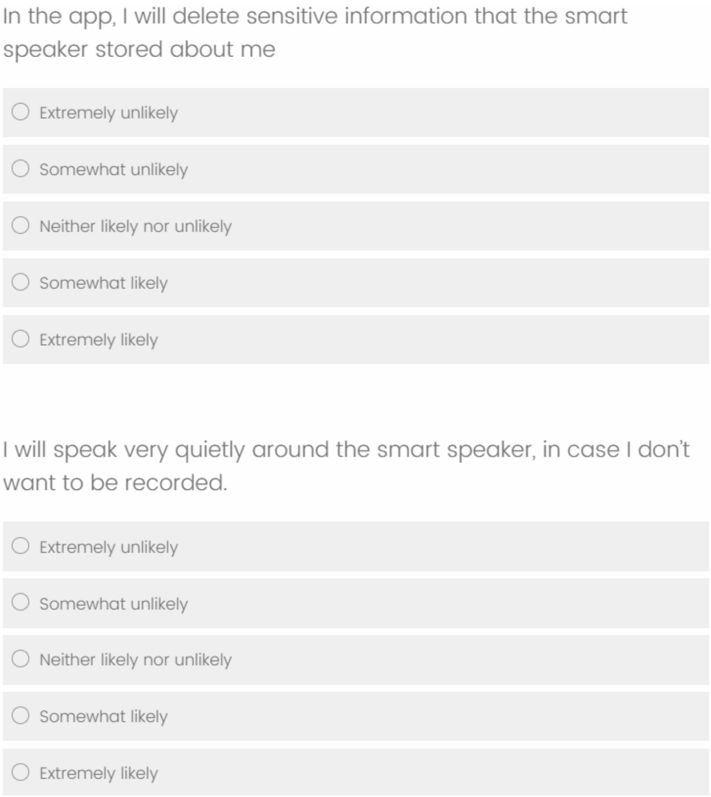
○ Extremely unlikely

○ Somewhat unlikely

○ Neither likely nor unlikely

○ Somewhat likely

○ Extremely likely



In the app, I will delete sensitive information that the smart speaker stored about me

○ Extremely unlikely

○ Somewhat unlikely

○ Neither likely nor unlikely

○ Somewhat likely

○ Extremely likely

I will speak very quietly around the smart speaker, in case I don't want to be recorded.

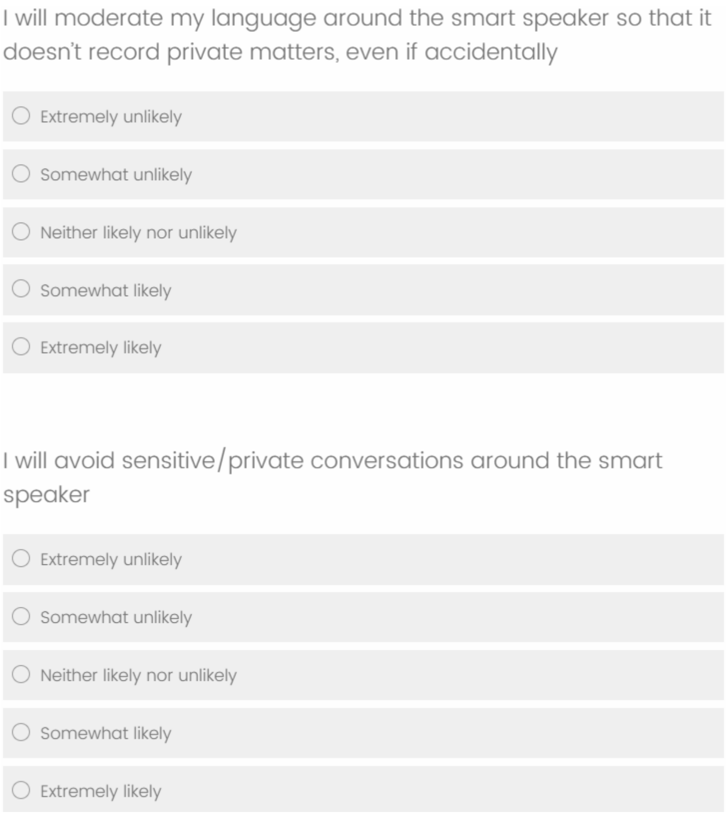
○ Extremely unlikely

○ Somewhat unlikely

○ Neither likely nor unlikely

○ Somewhat likely

○ Extremely likely



I will moderate my language around the smart speaker so that it doesn't record private matters, even if accidentally

○ Extremely unlikely

○ Somewhat unlikely

○ Neither likely nor unlikely

○ Somewhat likely

○ Extremely likely

I will avoid sensitive/private conversations around the smart speaker

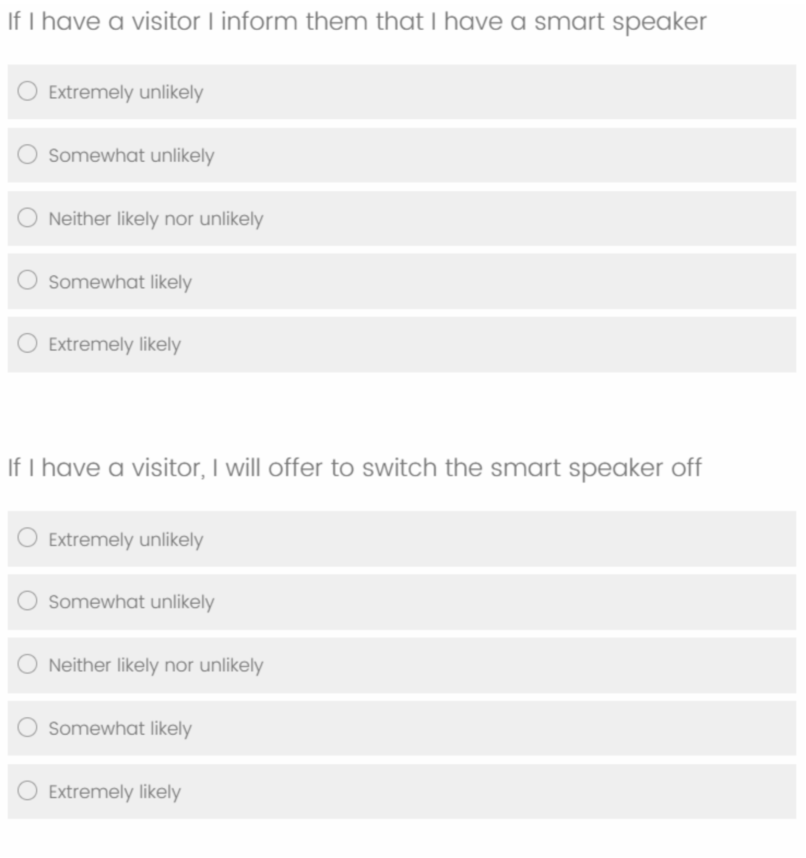
○ Extremely unlikely

○ Somewhat unlikely

○ Neither likely nor unlikely

○ Somewhat likely

○ Extremely likely



If I have a visitor I inform them that I have a smart speaker

○ Extremely unlikely

○ Somewhat unlikely

○ Neither likely nor unlikely

○ Somewhat likely

○ Extremely likely

If I have a visitor, I will offer to switch the smart speaker off

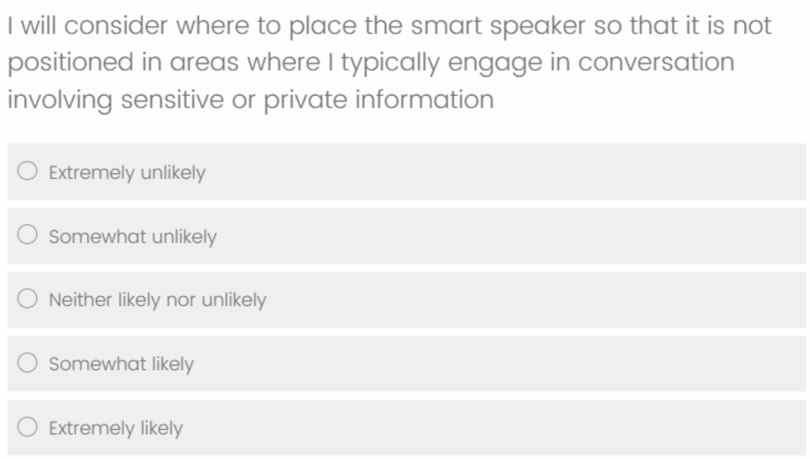
○ Extremely unlikely

○ Somewhat unlikely

○ Neither likely nor unlikely

○ Somewhat likely

○ Extremely likely



I will consider where to place the smart speaker so that it is not positioned in areas where I typically engage in conversation involving sensitive or private information

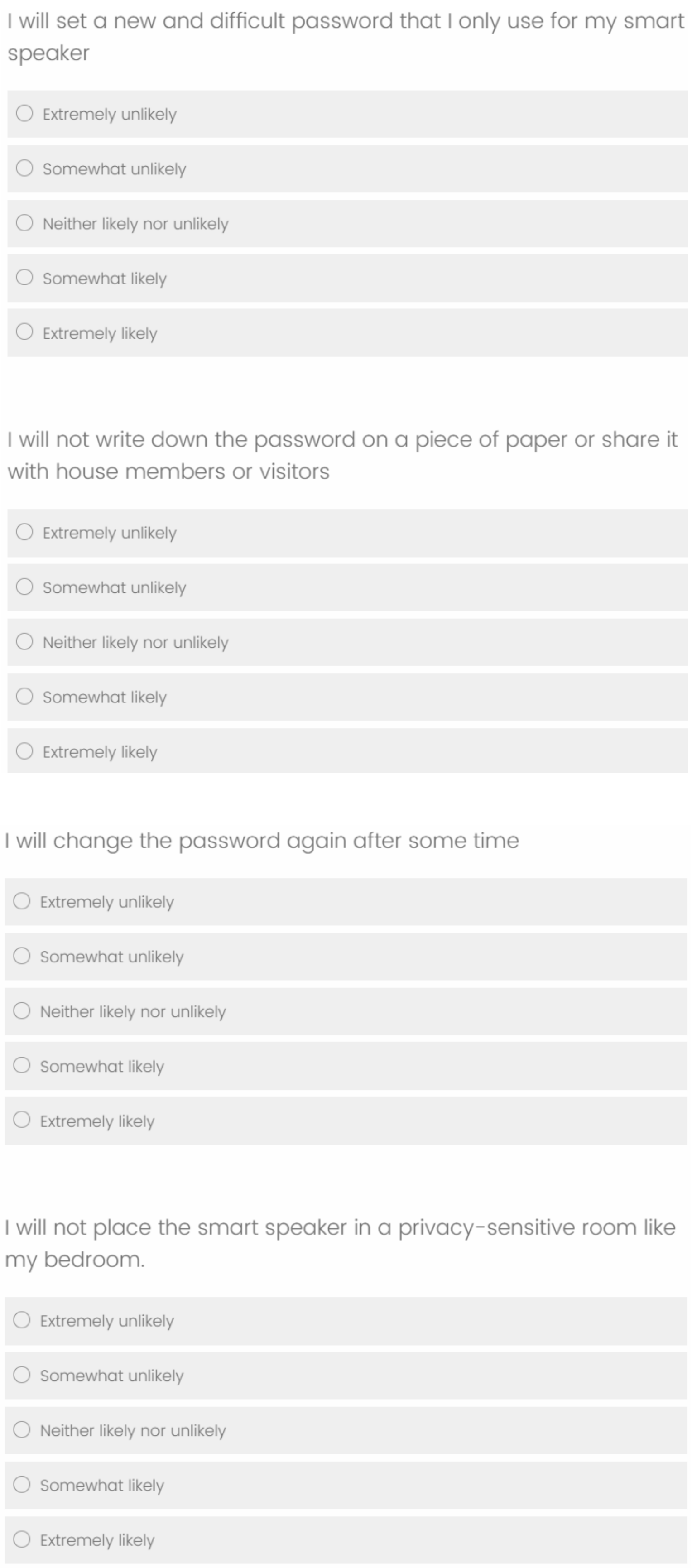
○ Extremely unlikely

○ Somewhat unlikely

○ Neither likely nor unlikely

○ Somewhat likely

○ Extremely likely



I will set a new and difficult password that I only use for my smart speaker

○ Extremely unlikely

○ Somewhat unlikely

○ Neither likely nor unlikely

○ Somewhat likely

○ Extremely likely

I will not write down the password on a piece of paper or share it with house members or visitors

○ Extremely unlikely

○ Somewhat unlikely

○ Neither likely nor unlikely

○ Somewhat likely

○ Extremely likely

I will change the password again after some time

○ Extremely unlikely

○ Somewhat unlikely

○ Neither likely nor unlikely

○ Somewhat likely

○ Extremely likely

I will not place the smart speaker in a privacy-sensitive room like my bedroom.

○ Extremely unlikely

○ Somewhat unlikely

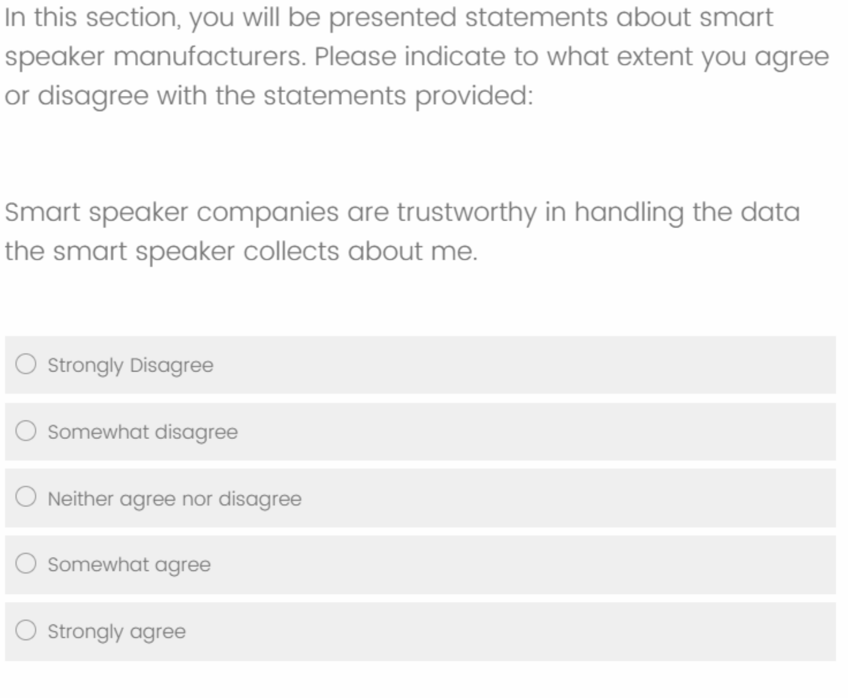
○ Neither likely nor unlikely

○ Somewhat likely

○ Extremely likely

Appendix G

Manufacturer trust scale



In this section, you will be presented statements about smart speaker manufacturers. Please indicate to what extent you agree or disagree with the statements provided:

Smart speaker companies are trustworthy in handling the data the smart speaker collects about me.

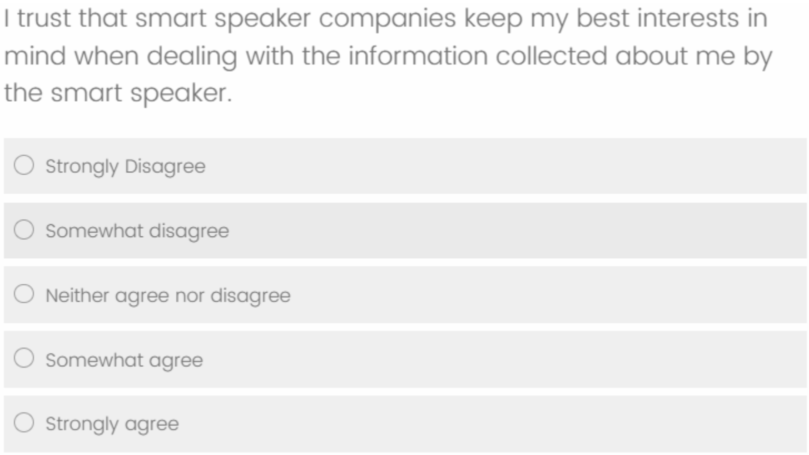
○ Strongly Disagree

○ Somewhat disagree

○ Neither agree nor disagree

○ Somewhat agree

○ Strongly agree



I trust that smart speaker companies keep my best interests in mind when dealing with the information collected about me by the smart speaker.

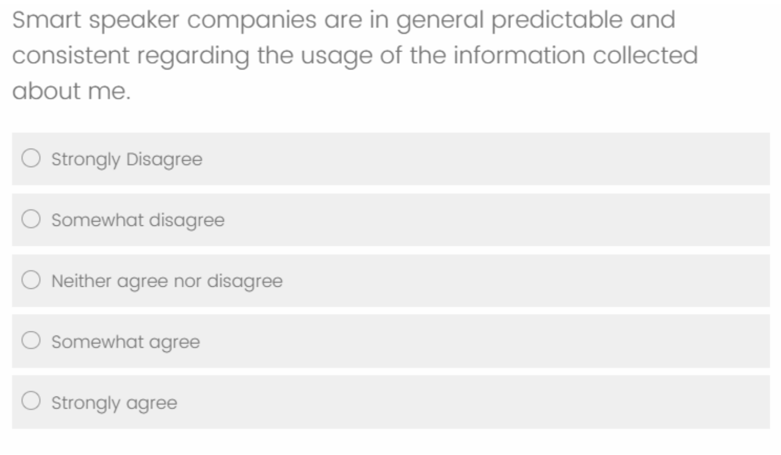
○ Strongly Disagree

○ Somewhat disagree

○ Neither agree nor disagree

○ Somewhat agree

○ Strongly agree


Smart speaker companies are in general predictable and consistent regarding the usage of the information collected about me.

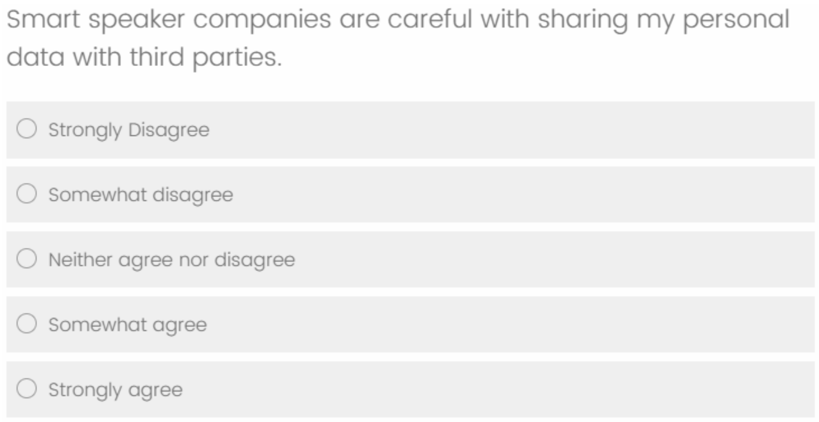
○ Strongly Disagree

○ Somewhat disagree

○ Neither agree nor disagree

○ Somewhat agree

○ Strongly agree


Smart speaker companies are careful with sharing my personal data with third parties.

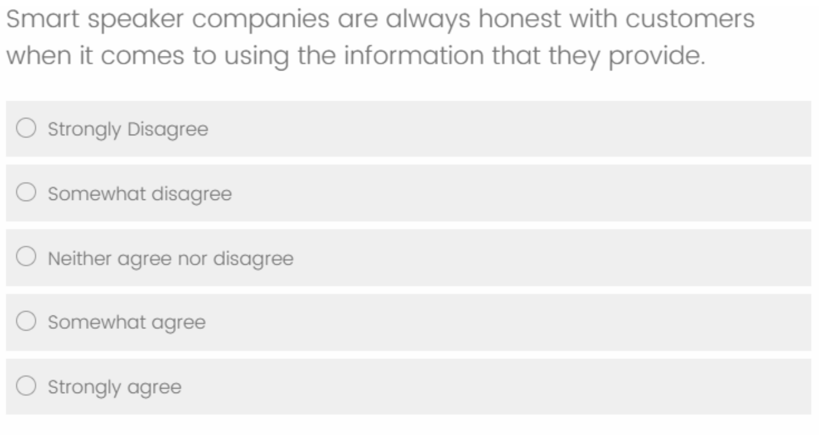
○ Strongly Disagree

○ Somewhat disagree

○ Neither agree nor disagree

○ Somewhat agree

○ Strongly agree


Smart speaker companies are always honest with customers when it comes to using the information that they provide.

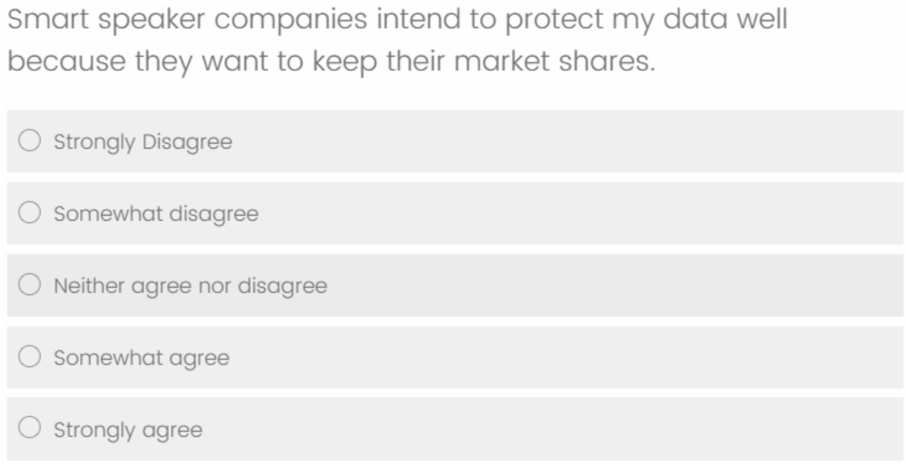
○ Strongly Disagree

○ Somewhat disagree

○ Neither agree nor disagree

○ Somewhat agree

○ Strongly agree



Smart speaker companies intend to protect my data well because they want to keep their market shares.

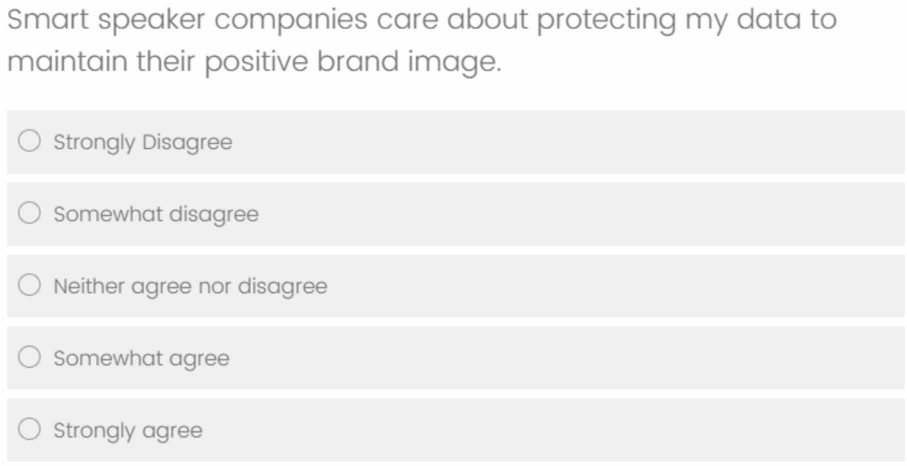
○ Strongly Disagree

○ Somewhat disagree

○ Neither agree nor disagree

○ Somewhat agree

○ Strongly agree



Smart speaker companies care about protecting my data to maintain their positive brand image.

○ Strongly Disagree

○ Somewhat disagree

○ Neither agree nor disagree

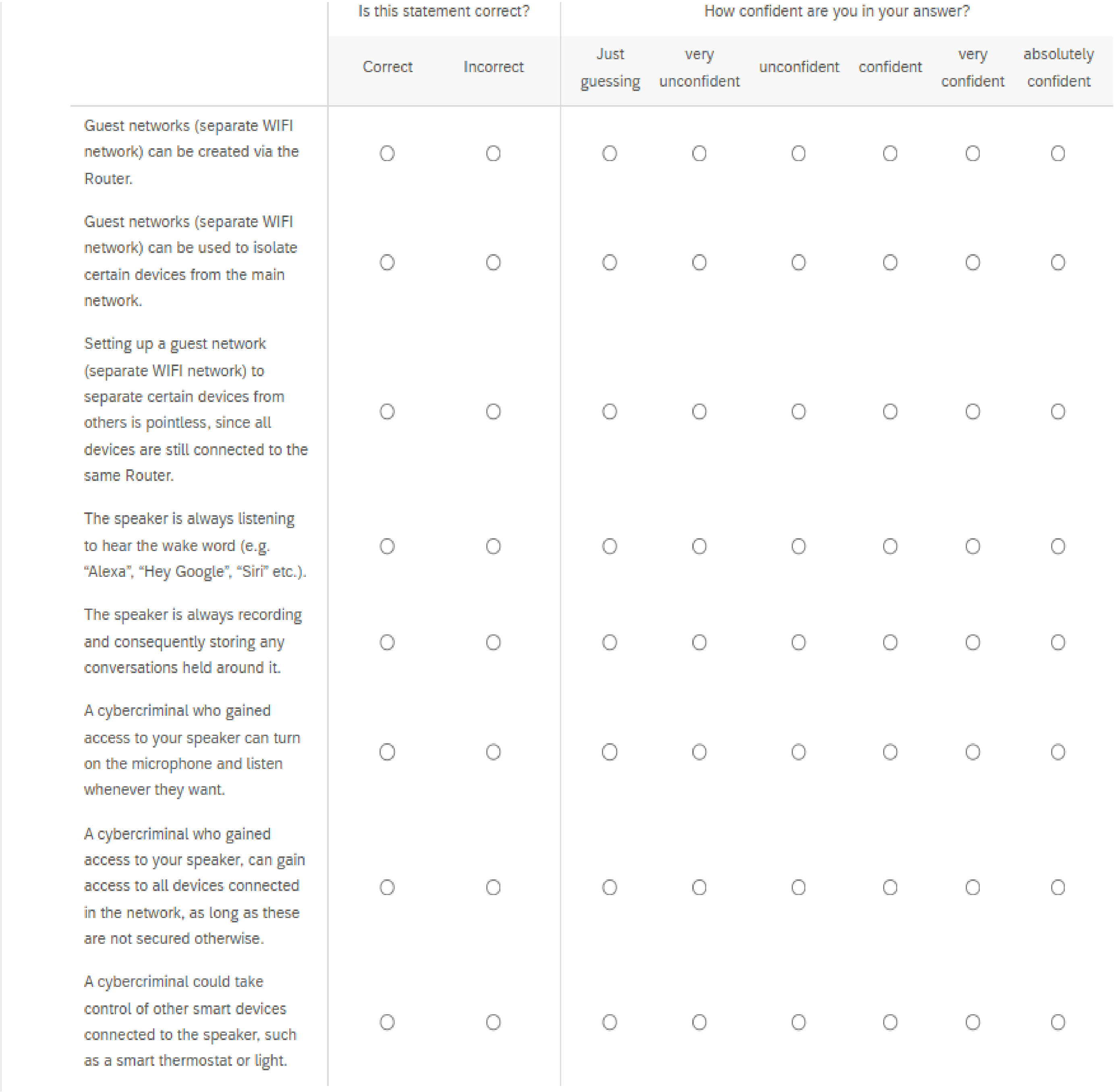
○ Somewhat agree

○ Strongly agree

Appendix H

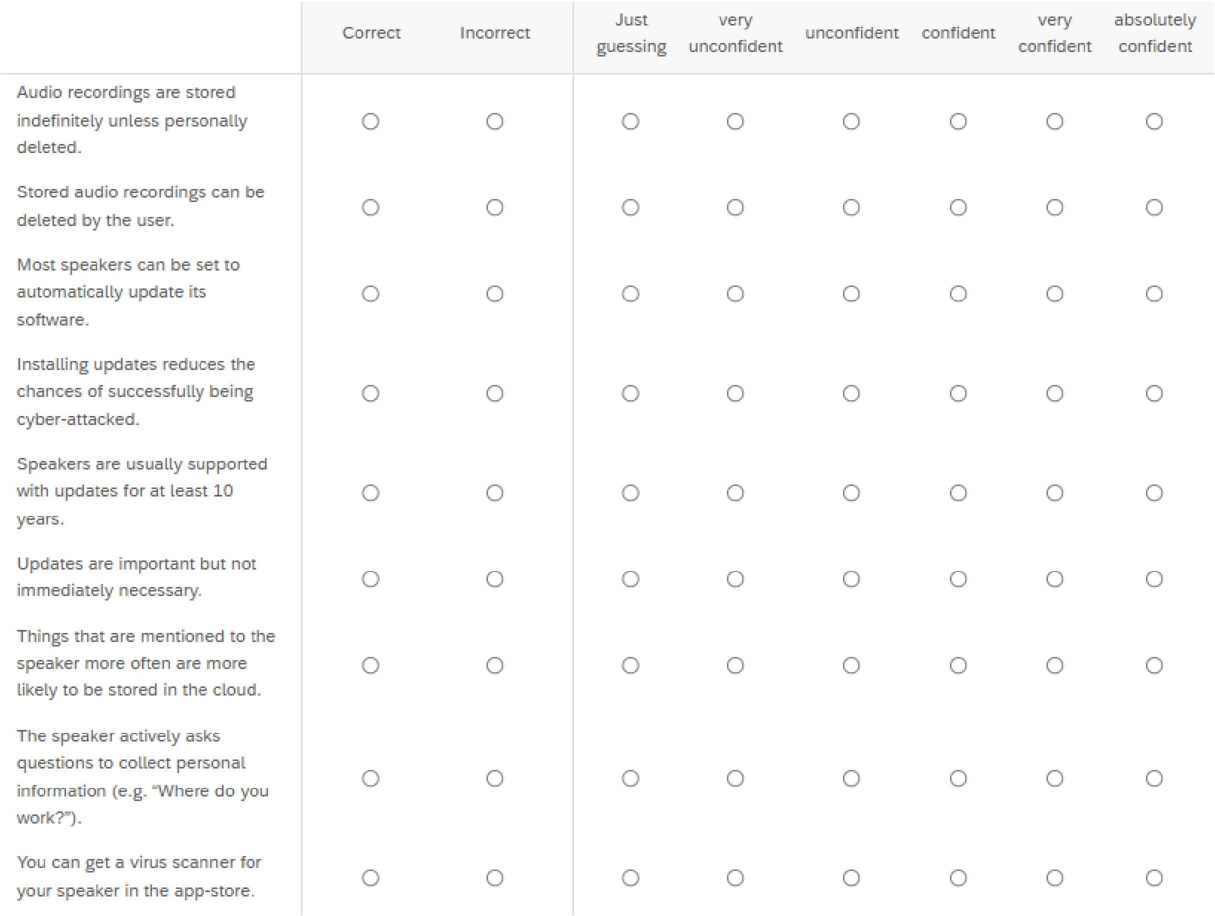
Misconception and knowledge scale

Device security:



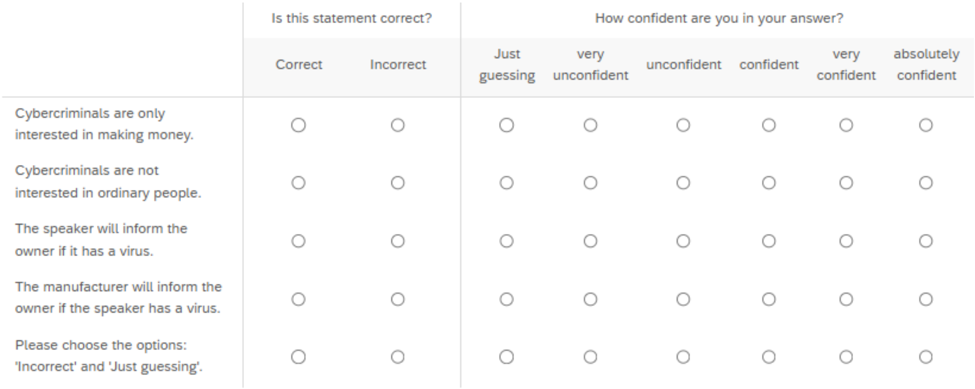
	Is this statement correct?		How confident are you in your answer?					
	Correct	Incorrect	Just guessing	very unconfident	unconfident	confident	very confident	absolutely confident
Guest networks (separate WIFI network) can be created via the Router.	○	○	○	○	○	○	○	○
Guest networks (separate WIFI network) can be used to isolate certain devices from the main network.	○	○	○	○	○	○	○	○
Setting up a guest network (separate WIFI network) to separate certain devices from others is pointless, since all devices are still connected to the same Router.	○	○	○	○	○	○	○	○
The speaker is always listening to hear the wake word (e.g. “Alexa”, “Hey Google”, “Siri” etc.).	○	○	○	○	○	○	○	○
The speaker is always recording and consequently storing any conversations held around it.	○	○	○	○	○	○	○	○
A cybercriminal who gained access to your speaker can turn on the microphone and listen whenever they want.	○	○	○	○	○	○	○	○
A cybercriminal who gained access to your speaker, can gain access to all devices connected in the network, as long as these are not secured otherwise.	○	○	○	○	○	○	○	○
A cybercriminal could take control of other smart devices connected to the speaker, such as a smart thermostat or light.	○	○	○	○	○	○	○	○

Online security:



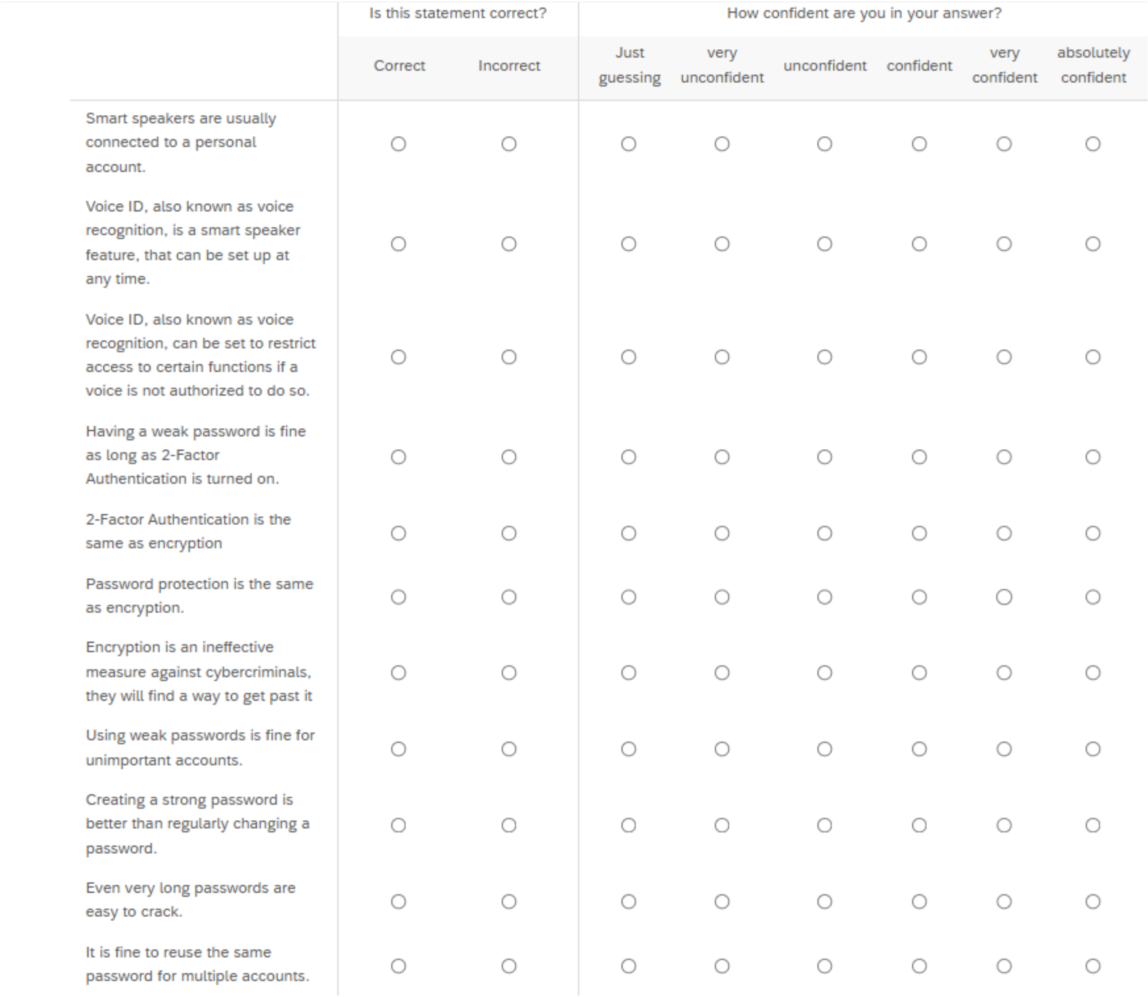
	Correct	Incorrect	Just guessing	very unconfident	unconfident	confident	very confident	absolutely confident
Audio recordings are stored indefinitely unless personally deleted.	○	○	○	○	○	○	○	○
Stored audio recordings can be deleted by the user.	○	○	○	○	○	○	○	○
Most speakers can be set to automatically update its software.	○	○	○	○	○	○	○	○
Installing updates reduces the chances of successfully being cyber-attacked.	○	○	○	○	○	○	○	○
Speakers are usually supported with updates for at least 10 years.	○	○	○	○	○	○	○	○
Updates are important but not immediately necessary.	○	○	○	○	○	○	○	○
Things that are mentioned to the speaker more often are more likely to be stored in the cloud.	○	○	○	○	○	○	○	○
The speaker actively asks questions to collect personal information (e.g. "Where do you work?").	○	○	○	○	○	○	○	○
You can get a virus scanner for your speaker in the app-store.	○	○	○	○	○	○	○	○

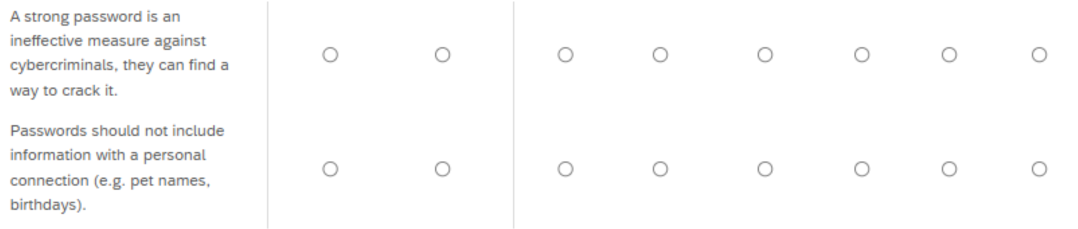
External security:



	Is this statement correct?		How confident are you in your answer?					
	Correct	Incorrect	Just guessing	very unconfident	unconfident	confident	very confident	absolutely confident
Cybercriminals are only interested in making money.	○	○	○	○	○	○	○	○
Cybercriminals are not interested in ordinary people.	○	○	○	○	○	○	○	○
The speaker will inform the owner if it has a virus.	○	○	○	○	○	○	○	○
The manufacturer will inform the owner if the speaker has a virus.	○	○	○	○	○	○	○	○
Please choose the options: 'Incorrect' and 'Just guessing'.	○	○	○	○	○	○	○	○

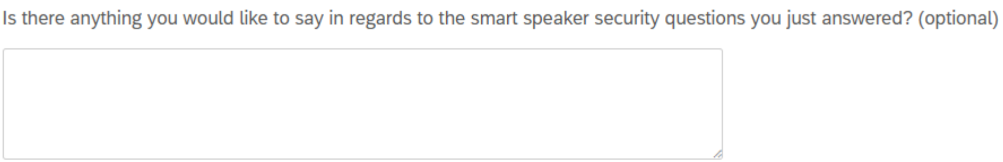
Account security:





	Is this statement correct?		How confident are you in your answer?					
	Correct	Incorrect	Just guessing	very unconfident	unconfident	confident	very confident	absolutely confident
Smart speakers are usually connected to a personal account.	○	○	○	○	○	○	○	○
Voice ID, also known as voice recognition, is a smart speaker feature, that can be set up at any time.	○	○	○	○	○	○	○	○
Voice ID, also known as voice recognition, can be set to restrict access to certain functions if a voice is not authorized to do so.	○	○	○	○	○	○	○	○
Having a weak password is fine as long as 2-Factor Authentication is turned on.	○	○	○	○	○	○	○	○
2-Factor Authentication is the same as encryption	○	○	○	○	○	○	○	○
Password protection is the same as encryption.	○	○	○	○	○	○	○	○
Encryption is an ineffective measure against cybercriminals, they will find a way to get past it	○	○	○	○	○	○	○	○
Using weak passwords is fine for unimportant accounts.	○	○	○	○	○	○	○	○
Creating a strong password is better than regularly changing a password.	○	○	○	○	○	○	○	○
Even very long passwords are easy to crack.	○	○	○	○	○	○	○	○
It is fine to reuse the same password for multiple accounts.	○	○	○	○	○	○	○	○
A strong password is an ineffective measure against cybercriminals, they can find a way to crack it.	○	○	○	○	○	○	○	○
Passwords should not include information with a personal connection (e.g. pet names, birthdays).	○	○	○	○	○	○	○	○

Open feedback:



Is there anything you would like to say in regards to the smart speaker security questions you just answered? (optional)

Appendix I

Participant nationality table

Participant Nationality

Nationality	Frequency
Jordanian	20
Dutch	14
Turkish	13
German	10
Spain	7
Greece	3
Indonesia	3
Russia	3
Poland	2
Ireland	2
Lebanon	2
Romania	2
Egypt	2
United Kingdom	1
United States	1
Canada	1
Cyprus	1

Denmark	1
France	1
Hungary	1
India	1
Iran	1
Italy	1
Japan	1
Austria	1
Palestine	1
Tunisia	1
Bulgaria	1
Bahrain	1
Belgium	1
United Arab Emirates	1

Appendix J

User Specific Protective Behaviours

The 21 items were averaged for non-users (M=2.88, SD=.88, α=.94) and users (M=1.85, SD=.71, α=.89

Non-Users

Since protective behaviour was reworded for and non-users, as such exploratory factor analysis was done separately for both groups. Beginning with non-users, principal axis factor analysis with varimax rotation was conducted on 21 protective behaviour items specific to non-users. Bartlett's test of sphericity was significant, $\chi^2(210) = 941.3$, $p < .001$, indicating sufficient correlations among variables. The Kaiser-Meyer-Olkin (KMO) measure verified sampling adequacy with an overall MSA of 0.81, suggesting the data was appropriate for factor analysis.

Parallel analysis recommended a three-factor solution. The three extracted factors accounted for 49% of the total variance (Proactive Security behaviours = 22%, Threat Recognition = 17%, Avoidance Tactics = 11%) with an item cutoff of 0.3.

Factor Loadings Non-User Protective Behaviour

Item	Physical Security Behaviours	Account Management	Avoidance tactics
I will turn off the smart speaker when I am not using it	.55		
I will unplug the smart speaker when I am not using it	.55		

I will unplug the smart speaker when I am having serious/private conversations	.77		
I will turn off the smart speaker when I am having serious/private conversations	.73		
I will mute the smart speaker's microphone when I am not using it	.60		
I will review the privacy settings of my smart speaker in the provider's (e.g., Alexa or Google) account		.73	
I will review which applications/services have access to my smart speaker		.91	
I will regularly spend time reviewing audio logs in the mobile app and delete those that I want to be deleted.		.51	
I will restrict the amount of data that the device is allowed to collect through the smart speaker's settings		.73	
I will delete my smart speaker recordings		.56	
In the app, I will delete sensitive information that the smart speaker stored about me	.35	.45	
I will speak very quietly around the smart speaker, in case I don't want to be recorded	.52		
I will moderate my language around the smart speaker so that it doesn't record private matters, even accidentally	.57		
I will avoid sensitive/private conversations around the smart speaker	.79		
If I have a visitor I will inform them that I have a smart speaker			.97
If I have a visitor, I will offer to switch the smart speaker off			.72
I will consider where to place the smart speaker so that it is not positioned in areas	.65		

where I typically engaged in sensitive or private conversations.			
I will set a new and difficult password that I only use for my smart speaker.			
I will not write down the password on a piece of paper or share it with house members or visitors.			
I will change the password again after some time			.45
I will not place the smart speaker in a privacy sensitive room like my bedroom		.35	

Factor loadings meaningful interpretation of all three components. The first factor was identified as Proactive Security behaviour included items reflecting pre-emptive behaviours such as unplugging or muting the smart speaker when not in use (α = .9, *M* = 2.85, *SD* = 1.04). The second factor was identified as Avoidance Tactics, reflecting additional steps one takes to protect themselves, such as, deleting their smart speaker recordings(α = .79, *M* = 2.29, *SD* = 1.15). Threat Recognition & Filtering included behaviours that involved identifying and mitigating digital threats such as changing one's password (α = .85, *M* = 3.20, *SD* = 1.07).

As questions 18, and 19 did not load on any factors, they were removed and factor analysis was repeated without them. However, an argument can be made to retain questions 18 and 19 as they relate to smart speaker setup and are theoretically sound. On the other hand, questions 20 and 21 also pertain to smart speaker setup and sufficiently load on factors. After conducting parallel analysis on the updated scale, the scree plot suggested that there were now only two factors being measured. After which, factor analysis on the updated scale with both two factors and three factors. With two factors, the fit became worse, and accounted for 48% of the variance. Whereas, factor analysis with 3 factors on the updated scale yielded a greater proportion of variance explained at 54%.

Users

An exploratory factor analysis (EFA) was conducted on the 21 items assessing protective behaviours for smart speaker users. This included both behavioural protection practices and smart speaker setup questions, which were originally coded as binary (1 = No, 2 = Yes). To align the scaling of these binary items with the 5-point Likert-scale items, responses were recoded: 1 was converted to 1, and 2 was recoded to 5, reflecting stronger protective engagement or no protective engagement.

Initial analysis of all 21 user-protective items revealed poor psychometric properties. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was 0.33, far below the commonly accepted threshold of 0.6 (Kaiser, 1974). Bartlett's Test of Sphericity was significant, $\chi^2(210) = 580.93$, $p < .001$, supporting the factorability of the correlation matrix. However, many items had individual MSA values below 0.5, suggesting they shared little variance with the others. Parallel analysis suggested a two-factor structure and factors were extracted with an item cutoff of 0.3.

Factor Loadings User Protective Behaviour

Item	Physical Security Behaviours	Account and Behaviour moderation
I turned off the smart speaker when I was not using it		
I unplugged the smart speaker when I was not using it	.32	.38

I unplugged the smart speaker when I was having serious/private conversations	.58	
I turned off the smart speaker when I was having serious/private conversations	.42	
I muted the smart speaker's microphone when I was not using it	.33	
I reviewed the privacy settings of my smart speaker in the provider's (e.g., Alexa or Google) account		.96
I reviewed which applications/services have access to my smart speaker		.89
I regularly spend time reviewing audio logs in the mobile app and delete those that I want to be deleted.	.85	
I restricted the amount of data that the device is allowed to collect through the smart speaker's settings	.34	.38
I deleted my smart speaker recordings	1.00	
In the app, I deleted sensitive information that the smart speaker stored about me	.79	
I spoke very quietly around the smart speaker, in case I didn't want to be recorded	.41	
I moderated my language around the smart speaker so that it didn't record private matters, even accidentally		.53
I avoided sensitive/private conversations around the smart speaker	.38	.49
When I had a visitor, I informed them that I have a smart speaker	.50	
When I had a visitor, I offered to switch the smart speaker off	.97	

I placed the smart speaker so that it is wasn’t positioned in areas where I typically engaged in sensitive or private conversations.		
I set a new and difficult password that I only use for my smart speaker.		
I did not write down the password on a piece of paper or share it with house members or visitors.	.33	
I changed the password again after some time		.48
I did not place the smart speaker in a privacy sensitive room like my bedroom		

Items such as question one and the three of the five smart speaker setup questions did not load. Although, this may be due to differences in measuring intended behaviour in non-users and actual behaviour in users. Thus the speaker setup ownership questions were retained due to their theoretical support. The rest of the questions had sufficient factor loadings. Similarly to the other metrics, the user scale explained less variance as opposed to the non-user scale, with the total accounted variance being 42% (Proactive Security Behaviour: 26%, Threat Recognition:16%) The two factors were identified as Proactive security behaviour and Threat Recognition, and are similar to the factors identified for non-users

Appendix K

Factor Analysis of Misconceptions

To explore the latent dimensions underlying misconceptions, an exploratory factor analysis (EFA) was conducted using the 34 misconception items across all four security domains (account, device, online, external). Misconceptions were defined as incorrect answers held with high confidence (ratings of 4–6).

Given the binary nature of the data, a tetrachoric correlation matrix was computed. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy also fell below the recommended minimum threshold of 0.60 (Kaiser, 1974), with an overall (MSA = .57), with individual item MSAs ranging from .33 to .78. Bartlett's test of sphericity was significant, $\chi^2(528) = p < .001$, supporting the factorability of the correlation matrix.

A parallel analysis suggested up to 11 factors, though a 4-factor solution was retained to maintain theoretical consistency with the knowledge scale and to enhance interpretability. The factor analysis used maximum likelihood extraction with a varimax rotation with a 0.3 cutoff, as there are distinct unrelated domains. The four-factor solution explained 47% of the total variance, with acceptable model fit (RMSR = .11; off-diagonal fit = .78). The average item complexity was 1.9, indicating moderate cross-loading across dimensions.

Factor analysis of misconceptions

Item	Smart Speaker Security knowledge	Passwords and Cybercrime	Encryption	Virus protection
Account security				
Smart speakers are usually connected to a personal account.	0.52			0.31
Voice ID, also known as voice recognition, is a smart speaker feature, that can be set up at any time.	0.62	0.31		
Voice ID, also known as voice recognition, can be set to restrict access to certain functions if a voice is not authorized to do so.	0.64			
Having a weak password is fine as long as 2-Factor Authentication is turned on.				0.37
2-Factor Authentication is the same as encryption			0.64	
Password protection is the same as encryption.				0.55
Encryption is an ineffective measure against cybercriminals, they will find a way to get past it			0.36	
Using weak passwords is fine for unimportant accounts.		0.74		0.43
Creating a strong password is better than regularly changing a password.				
Even very long passwords are easy to crack.			0.35	

It is fine to reuse the same password for multiple accounts.		0.60	
A strong password is an ineffective measure against cybercriminals, they can find a way to crack it.			0.54
Passwords should not include information with a personal connection (e.g. pet names, birthdays).		0.75	0.33
Device Security			
Guest networks (separate WIFI network) can be created via the Router.	0.51		0.36
Guest networks (separate WIFI network) can be used to isolate certain devices from the main network.	0.34	-0.37	0.41
Setting up a guest network (separate WIFI network) to separate certain devices from others is pointless, since all devices are still connected to the same Router.	0.57	-0.38	
The speaker is always listening to hear the wake word (e.g. Alexa, Hey google, Siri, etc.).	0.48		0.33
The speaker is always recording and consequently storing any conversations held around it.	0.54		-0.31
A cybercriminal who gained access to your speaker can turn on the microphone and listen whenever they want.	0.45		0.51

A cybercriminal who gained access to your speaker, can gain access to all devices connected in the network, as long as these are not secured otherwise.			0.63	
A cybercriminal could take control of other smart devices connected to the speaker, such as a smart thermostat or light.	0.53		0.36	
Online Security				
Audio recordings are stored indefinitely unless personally deleted.		0.36	0.47	
Stored audio recordings can be deleted by the user.	0.44			0.41
Most speakers can be set to automatically update its software.	0.58			
Installing updates reduces the chances of successfully being cyber-attacked.	0.69			
Speakers are usually supported with updates for at least 10 years.	0.45		0.32	
Updates are important but not immediately necessary.				
Things that are mentioned to the speaker more often are more likely to be stored in the cloud.				
The speaker actively asks questions to collect personal information (e.g. ‘’Where do you work?”).		-0.42		
You can get a virus scanner for your speaker in the app-store.	0.32			0.39
External Security				

Cybercriminals are only interested in making money.	0.47		
Cybercriminals are not interested in ordinary people.	0.61		
The speaker will inform the owner if it has a virus.	0.51	0.33	0.44
The manufacturer will inform the owner if the speaker has a virus.			0.97

All the items loaded sufficiently on at least one factor. Additionally, some items loaded on multiple factors suggesting that these domains may not be as distinct as expected. Interestingly, some items loaded negatively on some factors indicating a particular misconception may influence the other misconceptions held. Interestingly, the factors were different from the factors extracted for the knowledge scale. The first factor identified revolved around general smart speaker knowledge, or in this case, misconceptions. This factor contained questions such as "Voice ID, also known as voice recognition, is a smart speaker feature, that can be set up at any time.". The next factor that was identified is directly related to passwords and cybercrime, with questions such as "Using weak passwords is fine for unimportant accounts", and "Cybercriminals are not interested in ordinary people". The third factor that was extracted was labelled Encryption and had questions that revolved around password, network, and data encryption, with questions like "2-Factor Authentication is the same as encryption". The final factor loaded very highly on the following question "The manufacturer will inform the owner if the speaker has a virus.", it also loaded on all the other virus questions in the scale.

Appendix L

R code

```
install.packages("psych")
install.packages("tidyverse")
install.packages("psychTools")
install.packages("janitor")
install.packages("tidyr")
install.packages("ggplot2")
install.packages("broom")
install.packages("dplyr")
install.packages("readr")
library(tidyverse)
library(janitor)
library(broom)
library(tidyr)
library(psych)
library(dplyr)
library(psychTools)
library(ggplot2)
#set working directory and load the file
setwd("C:/Users/karam/Downloads")
data <- read_csv("Exploring+factors+surrounding+smart+speaker+security+protective+behaviour_May+5,+2025_06.52.csv")
#Remove start date, end date, status, and IP address
cleaned_data <- data[ , -c(1:4)]
#Remove progress, recorded date, response ID, recipient first name, last name, email and external reference
cleaned_data <- cleaned_data[, -c(1,4,5,6,7,8,9)]
#Remove longitude, latitude, distribution channel
cleaned_data <- cleaned_data[ , -c(3:5)]
#Save question text row
question_text <- cleaned_data[1, ]
#Remove participants who did not accept or had N/A responses for informed consent
cleaned_data <- cleaned_data %>%
  filter(row_number() == 1 |(Informed_Consent_Sig == 1 &!is.na(Informed_Consent_Sig)))
#Removing N/A responses for gender
cleaned_data <- cleaned_data %>%
  filter(row_number() == 1 | !is.na(Gender))
```

```
#Removing people who failed the attention check for knowledge
cleaned_data <- cleaned_data %>%
  filter(row_number() == 1 |(attentioncheck_knowledge_1_correctness == 2 &!is.na(attentioncheck_knowledge_1_correctness)))
#Removing people who failed the attention check for confidence
cleaned_data <- cleaned_data %>%
  filter(row_number() == 1 |(attentioncheck_knowledge_2_confidence == 1 &!is.na(attentioncheck_knowledge_2_confidence)))
#Removing failed user attention checks
cleaned_data <- cleaned_data %>%
  filter(row_number() == 1 | Attention_check_user == 5 | is.na(Attention_check_user))

# removing failed non_user attention checks
cleaned_data <- cleaned_data %>%
  filter(row_number() == 1 | Attention_check_non == 1 | is.na(Attention_check_non))
#removing the last na responses
cleaned_data <- cleaned_data %>%
  filter(!is.na(online_security_1_1_correctness))
#count total responses after all attention checks
n_responses <- nrow(cleaned_data) - 1
print(paste("Total valid responses:", n_responses))
#Get gender information
table(cleaned_data$Gender[-1])
#Get dutch and german nationalities
table(cleaned_data$Nationality[-1])
#create data without questions
data_no_header <- cleaned_data[-1, ]
#get other nationalities
nationality_counts <- table(data_no_header$Nationality_3_TEXT)  # Change 'Country' if needed
nationality_counts_sorted <- sort(nationality_counts, decreasing = TRUE)
# Display
nationality_counts_sorted
#get age information
age_vals <- as.numeric(data_no_header$Age_1)
age_mean <- round(mean(age_vals, na.rm = TRUE), 2)
age_sd <- round(sd(age_vals, na.rm = TRUE), 2)
age_range <- range(age_vals, na.rm = TRUE)

cat("Age: Range =", age_range[1], "to", age_range[2],
    " | Mean =", age_mean, " | SD =", age_sd, "\n")
```

```
#Get education demographics
education_counts <- table(data_no_header$Education)
print(education_counts)
#find median duration
duration_vals <- as.numeric(pull(cleaned_data[-1, 1]))
median_duration <- median(duration_vals, na.rm = TRUE)
print(paste("Median of first column:", median_duration))
# Creating trust_items to run factor analysis
trust_items <- data_no_header %>%
  select(starts_with("Manufacturer_"))
#Convert to numeric
trust_items <- trust_items %>% mutate(across(everything(), as.numeric))
# Check Bartlett's test of sphericity
cortest.bartlett(cor(trust_items, use = "pairwise.complete.obs"), n = nrow(trust_items))
# Check KMO (sampling adequacy)
KMO(trust_items)
# Check number of factors
fa.parallel(trust_items, fa = "fa", n.iter = 101, show.legend = FALSE)
# Factor analysis on trust
trust_fa <- fa(trust_items, nfactors = 2, rotate = "varimax", fm = "ml")
print(trust_fa, cut = 0.3)

trust_data <- trust_items %>%
  mutate(across(everything(), as.numeric))  # Ensure all columns are numeric

# Mean (averaged across items per participant)
trust_data$trust_mean <- rowMeans(trust_data, na.rm = TRUE)

# Standard deviation of the means
trust_sd <- sd(trust_data$trust_mean, na.rm = TRUE)
trust_mean <- mean(trust_data$trust_mean, na.rm = TRUE)

# Cronbach's alpha
trust_alpha <- psych::alpha(trust_data)

# Print results
cat("Manufacturer Trust Scale:\n")
cat("Mean =", round(trust_mean, 2), "\n")
cat("SD =", round(trust_sd, 2), "\n")
cat("Cronbach's ?? =", round(trust_alpha$total$raw_alpha, 2), "\n")

# Factor 1 (General Trust): Items 1, 2, 4, 5
# Factor 2 (Technical Trust): Items 6, 7

general_trust <- trust_items %>% select(Manufacturer_Trust_1, Manufacturer_Trust_2, Manufacturer_Trust_4, Manufacturer_Trust_5)
technical_trust <- trust_items %>% select(Manufacturer_Trust_6, Manufacturer_Trust_7)

# Subscale scores
general_trust$general_mean <- rowMeans(general_trust, na.rm = TRUE)
technical_trust$technical_mean <- rowMeans(technical_trust, na.rm = TRUE)

# Reliability
alpha_general <- psych::alpha(general_trust[, 1:5])
alpha_technical <- psych::alpha(technical_trust[, 1:2])

cat("\nGeneral Trust Subscale:\n")
cat("Cronbach's ?? =", round(alpha_general$total$raw_alpha, 2), "\n")

cat("\nTechnical Trust Subscale:\n")
cat("Cronbach's ?? =", round(alpha_technical$total$raw_alpha, 2), "\n")
# General_Trust subscale (Items 1, 2, 4, 5)
general_trust <- trust_items %>%
  select(Manufacturer_Trust_1, Manufacturer_Trust_2, Manufacturer_Trust_3, Manufacturer_Trust_4, Manufacturer_Trust_5)

general_trust$general_mean <- rowMeans(general_trust, na.rm = TRUE)
general_mean <- mean(general_trust$general_mean, na.rm = TRUE)
general_sd <- sd(general_trust$general_mean, na.rm = TRUE)

# Technical Trust subscale (Items 6, 7)
technical_trust <- trust_items %>%
  select(Manufacturer_Trust_6, Manufacturer_Trust_7)

technical_trust$technical_mean <- rowMeans(technical_trust, na.rm = TRUE)
technical_mean <- mean(technical_trust$technical_mean, na.rm = TRUE)
technical_sd <- sd(technical_trust$technical_mean, na.rm = TRUE)
```

```
# Print results
cat("\nGeneral Trust Subscale:\n")
cat("Mean =", round(general_mean, 2), "\n")
cat("SD =", round(general_sd, 2), "\n")

cat("\nTechnical Trust Subscale:\n")
cat("Mean =", round(technical_mean, 2), "\n")
cat("SD =", round(technical_sd, 2), "\n")

## ?????? new answer keys
???????????????????????????????????????????????????????????????????????????????????????????????????????????????????
key_account  <- setNames(
  c(2,2,2,2,2,2,2,2,1,2,2,2,1),        # where '2' = False/Incorrect
  paste0("account_security_", 1:13, "_1_correctness")
)

key_device   <- setNames(
  c(2,2,1,2,1,2,2,2),
  paste0("device_security_", 1:8, "_1_correctness")
)

key_online   <- setNames(
  c(2,2,2,2,1,1,1,1,1),
  paste0("online_security_", 1:9, "_1_correctness")
)

key_external <- setNames(
  c(2,2,2,2),                          # all four are "incorrect" items
  paste0("external_security_", 1:4, "_1_correctness")
)
## helper stays the same ---------------------------------
safe_numeric <- function(x){
  if(is.factor(x)) as.numeric(as.character(x)) else as.numeric(x)
}

score_block <- function(df, key_vec){
  cols <- names(key_vec)
  mutate(df,
        across(all_of(cols),
               ~ safe_numeric(.) == key_vec[cur_column()]))
}

## pipeline ------------------------------------------------
# ?????? Score each knowledge block and create section totals + grand average ??????
data_scored <- data_no_header %>%
  ## 1. convert all *_correctness columns to numeric safely
  mutate(across(matches("_correctness$"), ~ as.numeric(.))) %>%

  ## 2. score every block against its answer key
  score_block(key_account)  %>%
  score_block(key_device)   %>%
  score_block(key_online)   %>%
  score_block(key_external) %>%

  ## 3. create counts of correct answers per section
  rowwise() %>%
  mutate(
    acc_know = sum(c_across(all_of(names(key_account))),  na.rm = TRUE),  # 13 items
    dev_know = sum(c_across(all_of(names(key_device))),   na.rm = TRUE),  #  8 items
    onl_know = sum(c_across(all_of(names(key_online))),   na.rm = TRUE),  #  9 items
    ext_know = sum(c_across(all_of(names(key_external))), na.rm = TRUE),  #  4 items



    ## total correct answers across ALL sections
    knowledge_total = acc_know + dev_know + onl_know + ext_know,

    ## proportion correct (grand average)
    knowledge_avg   = knowledge_total / (length(key_account) +
                                         length(key_device)  +
                                         length(key_online)  +
                                         length(key_external))
  ) %>%
  ungroup()
```

```
## quick check: min / max possible scores
data_scored %>%
  summarise(across(
    c(acc_know, dev_know, onl_know, ext_know, knowledge_total, knowledge_avg),
    list(min = min, max = max), na.rm = TRUE))

# Combine all scored knowledge items
all_knowledge_items <- data_scored %>%
  select(all_of(c(names(key_account), names(key_device), names(key_online), names(key_external))))

# Bartlett's Test
cortest.bartlett(cor(all_knowledge_items, use = "pairwise.complete.obs"),
                 n = nrow(all_knowledge_items))

# KMO
KMO(all_knowledge_items)

# Parallel analysis
fa.parallel(all_knowledge_items, fa = "fa", n.iter = 100, show.legend = FALSE)

# EFA (start with 4 factors based on theory)
knowledge_efa <- fa(all_knowledge_items, nfactors = 4, rotate = "oblimin", fm = "ml")
print(knowledge_efa, cut = 0.3)

# Overall mean and SD
data_scored$knowledge_mean <- rowMeans(all_knowledge_items, na.rm = TRUE)

# Cronbach's alpha for full scale
knowledge_alpha <- psych::alpha(all_knowledge_items)

# Output
cat("Knowledge Scale (All Items):\n")
cat("Mean =", round(mean(data_scored$knowledge_mean, na.rm = TRUE), 2), "\n")
cat("SD =", sd(data_scored$knowledge_mean, na.rm = TRUE), 2, "\n")
cat("Cronbach's alpha =", round(knowledge_alpha$total$raw_alpha, 2), "\n")

# Combine all scored knowledge items into one data frame
knowledge_items_all <- data_scored %>%
  select(all_of(c(names(key_account), names(key_device), names(key_online), names(key_external))))


# Calculate total correct per participant

data_scored$knowledge_total <- rowSums(knowledge_items_all, na.rm = TRUE)


# Calculate average score per person (proportion correct)

data_scored$knowledge_avg <- rowMeans(knowledge_items_all, na.rm = TRUE)


# Descriptive statistics

total_mean <- mean(data_scored$knowledge_total, na.rm = TRUE)

total_sd <- sd(data_scored$knowledge_total, na.rm = TRUE)


avg_mean <- mean(data_scored$knowledge_avg, na.rm = TRUE)

avg_sd <- sd(data_scored$knowledge_avg, na.rm = TRUE)


# Print results

cat("Knowledge Total Score:\n")

cat("Mean =", round(total_mean, 2), "| SD =", round(total_sd, 2), "\n")


cat("Knowledge Average Score:\n")

cat("Mean =", round(avg_mean, 2), "| SD =", round(avg_sd, 2), "\n")



##
???????????????????????????????????????????????????????????????????????????????????????????????????????????????????????????????????????????????????
??????????????????????????????????????????????????????????????

##  4.  Run EFA on the *scored* items  (one block at a time)

##
???????????????????????????????????????????????????????????????????????????????????????????????????????????????????????????????????????????????????
??????????????????????????????????????????????????????????????

library(psych)

#Account security

# Bartlett's Test

cortest.bartlett(cor(data_scored[ , names(key_account)], use = "pairwise.complete.obs"), n = 106)


# KMO

KMO(data_scored[ , names(key_account)])

fa.parallel(data_scored[ , names(key_account)], fa = "fa", n.iter = 100, show.legend = FALSE)

efa_account <- fa(data_scored[ , names(key_account)],

                  nfactors = 3, rotate = "oblimin", fm = "minres")
```

```
print(efa_account, cut = .30)
#Device security
cortest.bartlett(cor(data_scored[ , names(key_device)], use = "pairwise.complete.obs"), n = 106)

# KMO
KMO(data_scored[ , names(key_device)])
fa.parallel(data_scored[ , names(key_device)], fa = "fa", n.iter = 100, show.legend = FALSE)
efa_device  <- fa(data_scored[ , names(key_device)],
                  nfactors = 3, rotate = "oblimin", fm = "minres")
print(efa_device,  cut = .30)
#online security
cortest.bartlett(cor(data_scored[ , names(key_online)], use = "pairwise.complete.obs"), n = 106)

# KMO
KMO(data_scored[ , names(key_online)])
fa.parallel(data_scored[ , names(key_online)], fa = "fa", n.iter = 100, show.legend = FALSE)
efa_online  <- fa(data_scored[ , names(key_online)],
                  nfactors = 2, rotate = "oblimin", fm = "minres")
print(efa_online,  cut = .30)
#External security
cortest.bartlett(cor(data_scored[ , names(key_device)], use = "pairwise.complete.obs"), n = 106)

# KMO
KMO(data_scored[ , names(key_external)])
fa.parallel(data_scored[ , names(key_external)], fa = "fa", n.iter = 100, show.legend = FALSE)
efa_external<- fa(data_scored[ , names(key_external)],
                  nfactors = 1, rotate = "oblimin", fm = "minres")
print(efa_external,cut = .30)



#Split users and non-users
users <- data_no_header %>% filter(ownership == 1)
non_users <- data_no_header %>% filter(ownership == 0)

# For USERS: remove NON-user items
users_clean <- users %>%
  select(-matches("^protective_nonuse"))
```

```
# For NON-USERS: remove USER items
non_users_clean <- non_users %>%
  select(-matches("^protective_user"))
non_users_clean <- non_users_clean %>%
  select(-matches("^Speaker_Setup"))

# Select non-user protective items
nonuser_protective_items <- non_users_clean %>%
  select(starts_with("protective_nonuse")) %>%
  mutate(across(everything(), as.numeric))
#recode binary user items
users_clean <- users_clean %>%
  mutate(across(starts_with("Speaker_Setup"), ~ case_when(
    as.character(.) == "2" ~ 5,
    as.character(.) == "1" ~ 1,
    TRUE ~ NA_real_
  )))

#Select user protective items
user_protective_items <- users_clean %>%
  select(starts_with("Speaker_Setup"), starts_with("protective_user")) %>%
  mutate(across(everything(), as.numeric))
# Tests for non-users
# Bartlett's test
cortest.bartlett(cor(nonuser_protective_items, use = "pairwise.complete.obs"),
                 n = nrow(nonuser_protective_items))

# KMO measure
KMO(nonuser_protective_items)
#Paralell analysis
fa.parallel(nonuser_protective_items, fa = "fa", n.iter = 75, show.legend = FALSE)
# Factor analysis with 3 factors
nonuser_fa_3 <- fa(nonuser_protective_items, nfactors = 3, rotate = "oblimin", fm = "ml")
print(nonuser_fa_3, cut = 0.3)

# Factor 1: Proactive Security Strategies
f1_nonuser <- c("Protective_nonuse_1", "Protective_nonuse_2", "Protective_nonuse_3",
                "Protective_nonuse_4", "Protective_nonuse_5", "Protective_nonuse_11", "Protective_nonuse_12",
        "Protective_nonuse_13", "Protective_nonuse_14", "Protective_nonuse_17")
nonuser_protective_items$factor1 <- rowMeans(nonuser_protective_items[, f1_nonuser], na.rm = TRUE)

# Factor 2: Avoidance Tactics
f2_nonuser <- c("Protective_nonuse_15", "Protective_nonuse_16", "Protective_nonuse_20")
nonuser_protective_items$factor2 <- rowMeans(nonuser_protective_items[, f2_nonuser], na.rm = TRUE)

# Factor 3: Threat Recognition & Filtering
f3_nonuser <- c("Protective_nonuse_6", "Protective_nonuse_7", "Protective_nonuse_8",
        "Protective_nonuse_9", "Protective_nonuse_10", "Protective_nonuse_21")
nonuser_protective_items$factor3 <- rowMeans(nonuser_protective_items[, f3_nonuser], na.rm = TRUE)
# Factor 1
f1_mean <- mean(nonuser_protective_items$factor1, na.rm = TRUE)
f1_sd <- sd(nonuser_protective_items$factor1, na.rm = TRUE)
alpha_f1 <- psych::alpha(nonuser_protective_items[, f1_nonuser])

# Factor 2
f2_mean <- mean(nonuser_protective_items$factor2, na.rm = TRUE)
f2_sd <- sd(nonuser_protective_items$factor2, na.rm = TRUE)
alpha_f2 <- psych::alpha(nonuser_protective_items[, f2_nonuser])

# Factor 3
f3_mean <- mean(nonuser_protective_items$factor3, na.rm = TRUE)
f3_sd <- sd(nonuser_protective_items$factor3, na.rm = TRUE)
alpha_f3 <- psych::alpha(nonuser_protective_items[, f3_nonuser])
#Print results
cat("\nFactor 1: Proactive Security Strategies\n")
cat("Mean =", round(f1_mean, 2), " SD =", round(f1_sd, 2),
  " ?? =", round(alpha_f1$total$raw_alpha, 2), "\n")

cat("\nFactor 2: Avoidance Tactics\n")
cat("Mean =", round(f2_mean, 2), " SD =", round(f2_sd, 2),
  " ?? =", round(alpha_f2$total$raw_alpha, 2), "\n")

cat("\nFactor 3: Threat Recognition & Filtering\n")
cat("Mean =", round(f3_mean, 2), " SD =", round(f3_sd, 2),
  " ?? =", round(alpha_f3$total$raw_alpha, 2), "\n")
```

```
# Compute row-wise mean across all items (no exclusions)
nonuser_protective_items$nonuser_all_mean <- rowMeans(nonuser_protective_items, na.rm = TRUE)

# Reliability across all 21 items
nonuser_all_alpha <- psych::alpha(nonuser_protective_items)

# Output
cat("Non-User Protective Behaviour (All Items):\n")
cat("Mean =", round(mean(nonuser_protective_items$nonuser_all_mean, na.rm = TRUE), 2), "\n")
cat("SD =", round(sd(nonuser_protective_items$nonuser_all_mean, na.rm = TRUE), 2), "\n")
cat("Cronbach's alpha =", round(nonuser_all_alpha$total$raw_alpha, 2), "\n")



# Step 2: Run Bartlett's test and KMO again
cortest.bartlett(cor(nonuser_items_cleaned, use = "pairwise.complete.obs"), n = nrow(nonuser_items_cleaned))
KMO(nonuser_items_cleaned)

# Step 3: Parallel analysis to check number of factors again
fa.parallel(nonuser_items_cleaned, fa = "fa", n.iter = 100, show.legend = FALSE)

# Step 4: Run EFA with 3 factors
nonuser_fa_3_cleaned <- fa(nonuser_items_cleaned, nfactors = 3, rotate = "oblimin", fm = "ml")
print(nonuser_fa_3_cleaned, cut = 0.3)



#Factor analysis for users protective behavior
# Bartlett's Test
cortest.bartlett(cor(user_protective_items, use = "pairwise.complete.obs"),
                 n = nrow(user_protective_items))

# KMO Measure
KMO(user_protective_items)

# Parallel analysis
fa.parallel(user_protective_items, fa = "fa", n.iter = 100, show.legend = FALSE)
#factor analysis
user_fa_2 <- fa(user_protective_items, nfactors = 2, rotate = "oblimin", fm = "ml")
print(user_fa_2, cut = 0.3)
```

```
# Compute mean across all user protective items
user_protective_items$user_all_mean <- rowMeans(user_protective_items, na.rm = TRUE)

# Reliability across all items
user_all_alpha <- psych::alpha(user_protective_items)

# Output
cat("User Protective Behaviour (All Items):\n")
cat("Mean =", round(mean(user_protective_items$user_all_mean, na.rm = TRUE), 2), "\n")
cat("SD =", round(sd(user_protective_items$user_all_mean, na.rm = TRUE), 2), "\n")
cat("Cronbach's alpha =", round(user_all_alpha$total$raw_alpha, 2), "\n")



# First, make sure they're in the same order
data_scored$trust_mean <- trust_data$trust_mean
#ALSO HAVE TO CODE MISCONCEPTION SCORES
# Users only: Compute total behaviour score
users_clean <- users_clean %>%
  mutate(across(starts_with("protective_user"), as.numeric),
         across(starts_with("Speaker_Setup"), as.numeric)) %>%
  mutate(protective_behaviour_total = rowMeans(select(., starts_with("protective_user"), starts_with("Speaker_Setup")), na.rm = TRUE))

non_users_clean <- non_users_clean %>%
  mutate(across(starts_with("protective_nonuse_"), as.numeric)) %>%
  mutate(protective_behaviour_total = rowMeans(select(., starts_with("protective_nonuse_")), na.rm = TRUE))
# Add knowledge and trust to users
# Add knowledge and trust to users
users_clean <- users_clean %>%
  mutate(knowledge_total = data_scored$knowledge_total[match(rownames(users_clean), rownames(data_scored))],
         trust_mean = data_scored$trust_mean[match(rownames(users_clean), rownames(data_scored))],
         Age_1 = data_scored$Age_1[match(rownames(users_clean), rownames(data_scored))],
         Gender = data_scored$Gender[match(rownames(users_clean), rownames(data_scored))])

# Same for non-users
non_users_clean <- non_users_clean %>%
  mutate(knowledge_total = data_scored$knowledge_total[match(rownames(non_users_clean), rownames(data_scored))],
        trust_mean = data_scored$trust_mean[match(rownames(non_users_clean), rownames(data_scored))],
        Age_1 = data_scored$Age_1[match(rownames(non_users_clean), rownames(data_scored))],
        Gender = data_scored$Gender[match(rownames(non_users_clean), rownames(data_scored))])

combined_data <- bind_rows(
  users_clean %>% mutate(group = "user"),
  non_users_clean %>% mutate(group = "nonuser")
)
combined_data$Age_1 <- as.numeric(as.character(combined_data$Age_1))
combined_data$ownership <- as.numeric(as.character(combined_data$ownership))
# Return 1 if the participant was wrong AND confident (confidence 4,5,6), else 0
misconception_flag <- function(correct_col, conf_col) {
  as.numeric((correct_col == 0 | correct_col == 2) & conf_col %in% c(4, 5, 6))
}
# Add misconception flags for all relevant items
data_misconceptions <- data_scored %>%
  mutate(across(ends_with("_correctness"), as.numeric),
        across(ends_with("_confidence"), as.numeric)) %>%
  rowwise() %>%
  mutate(
    misconceptions_account = sum(c_across(paste0("account_security_", 1:13, "_1_correctness")) == 0 &
                                c_across(paste0("account_security_", 1:13, "_2_confidence")) %in% 4:6, na.rm = TRUE),

    misconceptions_device = sum(c_across(paste0("device_security_", 1:8, "_1_correctness")) == 0 &
                               c_across(paste0("device_security_", 1:8, "_2_confidence")) %in% 4:6, na.rm = TRUE),

    misconceptions_online = sum(c_across(paste0("online_security_", 1:9, "_1_correctness")) == 0 &
                               c_across(paste0("online_security_", 1:9, "_2_confidence")) %in% 4:6, na.rm = TRUE),

    misconceptions_external = sum(c_across(paste0("external_security_", 1:4, "_1_correctness")) == 0 &
                                 c_across(paste0("external_security_", 1:4, "_2_confidence")) %in% 4:6, na.rm = TRUE)
  ) %>%
  ungroup() %>%
  mutate(misconceptions_total = misconceptions_account +
          misconceptions_device +
          misconceptions_online +
          misconceptions_external)
  combined_data$misconceptions_total <- data_misconceptions$misconceptions_total[match(rownames(combined_data),
rownames(data_misconceptions))]


#Linear model for H1

model_h1_misconceptions <- lm(protective_behaviour_total ~ misconceptions_total + ownership + Age_1 + Gender, data = combined_data)

  summary(model_h1_misconceptions)

#Trying without ownership as a variable

  model_h1_misconceptionsnoowner <- lm(protective_behaviour_total ~ misconceptions_total + Age_1 + Gender, data = combined_data)

  summary(model_h1_misconceptionsnoowner)

#Graph

  library(ggplot2)

  ggplot(combined_data, aes(x = misconceptions_total, y = protective_behaviour_total)) +

    geom_point() +

    geom_smooth(method = "lm", se = TRUE, color = "blue") +

    labs(x = "Misconceptions (total)", y = "Protective Behaviour", title = "Regression: Misconceptions vs Behaviour")


#PEARSONS COR TEST

  cor.test(combined_data$misconceptions_total, combined_data$protective_behaviour_total)


#H2 Independent samples T-Test

t.test(knowledge_total ~ ownership, data = combined_data)

wilcox.test(knowledge_total ~ ownership, data = combined_data)


#H3: Independent Samples T-Test

t.test(protective_behaviour_total ~ ownership, data = combined_data)
#ANCOVA MODEL CONTROLLING FOR OTHER FACTORS

model_h3 <- aov(protective_behaviour_total ~ ownership + knowledge_total + trust_mean, data = combined_data)

summary(model_h3)

model_h3_anova <- aov(protective_behaviour_total ~ ownership, data = combined_data)

summary(model_h3_anova)

#H4 Binary Logistic Regression


model_h4 <- glm(ownership ~ general_mean + technical_mean, data = combined_data, family = binomial)

summary(model_h4)

model_h4a <- glm(ownership ~ general_mean, data = combined_data, family = binomial)

summary(model_h4a)
```

```
model_h4b <- glm(ownership ~ trust_mean + knowledge_total + Age_1, data = combined_data, family = binomial)
summary(model_h4b)

#H5 Trust and protective behaviour linear regression
model_h5 <- lm(protective_behaviour_total ~ trust_mean + Age_1 + Gender, data = combined_data)
summary(model_h5)
#Pearsons test
cor.test(combined_data$trust_mean, combined_data$protective_behaviour_total)


#CHECKING ASSUMPTIONS OF NORMALITY
# Histogram
hist(combined_data$misconceptions_total, main = "Misconceptions Total", xlab = "Score")
hist(combined_data$protective_behaviour_total, main = "Protective Behaviour", xlab = "Score")

# Q-Q plot
qqnorm(combined_data$misconceptions_total); qqline(combined_data$misconceptions_total)
qqnorm(combined_data$protective_behaviour_total); qqline(combined_data$protective_behaviour_total)

# Shapiro-Wilk Test
shapiro.test(combined_data$misconceptions_total)
shapiro.test(combined_data$protective_behaviour_total)
shapiro.test(combined_data$knowledge_total)

#Checking if residulas are normally distributed
model <- lm(protective_behaviour_total ~ misconceptions_total + ownership + Age_1 + Gender, data = combined_data)

# Shapiro-Wilk test on residuals
shapiro.test(residuals(model))

# Q-Q plot
qqnorm(residuals(model)); qqline(residuals(model))

#Homoscedacity
model <- lm(protective_behaviour_total ~ misconceptions_total + ownership + Age_1 + Gender, data = combined_data)

# Residual plot
plot(model$fitted.values, model$residuals, main = "Residuals vs Fitted",
     xlab = "Fitted values", ylab = "Residuals")
abline(h = 0, col = "red")

# Q-Q plot of residuals
qqnorm(residuals(model)); qqline(residuals(model))

# Breusch-Pagan test for heteroscedasticity
library(lmtest)
bptest(model)
#multicolinearity
library(car)
vif(model)

#EXPLORATORY ANALYSIS

# Split dataset
# EXPLORATORY ANALYSIS

# Recalculate trust subscales directly into combined_data
combined_data$general_mean <- rowMeans(select(combined_data, Manufacturer_Trust_1, Manufacturer_Trust_2, Manufacturer_Trust_3, Manufacturer_Trust_4, Manufacturer_Trust_5), na.rm = TRUE)
combined_data$technical_mean <- rowMeans(select(combined_data, Manufacturer_Trust_6, Manufacturer_Trust_7), na.rm = TRUE)
# Calculate trust subscales from trust_items
general_mean <- rowMeans(trust_items[, c("Manufacturer_Trust_1", "Manufacturer_Trust_2", "Manufacturer_Trust_3",
                                         "Manufacturer_Trust_4", "Manufacturer_Trust_5")], na.rm = TRUE)

technical_mean <- rowMeans(trust_items[, c("Manufacturer_Trust_6", "Manufacturer_Trust_7")], na.rm = TRUE)

# Assign these to combined_data using matching row names
combined_data$general_mean <- general_mean[match(rownames(combined_data), rownames(data_no_header))]
combined_data$technical_mean <- technical_mean[match(rownames(combined_data), rownames(data_no_header))]

# Regressions for users and non-users
users_data <- combined_data %>% filter(ownership == 1)
nonusers_data <- combined_data %>% filter(ownership == 0)

# USERS
model_users <- lm(protective_behaviour_total ~ misconceptions_total + general_mean + technical_mean +
                knowledge_total + Age_1 + Gender, data = users_data)
summary(model_users)


# NON-USERS
model_nonusers <- lm(protective_behaviour_total ~ misconceptions_total + general_mean + technical_mean +
                  knowledge_total + Age_1 + Gender, data = nonusers_data)
summary(model_nonusers)


# Trust-only models
# General trust only
model_trust_general <- lm(protective_behaviour_total ~ general_mean + Age_1 + Gender, data = combined_data)
summary(model_trust_general)


# Technical trust only
model_trust_technical <- lm(protective_behaviour_total ~ technical_mean + Age_1 + Gender, data = combined_data)
summary(model_trust_technical)


# Combined trust model
model_trust_both <- lm(protective_behaviour_total ~ general_mean + technical_mean + Age_1 + Gender, data = combined_data)
summary(model_trust_both)


# Ensure ownership is numeric (0 = non-user, 1 = user)
combined_data$ownership <- as.numeric(as.character(combined_data$ownership))


# Pearson correlation
cor.test(combined_data$ownership, combined_data$misconceptions_total)
# Compare misconceptions between users and non-users
t.test(misconceptions_total ~ ownership, data = combined_data)


# Non-parametric alternative
wilcox.test(misconceptions_total ~ ownership, data = combined_data)


lm(protective_behaviour_total ~ misconceptions_total + general_mean + technical_mean + knowledge_total + ownership + Age_1 + Gender, data
= combined_data)


# Make sure ownership is a factor
combined_data$ownership <- as.factor(combined_data$ownership)
```

```
# Create interaction terms

combined_data$misconceptions_x_ownership <- combined_data$misconceptions_total * as.numeric(combined_data$ownership)

combined_data$trust1_x_ownership <- combined_data$general_mean * as.numeric(combined_data$ownership)

combined_data$trust2_x_ownership <- combined_data$technical_mean * as.numeric(combined_data$ownership)


model_moderated <- lm(protective_behaviour_total ~ misconceptions_total + general_mean + technical_mean +

                      ownership + misconceptions_x_ownership + trust1_x_ownership + trust2_x_ownership +

                      Age_1 + Gender, data = combined_data)

summary(model_moderated)


# Misconceptions only

model_misconceptions_only <- lm(protective_behaviour_total ~ misconceptions_total + Age_1 + Gender, data = combined_data)

summary(model_misconceptions_only)


# General Trust only

model_trust1_only <- lm(protective_behaviour_total ~ general_mean + Age_1 + Gender, data = combined_data)

summary(model_trust1_only)


# Technical Trust only

model_trust2_only <- lm(protective_behaviour_total ~ technical_mean + Age_1 + Gender, data = combined_data)

summary(model_trust2_only)


# Misconceptions + Ownership

model_misconceptions_own <- lm(protective_behaviour_total ~ misconceptions_total + ownership + Age_1 + Gender, data = combined_data)

summary(model_misconceptions_own)


# General Trust + Ownership

model_trust1_own <- lm(protective_behaviour_total ~ general_mean + ownership + Age_1 + Gender, data = combined_data)

summary(model_trust1_own)


# Technical Trust + Ownership

model_trust2_own <- lm(protective_behaviour_total ~ technical_mean + ownership + Age_1 + Gender, data = combined_data)

summary(model_trust2_own)


combined_data <- combined_data %>%

  mutate(across(starts_with("Manufacturer_Trust"), ~ as.numeric(as.character(.))))


combined_data$trust_total <- rowMeans(
  select(combined_data, Manufacturer_Trust_1, Manufacturer_Trust_2,
         Manufacturer_Trust_3, Manufacturer_Trust_4,
         Manufacturer_Trust_5, Manufacturer_Trust_6,
         Manufacturer_Trust_7),
  na.rm = TRUE
)

# Create single trust scale
combined_data$trust_total <- rowMeans(
  select(combined_data, Manufacturer_Trust_1, Manufacturer_Trust_2,
         Manufacturer_Trust_3, Manufacturer_Trust_4,
         Manufacturer_Trust_5, Manufacturer_Trust_6,
         Manufacturer_Trust_7),
  na.rm = TRUE
)
# Main effect
model_trust_total <- lm(protective_behaviour_total ~ trust_total + Age_1 + Gender, data = combined_data)
summary(model_trust_total)

# With ownership
model_trust_total_own <- lm(protective_behaviour_total ~ trust_total + ownership + Age_1 + Gender, data = combined_data)
summary(model_trust_total_own)

# Interaction model
combined_data$trust_total_x_ownership <- combined_data$trust_total * as.numeric(combined_data$ownership)

model_trust_interaction <- lm(protective_behaviour_total ~ trust_total + ownership + trust_total_x_ownership +
                              Age_1 + Gender, data = combined_data)
summary(model_trust_interaction)

# Step 1: Interaction terms
combined_data$trust_total_x_ownership <- combined_data$trust_total * as.numeric(combined_data$ownership)
combined_data$misconceptions_x_ownership <- combined_data$misconceptions_total * as.numeric(combined_data$ownership)

# Step 2: Main effect model
model_trust_total <- lm(protective_behaviour_total ~ trust_total + Age_1 + Gender, data = combined_data)
summary(model_trust_total)
```

```
# Step 3: Add ownership as a covariate
model_trust_total_own <- lm(protective_behaviour_total ~ trust_total + ownership + Age_1 + Gender, data = combined_data)
summary(model_trust_total_own)

# Step 4: Full interaction model (trust + misconceptions × ownership)
model_full_interaction <- lm(protective_behaviour_total ~ trust_total + misconceptions_total +
                             ownership + trust_total_x_ownership + misconceptions_x_ownership +
                             Age_1 + Gender, data = combined_data)
summary(model_full_interaction)
# Ensure ownership is numeric if needed (0 = non-user, 1 = user)
combined_data$ownership <- as.numeric(as.character(combined_data$ownership))

# 1. Pearson correlation
cor.test(combined_data$ownership, combined_data$misconceptions_total)

# 2. Welch Two-Sample t-test
t.test(misconceptions_total ~ ownership, data = combined_data)

# 3. Wilcoxon rank-sum test (non-parametric alternative)
wilcox.test(misconceptions_total ~ ownership, data = combined_data)
model_h2_linear <- lm(misconceptions_total ~ ownership + Age_1 + Gender, data = combined_data)
summary(model_h2_linear)
# Correlation
cor.test(combined_data$knowledge_total, combined_data$misconceptions_total)

# Linear regression
model_know_vs_mis <- lm(misconceptions_total ~ knowledge_total + Age_1 + Gender, data = combined_data)
summary(model_know_vs_mis)

# 1. Reconstruct individual misconception items using your scoring logic
#    Each TRUE = misconception (high confidence + incorrect)
# First: convert correctness and confidence to numeric
misconception_items <- data_scored %>%
  mutate(across(ends_with("_correctness"), as.numeric),
         across(ends_with("_confidence"), as.numeric))

# Then: apply misconception rule (incorrect AND confident)
misconception_matrix <- misconception_items %>%
  mutate(
    # Account security items
    across(paste0("account_security_", 1:13, "_1_correctness"),
        ~ . == 0 & get(sub("_1_correctness", "_2_confidence", cur_column())) %in% 4:6,
        .names = "{.col}_misconception"),

    # Device security items
    across(paste0("device_security_", 1:8, "_1_correctness"),
        ~ . == 0 & get(sub("_1_correctness", "_2_confidence", cur_column())) %in% 4:6,
        .names = "{.col}_misconception"),

    # Online security items
    across(paste0("online_security_", 1:9, "_1_correctness"),
        ~ . == 0 & get(sub("_1_correctness", "_2_confidence", cur_column())) %in% 4:6,
        .names = "{.col}_misconception"),

    # External security items
    across(paste0("external_security_", 1:4, "_1_correctness"),
        ~ . == 0 & get(sub("_1_correctness", "_2_confidence", cur_column())) %in% 4:6,
        .names = "{.col}_misconception")
  ) %>%
  # Select only misconception columns and convert to numeric
  select(ends_with("_misconception")) %>%
  mutate(across(everything(), as.numeric))


# Tetrachoric correlation matrix
tetra_corr <- tetrachoric(misconception_matrix)$rho

# KMO & Bartlett
KMO(tetra_corr)
cortest.bartlett(tetra_corr, n = nrow(misconception_matrix))

# Parallel analysis
fa.parallel(tetra_corr, fa = "fa", n.obs = nrow(misconception_matrix), show.legend = FALSE)

# Run EFA with 4 factors
misconception_fa <- fa(tetra_corr, nfactors = 6, rotate = "varimax", fm = "ml")
```

```
print(misconception_fa, cut = 0.3)



# Standardize PB_1 to PB_21 across users and non-users
combined_data <- combined_data %>%
  mutate(
    PB_1 = coalesce(Protective_user_1, Protective_nonuse_1),
    PB_2 = coalesce(Protective_user_2, Protective_nonuse_2),
    PB_3 = coalesce(Protective_user_3, Protective_nonuse_3),
    PB_4 = coalesce(Protective_user_4, Protective_nonuse_4),
    PB_5 = coalesce(Protective_user_5, Protective_nonuse_5),
    PB_6 = coalesce(Protective_user_6, Protective_nonuse_6),
    PB_7 = coalesce(Protective_user_7, Protective_nonuse_7),
    PB_8 = coalesce(Protective_user_8, Protective_nonuse_8),
    PB_9 = coalesce(Protective_user_9, Protective_nonuse_9),
    PB_10 = coalesce(Protective_user_10, Protective_nonuse_10),
    PB_11 = coalesce(Protective_user_11, Protective_nonuse_11),
    PB_12 = coalesce(Protective_user_12, Protective_nonuse_12),
    PB_13 = coalesce(Protective_user_13, Protective_nonuse_13),
    PB_14 = coalesce(Protective_user_14, Protective_nonuse_14),
    PB_15 = coalesce(Protective_user_15, Protective_nonuse_15),
    PB_16 = coalesce(Protective_user_16, Protective_nonuse_16),
    PB_17 = coalesce(speaker_Setup_user_1, Protective_nonuse_17),
    PB_18 = coalesce(Speaker_Setup_user_2, Protective_nonuse_18),
    PB_19 = coalesce(Speaker_Setup_user_3, Protective_nonuse_19),
    PB_20 = coalesce(Speaker_Setup_user_4, Protective_nonuse_20),
    PB_21 = coalesce(Speaker_Setup_user_5, Protective_nonuse_21)
  )

# Select only the standardized protective behaviour items
protective_items <- combined_data %>%
  select(starts_with("PB_"))
# Check KMO
KMO(cor(protective_items))

# Bartlett's test of sphericity
cortest.bartlett(cor(protective_items), n = nrow(protective_items))
```

```
# Parallel analysis to determine optimal number of factors
fa.parallel(protective_items, fa = "fa", fm = "ml", n.obs = nrow(protective_items))
# Run EFA with 1-4 factors depending on parallel results
protective_fa <- fa(protective_items, nfactors = 3, rotate = "oblimin", fm = "ml")

# Print output
print(protective_fa, cut = 0.3)



# Remove PB_19 from the dataset
protective_items_filtered <- protective_items %>% select(-PB_19)

# Re-run KMO and Bartlett's test
KMO(cor(protective_items_filtered))
cortest.bartlett(cor(protective_items_filtered), n = nrow(protective_items_filtered))

# Run parallel analysis to reassess the number of factors
fa.parallel(protective_items_filtered, fa = "fa", fm = "ml", n.obs = nrow(protective_items_filtered))

# Rerun the factor analysis (adjust nfactors based on parallel output)
protective_fa_updated <- fa(protective_items_filtered, nfactors = 3, rotate = "oblimin", fm = "ml")

# View the output
print(protective_fa_updated, cut = 0.3)



combined_data$PB_factor1 <- rowMeans(select(protective_items_filtered, PB_1, PB_2, PB_3, PB_4, PB_5, PB_10, PB_11, PB_12, PB_13,
PB_14, PB_17), na.rm = TRUE)
combined_data$PB_factor2 <- rowMeans(select(protective_items_filtered, PB_6, PB_7, PB_8, PB_9, PB_21), na.rm = TRUE)
combined_data$PB_factor3 <- rowMeans(select(protective_items_filtered, PB_15, PB_16, PB_18, PB_20), na.rm = TRUE)

# H1: Misconceptions predicting each protective behaviour factor
model_H1_f1 <- summary(lm(PB_factor1 ~ misconceptions_total + ownership + Age_1 + Gender, data = combined_data))
model_H1_f2 <- summary(lm(PB_factor2 ~ misconceptions_total + ownership + Age_1 + Gender, data = combined_data))
model_H1_f3 <- summary(lm(PB_factor3 ~ misconceptions_total + ownership + Age_1 + Gender, data = combined_data))

# H3: Ownership predicting each protective behaviour factor (controlling for knowledge and trust)
model_H3_f1 <- summary(aov(PB_factor1 ~ ownership + knowledge_total + general_mean + technical_mean, data = combined_data))
```

```
model_H3_f2 <- summary(aov(PB_factor2 ~ ownership + knowledge_total + general_mean + technical_mean, data = combined_data))
model_H3_f3 <- summary(aov(PB_factor3 ~ ownership + knowledge_total + general_mean + technical_mean, data = combined_data))

# H5: Trust predicting each protective behaviour factor (controlling for age and gender)
model_H5_f1 <- summary(lm(PB_factor1 ~ general_mean + technical_mean + Age_1 + Gender, data = combined_data))
model_H5_f2 <- summary(lm(PB_factor2 ~ general_mean + technical_mean + Age_1 + Gender, data = combined_data))
model_H5_f3 <- summary(lm(PB_factor3 ~ general_mean + technical_mean + Age_1 + Gender, data = combined_data))

# Print results
model_H1_f1; model_H1_f2; model_H1_f3
model_H3_f1; model_H3_f2; model_H3_f3
model_H5_f1; model_H5_f2; model_H5_f3
shapiro.test(residuals(model_H1_f1))
shapiro.test(residuals(model_H1_f2))
shapiro.test(residuals(model_H1_f3))

shapiro.test(residuals(model_H3_f1))
shapiro.test(residuals(model_H3_f2))
shapiro.test(residuals(model_H3_f3))
qqnorm(residuals(model_H3_f1)); qqline(residuals(model_H3_f1))
qqnorm(residuals(model_H3_f2)); qqline(residuals(model_H3_f2))
qqnorm(residuals(model_H3_f3)); qqline(residuals(model_H3_f3))
shapiro.test(residuals(model_H5_f1))
shapiro.test(residuals(model_H5_f2))
shapiro.test(residuals(model_H5_f3))
qqnorm(residuals(model_H5_f1)); qqline(residuals(model_H5_f1))
qqnorm(residuals(model_H5_f2)); qqline(residuals(model_H5_f2))
qqnorm(residuals(model_H5_f3)); qqline(residuals(model_H5_f3))
# Create interaction terms
combined_data$misconceptions_x_ownership <- combined_data$misconceptions_total * as.numeric(combined_data$ownership)
combined_data$trust1_x_ownership <- combined_data$general_mean * as.numeric(combined_data$ownership)
combined_data$trust2_x_ownership <- combined_data$technical_mean * as.numeric(combined_data$ownership)

# Full interaction model for Factor 1
full_model_f1 <- lm(PB_factor1 ~ misconceptions_total + general_mean + technical_mean +
            ownership + misconceptions_x_ownership + trust1_x_ownership + trust2_x_ownership +
            Age_1 + Gender, data = combined_data)
summary(full_model_f1)
```

```
full_model_f1 <- lm(PB_factor1 ~ misconceptions_total + general_mean + technical_mean +
              ownership +
              Age_1 + Gender, data = combined_data)
summary(full_model_f1)

# Full interaction model for Factor 2
full_model_f2 <- lm(PB_factor2 ~ misconceptions_total + general_mean + technical_mean +
              ownership +
              Age_1 + Gender, data = combined_data)
summary(full_model_f2)
full_model_f2 <- lm(PB_factor2 ~ misconceptions_total + general_mean + technical_mean +
              ownership + misconceptions_x_ownership + trust1_x_ownership + trust2_x_ownership +
              Age_1 + Gender, data = combined_data)
summary(full_model_f2)



# Full interaction model for Factor 3
full_model_f3 <- lm(PB_factor3 ~ misconceptions_total + general_mean + technical_mean +
              ownership +
              Age_1 + Gender, data = combined_data)
summary(full_model_f3)
full_model_f3 <- lm(PB_factor3 ~ misconceptions_total + general_mean + technical_mean +
              ownership + misconceptions_x_ownership + trust1_x_ownership + trust2_x_ownership +
              Age_1 + Gender, data = combined_data)
summary(full_model_f3)

#full model


full_model_all_no_int <- lm(PB_factor1 + PB_factor2 + PB_factor3 ~ misconceptions_total + general_mean + technical_mean +
              ownership + Age_1 + Gender, data = combined_data)
summary(full_model_all_no_int)

full_model_f3 <- lm(PB_factor1 + PB_factor2 + PB_factor3 ~ misconceptions_total + general_mean + technical_mean +
              ownership + misconceptions_x_ownership + trust1_x_ownership + trust2_x_ownership +
              Age_1 + Gender, data = combined_data)
summary(full_model_f3)

manova_model <- manova(cbind(PB_factor1, PB_factor2, PB_factor3) ~ misconceptions_total + general_mean + technical_mean +
            ownership + misconceptions_x_ownership + trust1_x_ownership + trust2_x_ownership + knowledge_total +
            Age_1 + Gender, data = combined_data)

summary(manova_model)

combined_data$Total_Protective_Behavior <- (combined_data$PB_factor1 + combined_data$PB_factor2 + combined_data$PB_factor3) / 3

full_model_single_DV <- lm(Total_Protective_Behavior ~ misconceptions_total + general_mean + technical_mean +
              ownership + misconceptions_x_ownership + trust1_x_ownership + trust2_x_ownership +
              Age_1 + Gender, data = combined_data)

# Step 3: View the summary of the single DV model
summary(full_model_single_DV)
resids <- residuals(manova_model)
mvnResids <- mvn(data = resids)
print(mvnResids)



cor.test(combined_data$misconceptions_total, combined_data$knowledge_total, method = "pearson")

model_h3 <- aov(PB_factor1 ~ ownership + knowledge_total + trust_mean, data = combined_data)
summary(model_h3)
model_h3_anova <- aov(PB_factor1 ~ ownership, data = combined_data)
summary(model_h3_anova)
model_h3_anova <- aov(PB_factor2 ~ ownership, data = combined_data)
summary(model_h3_anova)
model_h3_anova <- aov(PB_factor3 ~ ownership, data = combined_data)
summary(model_h3_anova)

t.test(misconceptions_total ~ ownership, data = combined_data)
wilcox.test(misconceptions_total ~ ownership, data = combined_data)



# Extract individual misconception items (binary: 1 = misconception, 0 = no misconception)
misconception_individual <- misconception_matrix

# Extract individual protective behavior items
protective_individual <- protective_items_filtered
```

```
# Correlation matrix between all misconception and protective behavior items
cor_matrix <- cor(misconception_individual, protective_individual,
                  use = "complete.obs", method = "pearson")

# View the correlation matrix
print(cor_matrix)

item_corr_results <- corr.test(misconception_individual,protective_individual, use = "complete.obs", method = "pearson")

# Print correlation matrix
cat("Correlation Matrix:\n")
print(round(item_corr_results$r, 3))

# Print p-value matrix
cat("\nP-value Matrix:\n")
print(round(item_corr_results$p, 3))

# Find and display significant correlations
sig_mask <- item_corr_results$p < 0.05 & !is.na(item_corr_results$p)
sig_indices <- which(sig_mask, arr.ind = TRUE)

if(nrow(sig_indices) > 0) {
  cat("\nSignificant correlations (p < 0.05):\n")

  sig_results <- data.frame(
    Misconception_Item = rownames(item_corr_results$r)[sig_indices[,1]],
    Protective_Item = colnames(item_corr_results$r)[sig_indices[,2]],
    Correlation = round(item_corr_results$r[sig_indices], 3),
    P_value = round(item_corr_results$p[sig_indices], 3)
  )

  print(sig_results)
} else {
  cat("\nNo significant correlations found at p < 0.05\n")
}
```

```
# Prepare predictors (all your individual items)
# Get knowledge variables from data_scored and add to combined_data
# First, make sure we match the right rows
combined_data$acc_know <- data_scored$acc_know[match(rownames(combined_data), rownames(data_scored))]
combined_data$dev_know <- data_scored$dev_know[match(rownames(combined_data), rownames(data_scored))]
combined_data$onl_know <- data_scored$onl_know[match(rownames(combined_data), rownames(data_scored))]
combined_data$ext_know <- data_scored$ext_know[match(rownames(combined_data), rownames(data_scored))]

# Now prepare predictors with the correct knowledge variables
predictor_items <- combined_data %>%
  select(
    # Individual protective behavior items
    starts_with("PB_"),

    # Individual trust items
    starts_with("Manufacturer_Trust"),

    # Knowledge domain scores (now from the correct dataset)
    acc_know, dev_know, onl_know, ext_know,

    # Demographics
    Age_1, Gender, ownership
  ) %>%
  mutate(
    Gender = as.numeric(as.factor(Gender)),
    ownership = as.numeric(as.character(ownership)),
    Age_1 = as.numeric(as.character(Age_1))
  )

# Make sure both datasets have same number of rows
n_rows <- min(nrow(predictor_items), nrow(misconception_items_individual))
predictor_items <- predictor_items[1:n_rows, ]
misconception_items <- misconception_items_individual[1:n_rows, ]

# Check that knowledge variables are now included
cat("Knowledge variables included:\n")
print(summary(predictor_items[, c("acc_know", "dev_know", "onl_know", "ext_know")]))
```

```
# Make sure both datasets have same number of rows
n_rows <- min(nrow(predictor_items), nrow(misconception_individual))
predictor_items <- predictor_items[1:n_rows, ]
misconception_items <- misconception_individual[1:n_rows, ]

# Correlation analysis between predictors and individual misconceptions
item_misc_correlations <- corr.test(predictor_items, misconception_individual,
                                    use = "complete")

# Print correlation matrix
cat("Correlations between predictor items and individual misconceptions:\n")
print(round(item_misc_correlations$r, 3))

# Find significant correlations
sig_cors <- which(item_misc_correlations$p < 0.05, arr.ind = TRUE)

if(length(sig_cors) > 0) {
  cat("\nSignificant correlations (p < 0.05):\n")
  sig_results <- data.frame(
    Predictor = rownames(item_misc_correlations$r)[sig_cors[,1]],
    Misconception = colnames(item_misc_correlations$r)[sig_cors[,2]],
    Correlation = round(item_misc_correlations$r[sig_cors], 3),
    P_value = round(item_misc_correlations$p[sig_cors], 4)
  )

  # Sort by correlation strength
  sig_results <- sig_results[order(abs(sig_results$Correlation), decreasing = TRUE), ]
  print(sig_results)
}


# Extract individual knowledge items (scored as 0/1 for incorrect/correct)
individual_knowledge_items <- data_scored %>%
  select(
    # Account security items (13 items)
    all_of(names(key_account)),

    # Device security items (8 items)
    all_of(names(key_device)),

    # Online security items (9 items)
    all_of(names(key_online)),

    # External security items (4 items)
    all_of(names(key_external))
  )

# Rename for clarity
colnames(individual_knowledge_items) <- c(
  # Account items
  paste0("know_acc_", 1:13),

  # Device items
  paste0("know_dev_", 1:8),

  # Online items
  paste0("know_onl_", 1:9),

  # External items
  paste0("know_ext_", 1:4)
)

# Check the structure
cat("Individual knowledge items extracted:\n")
cat("Dimensions:", dim(individual_knowledge_items), "\n")
cat("Sample of first few items:\n")
print(head(individual_knowledge_items[, 1:6]))

# Create comprehensive predictor dataset
predictor_items_complete <- data_scored %>%
  select(
    # Demographics
    Age_1, Gender, ownership
  ) %>%
  mutate(
    Gender = as.numeric(as.factor(Gender)),
    ownership = as.numeric(as.character(ownership)),
    Age_1 = as.numeric(as.character(Age_1))
  )

# Add individual knowledge items
predictor_items_complete <- cbind(predictor_items_complete, individual_knowledge_items)

# Add trust variables from combined_data
trust_cols <- paste0("Manufacturer_Trust_", 1:7)
for(col in trust_cols) {
  if(col %in% names(combined_data)) {
    predictor_items_complete[[col]] <- combined_data[[col]][match(rownames(data_scored), rownames(combined_data))]
  }
}

# Add protective behavior items
pb_cols <- paste0("PB_", 1:21)
for(col in pb_cols) {
  if(col %in% names(combined_data)) {
    predictor_items_complete[[col]] <- combined_data[[col]][match(rownames(data_scored), rownames(combined_data))]
  }
}

# Ensure numeric conversion for all items
predictor_items_complete <- predictor_items_complete %>%
  mutate(across(starts_with("know_"), as.numeric),
         across(starts_with("Manufacturer_Trust"), as.numeric),
         across(starts_with("PB_"), as.numeric))

# Remove any columns with all NA values
predictor_items_complete <- predictor_items_complete[, colSums(is.na(predictor_items_complete)) < nrow(predictor_items_complete)]

# Align with misconception data
n_rows <- min(nrow(predictor_items_complete), nrow(misconception_individual))
predictor_items_final <- predictor_items_complete[1:n_rows, ]
misconception_items_final <- misconception_individual[1:n_rows, ]

cat("Final dataset dimensions:\n")
cat("Predictors:", dim(predictor_items_final), "\n")
cat("Misconceptions:", dim(misconception_items_final), "\n")


# Extract just the knowledge items for focused analysis
knowledge_items_only <- predictor_items_final %>%
  select(starts_with("know_"))

# Correlation between individual knowledge items and individual misconceptions
knowledge_misc_correlations <- corr.test(knowledge_items_only, misconception_items_final,
                                         use = "complete")

# Print correlation matrix
cat("Correlations between individual knowledge items and misconceptions:\n")
print(round(knowledge_misc_correlations$r, 3))

# Find significant correlations
sig_knowledge_cors <- which(knowledge_misc_correlations$p < 0.05, arr.ind = TRUE)

if(length(sig_knowledge_cors) > 0) {
  cat("\nSignificant knowledge-misconception correlations (p < 0.05):\n")
  sig_knowledge_results <- data.frame(
    Knowledge_Item = rownames(knowledge_misc_correlations$r)[sig_knowledge_cors[,1]],
    Misconception_Item = colnames(knowledge_misc_correlations$r)[sig_knowledge_cors[,2]],
    Correlation = round(knowledge_misc_correlations$r[sig_knowledge_cors], 3),
    P_value = round(knowledge_misc_correlations$p[sig_knowledge_cors], 4)
  )

  # Sort by correlation strength
  sig_knowledge_results <- sig_knowledge_results[order(abs(sig_knowledge_results$Correlation), decreasing = TRUE), ]
  print(sig_knowledge_results)
}

# Logistic regression for each knowledge item predicting each misconception
knowledge_logistic_results <- data.frame()

for(knowledge_item in names(knowledge_items_only)) {
  for(misconception_item in names(misconception_items_final)) {
```

```
# Create dataset for this analysis
analysis_data <- data.frame(
  knowledge = knowledge_items_only[[knowledge_item]],
  misconception = misconception_items_final[[misconception_item]],
  Age_1 = predictor_items_final$Age_1,
  Gender = predictor_items_final$Gender,
  ownership = predictor_items_final$ownership
)

# Remove rows with missing data
analysis_data <- analysis_data[complete.cases(analysis_data), ]

# Run logistic regression if we have enough data
if(nrow(analysis_data) > 10 && sum(analysis_data$misconception) > 0) {
  tryCatch({
    model <- glm(misconception ~ knowledge + Age_1 + Gender + ownership,
                 data = analysis_data, family = binomial)

    # Extract results for knowledge predictor
    coef_summary <- summary(model)$coefficients
    if("knowledge" %in% rownames(coef_summary)) {
      knowledge_row <- coef_summary["knowledge", ]

      # Calculate odds ratio
      odds_ratio <- exp(knowledge_row["Estimate"])

      # Store results
      knowledge_logistic_results <- rbind(knowledge_logistic_results, data.frame(
        Knowledge_Item = knowledge_item,
        Misconception_Item = misconception_item,
        Beta = round(knowledge_row["Estimate"], 4),
        SE = round(knowledge_row["Std. Error"], 4),
        z_value = round(knowledge_row["z value"], 3),
        p_value = round(knowledge_row["Pr(>|z|)"], 4),
        Odds_Ratio = round(odds_ratio, 3),
        N = nrow(analysis_data)
      ))
```

```
        }
      }, error = function(e) {
        # Skip problematic models
      })
    }
  }
}

# Sort by p-value
knowledge_logistic_results <- knowledge_logistic_results[order(knowledge_logistic_results$p_value), ]

# Show significant results
significant_knowledge <- knowledge_logistic_results[knowledge_logistic_results$p_value < 0.05, ]
cat("Significant individual knowledge predictors of misconceptions:\n")
print(significant_knowledge)

library(psych)

# Method 1: Using the individual misconception items you already created
misconception_alpha <- psych::alpha(misconception_individual)

# Print results
cat("Misconceptions Scale Reliability:\n")
cat("Cronbach's α =", round(misconception_alpha$total$raw_alpha, 3), "\n")
cat("Standardized α =", round(misconception_alpha$total$std.alpha, 3), "\n")
cat("Number of items =", misconception_alpha$total$nvar, "\n")

# Print the full output for more details
print(misconception_alpha)

mean(combined_data$PB_factor1, na.rm = TRUE)
sd(combined_data$PB_factor1, na.rm = TRUE)
mean(combined_data$PB_factor2, na.rm = TRUE)
sd(combined_data$PB_factor2, na.rm = TRUE)

mean(combined_data$PB_factor3, na.rm = TRUE)
sd(combined_data$PB_factor3, na.rm = TRUE)
```

```
PB_factor1_items_for_alpha <- protective_items_filtered %>%
  select(PB_1, PB_2, PB_3, PB_4, PB_5, PB_10, PB_11, PB_12, PB_13, PB_14, PB_17)

alpha_PB1_results <- alpha(PB_factor1_items_for_alpha)
print("Alpha for PB_factor1 (Physical Security Behaviors):")
print(alpha_PB1_results$alpha.drop) # This gives alpha if item is dropped and overall alpha.
print(alpha_PB1_results$total$raw_alpha) # This gives the overall Cronbach's alpha for the factor

# --- Calculate Cronbach's Alpha for PB_factor2 (Account Management) ---
# Use the items that you defined for PB_factor2
PB_factor2_items_for_alpha <- protective_items_filtered %>%
  select(PB_6, PB_7, PB_8, PB_9, PB_21)

alpha_PB2_results <- alpha(PB_factor2_items_for_alpha)
print("Alpha for PB_factor2 (Account Management):")
print(alpha_PB2_results$total$raw_alpha)
alpha_PB2_results <- alpha(PB_factor2_items_for_alpha)
print("Alpha for PB_factor1 (Physical Security Behaviors):")
print(alpha_PB2_results$alpha.drop)

# --- Calculate Cronbach's Alpha for PB_factor3 (Password Management) ---
# Use the items that you defined for PB_factor3
PB_factor3_items_for_alpha <- protective_items_filtered %>%
  select(PB_15, PB_16, PB_18, PB_20)

alpha_PB3_results <- alpha(PB_factor3_items_for_alpha)
print("Alpha for PB_factor3 (Password Management):")
print(alpha_PB3_results$total$raw_alpha)
# Make sure you're working with your data frame, assuming it's named 'combined_data'

# 1. Center the continuous predictor variables
# You can create new columns for the centered versions
combined_data$misconceptions_total_c <- combined_data$misconceptions_total - mean(combined_data$misconceptions_total, na.rm = TRUE)
combined_data$ownership_c <- combined_data$ownership - mean(combined_data$ownership, na.rm = TRUE)
combined_data$trust1_c <- combined_data$general_mean - mean(combined_data$general_mean, na.rm = TRUE) # Assuming trust1 was the original variable for trust1_x_ownership
combined_data$trust2_c <- combined_data$technical_mean - mean(combined_data$technical_mean, na.rm = TRUE) # Assuming trust2 was the original variable for trust2_x_ownership
```

```
# You might also center other continuous predictors not involved in interactions,
# but it's primarily crucial for those that form interaction terms.
combined_data$general_mean_c <- combined_data$general_mean - mean(combined_data$general_mean, na.rm = TRUE)
combined_data$technical_mean_c <- combined_data$technical_mean - mean(combined_data$technical_mean, na.rm = TRUE)
combined_data$Age_1_c <- combined_data$Age_1 - mean(combined_data$Age_1, na.rm = TRUE)

# 2. Create the interaction terms using the CENTERED variables
# Use the colon (:) operator for interactions in lm() directly, it's generally
# cleaner and R handles the product for you.
# If you want to include all main effects AND interaction terms, use the asterisk (*)
# Example: A * B is equivalent to A + B + A:B
# This is usually preferred for interpretability of the interaction effect.

# So, your full model formula would look like this:

# Assuming you've created all the centered variables like:
# combined_data$misconceptions_total_c <- combined_data$misconceptions_total - mean(combined_data$misconceptions_total, na.rm = TRUE)
# combined_data$ownership_c <- combined_data$ownership - mean(combined_data$ownership, na.rm = TRUE)
# combined_data$general_mean_c <- combined_data$general_mean - mean(combined_data$general_mean, na.rm = TRUE) # This is trust1_c
# combined_data$technical_mean_c <- combined_data$technical_mean - mean(combined_data$technical_mean, na.rm = TRUE) # This is trust2_c
# combined_data$Age_1_c <- combined_data$Age_1 - mean(combined_data$Age_1, na.rm = TRUE)

# Correct way to run multivariate multiple regression in R with lm()
# This will produce an 'mlm' object, and summary() on this object
# will provide separate output for each response.

# Main Effects Model (no interaction terms)
# Still use centered variables for main effects, especially if you plan to
# later add interaction terms. Centering helps interpret main effects
# as the effect when other predictors are at their mean.
library(dplyr)

combined_data <- combined_data %>%
  mutate(Gender = case_when(
    Gender == 3 ~ 1.5,
    Gender == 4 ~ 1.5,
    TRUE ~ as.numeric(Gender)
  ))



main_effects_model_mlm <- lm(cbind(PB_factor1, PB_factor2, PB_factor3) ~
                            misconceptions_total_c +
                            ownership_c +
                            general_mean_c +
                            technical_mean_c +
                            Age_1_c +
                            Gender,
                           data = combined_data)

summary(main_effects_model_mlm)

full_model_mlm <- lm(cbind(PB_factor1, PB_factor2, PB_factor3) ~
                     misconceptions_total_c * ownership_c +
                     general_mean_c * ownership_c +
                     technical_mean_c * ownership_c +
                     Age_1_c + Gender,
                   data = combined_data)

summary(full_model_mlm)



# Load the necessary package for a nicely formatted correlation matrix (optional but recommended)
# If you don't have it, install it: install.packages("corrplot")
# Install if you don't have it:
# install.packages("Hmisc")
library(Hmisc)
library(dplyr) # For select if you prefer

# Select only the continuous variables you want in the correlation matrix
continuous_vars <- combined_data %>%
  select(misconceptions_total,
         ownership,
         general_mean,
         technical_mean,
         Age_1,
    PB_factor1,
    PB_factor2,
    PB_factor3)

# Calculate correlation matrix with p-values
# rcorr() requires a matrix as input, so convert the data frame
cor_results <- rcorr(as.matrix(continuous_vars), type = "pearson") # "pearson" is default but good to specify

# Print the correlation coefficients
print("Correlation Coefficients (r):")
print(round(cor_results$r, 2))

# Print the p-values
print("P-values:")
print(round(cor_results$P, 3)) # Often round p-values to 3 decimal places

# --- Step 1: Run Main Effects Models for EACH PB_factor individually ---
# This ensures you have separate lm objects for each PB factor's main effects model.

main_effects_PB1 <- lm(PB_factor1 ~ misconceptions_total_c + ownership_c +
                        general_mean_c + technical_mean_c + Age_1_c + Gender,
                      data = combined_data)

main_effects_PB2 <- lm(PB_factor2 ~ misconceptions_total_c + ownership_c +
                        general_mean_c + technical_mean_c + Age_1_c + Gender,
                      data = combined_data)

main_effects_PB3 <- lm(PB_factor3 ~ misconceptions_total_c + ownership_c +
                        general_mean_c + technical_mean_c + Age_1_c + Gender,
                      data = combined_data)



# --- Step 2: Run Full Interaction Models for EACH PB_factor individually ---
# This ensures you have separate lm objects for each PB factor's full model.

full_interactions_PB1 <- lm(PB_factor1 ~ misconceptions_total_c * ownership_c +
                              general_mean_c * ownership_c +
                              technical_mean_c * ownership_c +
                              Age_1_c + Gender,
                             data = combined_data)

full_interactions_PB2 <- lm(PB_factor2 ~ misconceptions_total_c * ownership_c +
                              general_mean_c * ownership_c +
                              technical_mean_c * ownership_c +
                              Age_1_c + Gender,
                             data = combined_data)

full_interactions_PB3 <- lm(PB_factor3 ~ misconceptions_total_c * ownership_c +
                              general_mean_c * ownership_c +
                              technical_mean_c * ownership_c +
                              Age_1_c + Gender,
                             data = combined_data)


# --- Step 3: Now, perform the anova() comparisons using these individual lm objects ---

# For PB_factor1:
anova(main_effects_PB1, full_interactions_PB1)

# For PB_factor2:
anova(main_effects_PB2, full_interactions_PB2)

# For PB_factor3:
anova(main_effects_PB3, full_interactions_PB3)
```

Artificial Intelligence statement:

During the preparation of this report, I (Karam Altabbaa) utilized ChatGPT 4 as a data analysis tool solely to help generate R code. While using this AI, I thoroughly reviewed the code and output, ensuring no mistakes were made during analysis, I take full responsibility for the final outcome.