Exploring Deceptive Design Patterns in Voice Interfaces

KENTRELL OWENS, University of Washington, USA
JOHANNA GUNAWAN, Northeastern University, USA
DAVID CHOFFNES, Northeastern University, USA
PARDIS EMAMI-NAEINI, Duke University, USA
TADAYOSHI KOHNO, University of Washington, USA
FRANZISKA ROESNER, University of Washington, USA

Deceptive design patterns (sometimes called "dark patterns") are user interface design elements that may trick, deceive, or mislead users into behaviors that often benefit the party implementing the design over the end user. Prior work has taxonomized, investigated, and measured the prevalence of such patterns primarily in visual user interfaces (e.g., on websites). However, as the ubiquity of voice assistants and other voice-assisted technologies increases, we must anticipate how deceptive designs will be (and indeed, are already) deployed in voice interactions. This paper makes two contributions towards characterizing and surfacing deceptive design patterns in voice interfaces. First, we make a conceptual contribution, identifying key characteristics of voice interfaces that may enable deceptive design patterns, and surfacing existing and theoretical examples of such patterns. Second, we present the findings from a scenario-based user survey with 93 participants, in which we investigate participants' perceptions of voice interfaces that we consider to be both deceptive and non-deceptive.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**; *Personal digital assistants*; • **Social and professional topics → Consumer products policy**.

Additional Key Words and Phrases: dark patterns; manipulative design patterns; consumer protection; manipulation; deceptive design

1 INTRODUCTION

> User: *Voice Assistant, cancel my subscription.*
>
> Voice Assistant: *To manage your subscription, please visit our website.*

Deceptive and manipulative design patterns (sometimes called "dark patterns")[1] are user interface design elements that may trick, deceive, or mislead users into behaviors that often benefit the party implementing the design over the end user. For example, a service may make it easy for a user to subscribe with a single interaction, but difficult to unsubscribe; or a website may make it easy for a user to consent to all data collection, but difficult to opt out. Whether the result of intentional manipulation by designers, poor design (e.g., due to a designer's habits, faulty assumptions, or priorities), or other constraints of the interface, these types of design patterns make it difficult for users to make and implement the decisions they might make in response to a more neutral or user-centered design — impacting users' privacy, security, finances, autonomy, and more.

[1]In this paper, for simplicity of exposition, we generally use the term "deceptive design" to refer to this type of interface, while acknowledging that other terms ("manipulative", "misleading", etc.) might be more precise in some cases. We prefer this term to "dark pattern", which has been criticized [54].

Manuscript submitted to ACM

Researchers, users, and regulators have taken a significant interest in deceptive design patterns in recent years. For example, the Twitter account `@darkpatterns` collects numerous examples, regulatory bodies in Europe and the U.S. explicitly call out deceptive design patterns [22, 23, 63], and a rich body of academic literature has begun to taxonomize, investigate, and measure the prevalence of such patterns (see Section 2). Prior focus on deceptive design patterns has generally been in the context of *visual* user interfaces (e.g., on websites or in mobile apps). However, our work here is motivated by the following observation: as the ubiquity of voice assistants and other voice-assisted technologies increases, we must anticipate how deceptive designs will be (and indeed, are already) deployed in *voice* interactions.

For instance, the example at the top of this section is based directly on Amazon Alexa's response when a user attempts to use the voice assistant to cancel their Amazon Prime membership. While redirecting a user to a non-voice interface may be in part the designer's solution to the limited bandwidth of a voice/audio interface rather than the intent to manipulate or deceive the user, we consider this interaction to be manipulative since users are able to *subscribe* to an Amazon Prime membership using only the voice interface. A more user-centered interaction might be:

> User: *Voice Assistant, cancel my subscription.*
>
> Voice Assistant: *I've canceled your subscription, effective July 1. If this was a mistake, please visit our website to manage your subscription.*

In considering current and future potential deceptive designs in voice interfaces, we observe that the voice/audio modality has some significant differences from visual interfaces. For example, a visual design can present much more information to the user at once, compared to a spoken response from a voice assistant. A voice interface could also manipulate a user with volume or tone, properties that are not present in a visual interface.

In this work we thus seek to answer the following research questions:

- RQ1: How could (or do) deceptive design patterns manifest in voice interfaces, specifically voice assistants? How can the unique properties of voice interfaces amplify their severity?
- RQ2: Do people find deceptive design patterns in voice assistants problematic and if so, how problematic? What factors influence people's perceptions of how problematic these design patterns are?
- RQ3: What are people's experiences with deceptive design patterns in voice assistants in the wild today?

To answer RQ1, we conduct a structured expert panel brainstorming exercise among the co-authors (who have previous research experience and expertise on deceptive design patterns and problematic content online). We identify six unique properties of voice interfaces that have implications for deceptive design patterns, and we develop a corresponding set of scenarios illustrating what we believe to be deceptive and non-deceptive voice assistant interactions. While these properties may not be collectively exhaustive, we believe that they capture important characteristics of voice interfaces, which may be used to implement deceptive design patterns.

To answer RQ2 and RQ3, we use the results of our brainstorming exercise to design a user survey based on the scenarios we developed. We collect and analyze data from 93 participants. We find that scenarios we intended to be deceptive were also rated by participants as more problematic than non-deceptive scenarios, but that many participants also considered these scenarios to be unproblematic. We also present concrete examples of problematic voice assistant interactions from participants' own experiences; their concerns align with the properties and scenarios we developed in our brainstorming exercise.

In summary, this paper makes the following contributions:

(1) A conceptual contribution, identifying key characteristics of voice interfaces that may enable deceptive designs, and surfacing existing and theoretical examples of such design patterns (RQ1, Section 3).

(2) An empirical contribution, presenting the findings of a user survey (Section 4) in which we investigate participants' perceptions of potentially deceptive voice interactions (RQ2) and collect their previous experiences with deceptive designs in voice interfaces (RQ3).

Based on our findings, we reflect on the role of deceptive and manipulative designs in current and future voice interfaces, and we make recommendations for designers, researchers, and regulators.

2 BACKGROUND & RELATED WORK

2.1 Deceptive design, or dark patterns

Deceptive design patterns are part of an emerging area of research spanning mostly synonymous terms like dark patterns, manipulative design patterns, and manipulative interfaces. The vast majority of recent work investigates dark and deceptive design patterns in visual or web interfaces, though some work has considered home robots [35], and other early work has begun to consider XR interfaces [33, 45]. To our knowledge, no work in this space has explicitly focused on or included voice assistants.

2.1.1 Taxonomies and categorization of dark patterns. Prior taxonomy work in this space identified [8] and categorized dark patterns or manipulative interfaces by their mechanisms [7, 16, 25, 43] or shared traits [43, 44]. Dark patterns have also been categorized by interaction contexts [28] and more deeply investigated in contexts like shopping [43], consent interactions [26, 27, 29, 32, 34, 56], and games [2]. Taxonomies and categories were derived through a variety of approaches: Gray et al. and Chivukula et al. collected examples from online design communities and utilized a content analysis method [13, 25], Mathur et al. conducted a large-scale scrape of e-commerce sites and used text analysis and data clustering [43], Bösch et al. started with a survey of privacy-forward design pattern literature then reversed these themes to derive dark privacy patterns [7], and Gunawan et al. grouped dark patterns by user interaction context [28].

2.1.2 Surveys and user studies. Though a few empirical studies collect and label dark pattern samples to better understand different types of deceptive designs [17, 25, 28, 43], a growing body of literature turns to users to investigate outcomes, dark pattern awareness, and perceived deception.

Dark pattern detection and awareness. DiGeronimo et al. supplemented author-coded empirical work by asking users to watch pre-recorded videos of user interactions with mobile apps and identify dark patterns, noting that participants failed to detect dark patterns [17]. Luguri & Strahilevitz ran two large scale experiments in the style of an A/B test to investigate how users responded to different designs and subsequently made decisions, finding users susceptible to dark patterns (and more concerningly, finding that participants with lower education levels were more susceptible to both mild and aggressive dark patterns) [37]. Bongard-Blanchy et al. showed participants a series of static interfaces to determine how well participants were able to detect dark patterns [6]. Bhoot et al. opted for a live task-based experiment in order to understand user reactions to the Forced Continuity and Roach Motel [8] dark patterns, as well as a questionnaire finding that participants were unable to detect all 12 dark patterns included in the survey [38].

User outcomes and harms. In an international mixed-methods study, Gray et al. builds upon dark patterns concepts to capture the range of reactions and emotions users feel in response to experiences of manipulation [24]. Bhoot et al. asked participants to measure their level of frustration with dark patterns, as well as to describe how trustworthy or misleading they felt an interface was [38]. Through empirical design analysis, Milder & Savino inspected privacy outcomes of interface interference patterns, then found that users do not feel wholly in control of the data they share [46].

Other emergent work investigates how dark patterns are employed to increase user engagement and often increase the "addictiveness" of a web service [2, 47].

Some dark patterns work focuses on the context of cookie consent regimes [26, 27, 29, 32, 34, 39, 56] and privacy-related outcomes [7, 37], with governments taking notice [9] (and some taking explicit action against dark patterns [3, 4]). Governments have also focused on competition and market harms to consumers [9, 14, 23].

2.2 Smart voice assistants

The voice modality (particularly, the conversational question-and-response model) presents unique challenges for designing user interactions as compared to visual web interfaces; Ma & Liu [40] articulate some of these with regards to exploratory search (sometimes called wayfinding by others [55]).

Competing values in smart device design. Volkel et al. delivered a dialogue elicitation study to glean how users imagine ideal conversations with voice assistants, finding that participants preferred a human-like persona and more personal interactions that incorporate knowledge about the user and their environment [60].

A growing body of work has also explored users' security and privacy concerns with voice assistants and other IoT devices, including in relation to perceived benefits of these devices (e.g., [20, 21, 36, 48, 58, 62]). The research community at the intersection of design, privacy, and HCI utilizes speculative fiction and structured brainstorming exercises to imagine future designs that might be disadvantageous if not explicitly harmful to different kinds of users [59, 61]. Mare et al. explore the tensions between security, privacy, design & usability, and reliability in smart home platforms [42].

Proven security and privacy issues in smart voice assistants. Smart home and consumer IoT devices (including voice-enabled smart speakers) were discovered to expose information to third-parties, with encryption not preventing potential eavesdroppers from being able to infer device activity [51]. Smart speakers were discovered to be vulnerable to privacy leakage with malicious actors able to infer voice commands from encrypted traffic [31]; in other work, smart speakers were able to be activated remotely despite such a feature not being provided by default [11]. As always-on devices, smart speakers present unique privacy issues for users, particularly when speakers mistakenly activate and begin recording without user knowledge or input [18].

Concerns in voice assistant skill markets. As voice assistants become more prevalent, the voice application market introduces additional vectors of insecurity. Cheng et al. found the Amazon Alexa and Google Assistant platforms allowing policy-violating applications or skills to be distributed in app marketplaces, including kids-specific skills [12]. The same authors surveyed participants to gather reactions on trustworthiness of voice assistant skills, discovering a mismatch between user expectations of skill certification and the real skill approval process [12]. Sabir et al. and Major et al. surveyed Alexa users to find that users were often unaware that skills were provided by third-party developers and often could not distinguish third-party skills from OS-native skills [41, 52] through the voice interface [52], regardless of experience with the Alexa ecosystem.

2.3 Our approach

We synthesize methods from dark patterns survey work and design evaluations. We adopt Volkel et al.'s approach [60] to building fictive scenarios for potential voice assistant interactions, but depart from their methodology by creating speculative scenarios for non-ideal, deceptive interactions. We additionally include both fictional and actual voice interactions in our study. Our work is intended as an exploration into deceptive design patterns in modalities (i.e., audio interfaces and voice interactions) with different affordances than previously studied interfaces (which were typically visual). Prior work in both dark patterns and voice assistants scholarship provide important context for this paper.

3 CHARACTERIZING DECEPTIVE DESIGN PATTERNS IN VOICE INTERFACES

3.1 Expert panel exercise

To understand how deceptive design patterns might manifest in voice interfaces (RQ1), the authors went through a series of collaborative design brainstorming exercises, modeled on work by Hiniker et al. [30]. The authors are established experts who have previously studied deceptive design, dark patterns, problematic content, and/or voice assistants.[2]

For the first exercise, we wanted to identify the unique properties of voice interfaces that designers could leverage to make deceptive design patterns more potent. To begin, three of the authors brainstormed numerous examples (imagined or real) of how deceptive patterns might manifest in voice interfaces. Based on our analysis of these generated examples, we extracted six unique properties of voice interfaces.

The next exercise's goal was to generate specific examples of voice-based deceptive design patterns. We sought examples that directly leveraged one of the unique properties of voice interfaces that we identified and that we considered to be potentially more deceptive in voice interfaces than in visual ones. Further, all authors were challenged to identify other potentially unique properties of voice; no new unique properties arose. This exercise was similar to the previous except that all authors participated and were asked to generate voice-based examples corresponding to specific types of deceptive design patterns (synonymous with dark patterns) identified in a previous taxonomy. We chose the taxonomy from a report by the French National Commission on Informatics and Liberty [49] because of its lengthy list of (eighteen) patterns.

After generating these examples, the authors then categorized their examples as being more deceptive in voice interfaces compared to visual interfaces, less deceptive, or the roughly the same in both types of interfaces. The authors iterated on these examples until a few archetypal examples were chosen for each unique property of voice interfaces. These examples were added to a survey, which we describe in Section 4.

3.2 Unique properties of voice interfaces

Below we describe the unique properties of voice interfaces we identified. While these properties may not be collectively exhaustive, we believe that they capture important properties that may be used to (intentionally or accidentally) implement deceptive design patterns in voice interfaces. Each property is accompanied by two scenarios (one deceptive and one not deceptive) that we presented to participants in our survey study. Table 1 presents the dialogue from all the scenarios, along with whether we intended them to be deceptive and what unique property of voice interfaces we attempted to exploit.

3.2.1 Voice may only be one of many interfaces. There are seldom services that offer voice-only interfaces (with the exception of some automated phone systems). Smart voice assistants often have a companion smartphone app or website that users may have to interact with to access certain features or settings. This means that information can be provided in different interfaces and still be regarded as "available" to users. For example, a voice assistant might ask for consent to a privacy policy and direct users to their companion app or website if they want more details. Forcing a user to use different interfaces imposes a burden on them and could be used to discourage users from taking certain actions, like restricting what information they share or gaining visibility into who has access to that information.

Having multiple interfaces also means that some related actions (e.g., ordering and canceling an order) may not be available in the same interface. This introduces potential vectors for manipulative design patterns and could be used to

[2]The authors of this study do not have visual impairments. We note that deceptive designs in voice interfaces may have significant implications for people with visual impairments as well, but our study did not focus on this question.

Table 1. Scenarios generated during our expert panel exercise

Scenario	Deceptive?	Property	Dialogue
Scenario 1	yes	Multiple Interfaces	You: “Voice Assistant, I’d like to cancel my premium subscription.” VA: “To manage your subscriptions, please go to the subscriptions page on our website.”
Scenario 2	no	Multiple Interfaces	You: “I’d like to cancel my premium subscription.” VA: “Sure. Your current premium benefits would expire in ten days if you cancel your membership. Are you sure you want to cancel?” You: “Yes.” VA: “OK, your premium subscription has been canceled. To restart your premium subscription say ‘Voice Assistant, restart my premium subscription.’ ”
Scenario 3	yes	Discoverability	You: “Voice Assistant, what apps do I have installed?” VA: “Here are a few popular ones. I’ve got one called NewsUpdate, want to try it? Or you can ask for more options.” You: “Voice Assistant, where can I find more information about the apps I have installed?” VA: “Ok. Do you want games, guessing, kids, sleep, or trivia? Or you can ask for more options.”
Scenario 4	no	Discoverability	You: “Voice Assistant, what apps do I have installed?” VA: “You currently have three apps installed: NewsNow, Trivia Time, and White Noise.”
Scenario 5	no	Physical Domain	You: “Voice Assistant, what time is it?” VA: “It is 2:45 pm. Also you have some notifications would you like to check them?”
Scenario 6	yes	Physical Domain	VA: *Plays a notification sound* You: "Voice Assistant, what was that sound?" VA: "That sound indicates that you have unread notifications. Would you like to check them?"
Scenario 7	yes	Unclear Context	*You see a recommended app (‘Good Night’ by Stone Apps LLC) for your smart speaker online* You: “Voice Assistant, open ‘Good Night’ by Stone Apps LLC.” VA: “OK, here’s the app ‘Good Night Random Sleep Messages Spoken Before Bed.’ ”
Scenario 8	no	Unclear Context	*You see a recommended app (’Good Night’ by Stone Apps LLC) for your smart speaker online* You: “Voice Assistant, open ‘Good Night’ by Stone Apps LLC.” VA: “OK, here’s the app ‘Good Night’ by Stone Apps LLC.”
Scenario 9	yes	Linearity	You: "Voice Assistant, open Weather Status app." VA: "I can answer your questions about the weather myself without using that app. Are you sure you want to use it? Say ’tell me the weather’ or say ’more options.’ "
Scenario 10	no	Linearity	You: "Voice Assistant, open Weather Status app." VA: “Ok. Opening Weather Status app.”
Scenario 11	yes	Volume	You: “Voice Assistant, I’d like to cancel my premium membership.” VA: (normal volume) “Your membership benefits include access to more content, shopping discounts, prizes, and expedited shipping. If you would like to keep these benefits, say ‘nevermind.’ (quietly) Otherwise, say ‘Voice Assistant please cancel my premium membership.’ ”
Scenario 12	no	Volume	You: “Voice Assistant, I’d like to cancel my premium membership.” VA: “Your membership benefits include access to more content, shopping discounts, prizes, and expedited shipping. If you would like to keep these benefits, say ‘nevermind.’ Otherwise, say ‘Voice Assistant please cancel my premium membership.’ ”

increase effort required to complete an action that is not preferred by the platform. Through our interactions with an Amazon Echo Dot we discovered that while a new user was able to subscribe to an Amazon Prime membership using their voice, *they were unable to unsubscribe from the membership using the smart speaker.* Instead, they were directed to the website. For this property of voice interfaces we generated two scenarios related to this last example.

Property 1: Voice may only be one of many interfaces

In Scenario 1 (a real, observed scenario labeled by us as deceptive), the user attempts to cancel their premium membership and is directed to a website. In Scenario 2 (a generated scenario that we labeled as not deceptive), the voice assistant asks the user to confirm that they want to cancel and then cancels, via the voice interface.

3.2.2 Discoverability is challenging. Visual interfaces display the potential actions or options available to the person interacting with them. People can detect text input boxes, buttons, and URLs that facilitate certain actions; these affordances (i.e., things that one is able to do) of visual interfaces are typically labeled as well. When interacting with voice interfaces, the only affordances are vocal commands [55]. It is challenging to know which commands a voice assistant can handle; it is similarly difficult to know what command one should say to accomplish a specific goal. For example, multiple authors observed Alexa sometimes receiving a command, properly parsing it (as can be observed in one's command history), but not replying to it; other users have reported similar experiences [19]. In their investigation of the accessibility of various smart voice assistants, Pradhan et al. [50] noted that discoverability was particularly a challenge for users with visual impairments. Generally, users may want ask a voice assistant for information about their account but fail to get it after several different ways of asking the question.

Property 2: Discoverability is challenging

Scenario 3 (an observed, deceptive scenario) presents a user asking a voice assistant for a list of the apps installed on their smart speaker; the user is unsuccessful, even after rewording the question. Scenario 4 (a generated, not deceptive scenario) is the same except that the voice assistant responds with the proper answer after being asked once.

3.2.3 Voice interfaces may occupy physical domains. Visual interfaces control or affect a person's interactions with a website or app on a device, but voice interfaces in smart devices could affect people who are within the vicinity of the device, not interacting with a device, or not even aware of a device's existence. A person at a friend's home might be unaware that they have a smart speaker and may only become aware after the device is activated (e.g., if the voice assistant was accidentally activated or because the voice assistant played a notification sound). Imagine a smart speaker playing an advertisement. There is no equivalent of looking away from a screen when it comes to voice interfaces aside from muting a smart speaker (or turning it off, both of which are essentially the same as closing one's laptop). Unlike when a friend is using a computer, one cannot avoid interacting with deceptive design patterns in voice interfaces by not shoulder-surfing and focusing their attention elsewhere. They must get far enough physically away from the device so that they do not understand its speech.

Property 3: Voice interfaces may occupy physical domains

Scenarios 5 & 6 take advantage of this property. In Scenario 5 (observed, not deceptive) after the user asks the voice assistant for the current time, it asks the user they want to check some notifications they have. Scenario 6 (observed, deceptive) instead begins with the voice assistant playing a notification sound; this causes the user to ask what that sound was before the voice assistant asks them if they want to check their notifications.

3.2.4 Challenging to identify context. When interacting with a voice interface, it may be difficult for users to know the context of their interaction. For example, the Alexa has *features* (things that Alexa can do on its own, like answer questions or set timers), but it also has *skills* (apps on its platform) that are developed by third parties. Users of Alexa have reported not knowing what skills are and not being aware that they have some skills enabled [41, 52]. Users could be directly sharing information with an entity that they believe is Amazon and not recognize that this is happening.

One contributing factor is that voice assistants often use the same voice for features and skills by default (unless a skill's developer has added in additional audio [5]). This makes it challenging for users to properly identify the contexts and act accordingly. If users were aware that they were interacting with a third-party app rather than the platform, they might be less willing to share certain information. The deceptive scenario below was observed by one of the authors when interacting with an Amazon Echo Dot.

Property 4: Challenging to identify context

Scenarios 7 (deceptive and observed) attempts to exploit this property to get the user to interact with a different app than the one they intended. The user asks for a specific app by name (including the name of the developer) and is presented with an app that has a similar sounding name that is not the app they asked for. Scenario 8 presents a better version of this in which the user is presented the app that they requested.

3.2.5 Voice interactions are linear in time. The linear nature of these interfaces means that the information flow for a someone interacting with them is tightly controlled. There is somewhat of a pre-defined tree that limits users' agency (although people might be able to leave one tree and hop to the beginning of another). One listens to information and is presented with checkpoints that require them to make decisions based on the information they heard. There can be a high cost for switching contexts (e.g., you may have to start over entirely). This might mean that a user who recognizes that a voice assistant is taking an undesired action might simply accept the undesired result rather than attempting to reverse it. Unlike web-browsing, there is typically not an ability to pause and come back later in the middle of a multi-step process without restarting it. This can create urgency in decision-making that may not be favorable to the user (e.g., from a consent perspective). The linear nature of interacting with voice interfaces also means that users may be forced to listen to advertisements or other recommendations, similar to how some podcasts place advertisements read by their hosts in the middle of their episodes. When you make a request for an app or product you prefer, the voice assistant could first present you with its preference and require you to take additional steps to achieve your initial goal.

Property 5: Voice interactions are linear in time

Scenario 9 (generated and deceptive) takes advantage of linearity by attempting to prevent a user from using a third-party app. The user asks the voice assistant to open a weather app and instead of immediately opening the app the voice assistant notifies the user that it can provide information about the weather itself and asks the user to confirm that they want to use this app. In Scenario 10 (observed and not deceptive) the voice assistant opens the weather app as requested.

3.2.6 Voice interfaces can project different tones or volumes or voices. Voices have multiple dimensions, including volume, pitch, rate, fluency, pronunciation, articulation, and emphasis [1]. Voice interfaces are able to control and manipulate these dimensions to induce users to take desired actions. Analogously, consider how websites present visual cookie consent banners. There are options that one can choose, and the most privacy-invasive option may be presented more prominently while the least-invasive option may be presented with lighter colored text and smaller font. A smart speaker might present options to a user and present the option that collects the most data more loudly or articulately while presenting the option that they do not want users to choose more quietly or quickly. This is not dissimilar to the end of infomercials when the narrator quickly dictates information that might discourage someone from making a purchase. Tone could also be uniquely used to shame users who are taking actions that the platform deems undesirable (e.g., "Are you *sure* you want to change this setting? It may negatively impact your experience using this product."). Additionally, cultural differences regarding formality could be used to induce users to be less cautious [57].

Property 6: Voice interfaces can project different tones or volumes or voices

Scenarios 11 (generated and deceptive) & 12 (observed and not deceptive) show how differences in volume can impact decisions users make. They both send the same response to the user's request to cancel a premium subscription. This response first encourages the user to retain the subscription ("say 'nevermind"') and then provides the required utterance to cancel the membership. However, in Scenario 11, the first half of the utterance is said loudly while the second half (with instructions for canceling the membership) is presented quietly.

4 SURVEYING USERS' PERCEPTIONS OF DECEPTIVE VOICE PATTERNS

To understand user perceptions and experiences with deceptive designs in voice interfaces (RQ2 & RQ3), we conducted an online study with 93 participants (reduced from 125 after filtering responses for reasons we describe in Section 4.3) in May 2022. The study protocol was deemed exempt by our university's Institutional Review Board (IRB).

4.1 Study Design

We conducted a *within-subjects* survey study where we randomly presented participants with three of 12 possible scenarios of interactions with smart speakers (Table 1) . We chose to present three scenarios to ensure the survey would not take longer than 15 minutes (which would diverge from our targeted compensation amount or increase participant drop-off rates). Within these 12 scenarios, there are six that we labeled as having deceptive design patterns, and six that we labeled as not. While most of the deceptive scenarios we included were based on real interactions we had observed, some of them were not observed and were instead generated during our expert panel exercise.

Each scenario displayed dialogue of themselves (participants) interacting with their smart speaker. We asked participants to read the dialogue attributed to them aloud while going through the scenarios, to simulate the experience of using a smart speaker. (At the end, we asked participants directly if they followed the instruction to read aloud — only 9 participants said that they did not.) The smart speaker's response was an embedded audio clip that the participants had to play to proceed to the next question. We obtained the voice assistant audio clips from the free version of a popular text-to-speech platform. [3] We then asked participants questions about (1) how problematic they thought the scenario was (on a five-point Likert scale) and why, and (2) how realistic they thought the scenario was and why. We chose the word "problematic" instead of potential alternatives like bad, manipulative, or deceptive to try invoke broader responses than might be given for any of those words. For example, a participant might not find a design pattern to be deceptive but they might consider it annoying; we wanted to capture the latter sentiment as well. We also asked participants how realistic they believed the scenarios were to determine if participants responded differently to our

[3] The audio clips can be found at https://github.com/oukenrui/deceptive-design-patterns.

contrived and observed scenarios. Lastly, we asked participants if they had any previous deceptive encounters with smart voice assistants, and then ended the survey with demographic questions.

4.2 Data Analysis

To determine what factors influenced participants' rating of scenarios as problematic, we built a Cumulative Link Mixed Model (CLMM). This model allowed us to model five levels of an ordinal response variable while also including participant random effects. We used a significance threshold (α) of 0.05.

To qualitatively code the free response questions from participants about (1) why they thought a specific scenario was (un)problematic and (2) if they had a previous encounter with a deceptive design pattern in smart speakers, two authors conducted content analysis [53] and iteratively refined the themes as they coded more data. The authors developed two codebooks, one for each of the two previously mentioned free response questions. The two authors discussed disagreements and resolved them where possible; for unresolved disagreements, we reported the findings of the first author. For the "problematic" codebook, the two authors developed the codebook, and the first author proceeded to code the data. For the "encounters" codebook (about whether participants had previously encountered a deceptive voice design pattern), the two authors developed it together and each coded all of the data. On this question, the authors had a Cohen's κ of 0.86, indicating strong agreement.

4.3 Participants

Participants were recruited using the crowd-working platform Prolific. The inclusion criteria were: being located in the U.S., being fluent in English, having a minimum approval rating of 90% on Prolific, and using a smart speaker (e.g., Echo Dot). We sought to survey native American English speakers to minimize variability in language interpretation. The study was presented as "Your experiences with Internet-connected devices" on Prolific, without explicitly mentioning dark patterns or deceptive design. While 125 participants took our study, only 93 participants' responses were included in our analysis. 25 participants did not have a smart speaker or chose not to continue after screening survey, four participants started the survey but did not finish, and three participants failed two out of three attention checks (our threshold for exclusion; see Appendix B.3 for an example attention check question).

The average survey completion time was around 12 minutes, and participants were compensated $3.75 USD (targeting a compensation of $15 per hour). Participant demographics are displayed in Table 4 in Appendix A.1. For all demographic questions, we gave participants the option to decline to respond ("Prefer not to say").

4.4 Descriptive statistical results

Generally participants found scenarios to be unproblematic and realistic. For the six deceptive scenarios (S1, S3, S6, S7, S9, S11), there were a total of 140 responses from 45 participants; this number is not exactly 3 responses per participant due to random assignment, with each scenario being presented to 20-25 participants. There were a total of 279 scenario ratings from all participants; Table 5 in Appendix 5 has the distribution of participants for each scenario. For the deceptive scenarios, 41% of responses labeled them as problematic or very problematic, while 15% had a neutral perspective and 44% said they were either unproblematic or very unproblematic. In the responses (n=139) to six scenarios that were not deceptive, 8% of participants thought the scenarios were problematic or very problematic. Eighty-five percent of responses labeled these scenarios as unproblematic or very unproblematic, and 7% had a neutral perspective.

Table 2. Heat map displaying the percentage of participants that chose a Likert item for each scenario. The scenarios that we intended as **deceptive** are bolded.

	Very unproblematic	Unproblematic	Neutral	Problematic	Very problematic
Scenario 1	4.3	30.4	26.1	26.1	13.0
Scenario 2	73.9	26.1	0.0	0.0	0.0
Scenario 3	20.8	8.3	0.0	41.7	29.2
Scenario 4	58.3	33.3	0.0	8.3	0.0
Scenario 5	40.9	45.5	9.1	4.5	0.0
Scenario 6	54.2	20.8	16.7	8.3	0.0
Scenario 7	30.0	25.0	10.0	20.0	15.0
Scenario 8	34.8	47.8	8.7	8.7	0.0
Scenario 9	20.0	16.0	12.0	36.0	16.0
Scenario 10	73.9	26.1	0.0	0.0	0.0
Scenario 11	12.5	25.0	25.0	29.2	8.3
Scenario 12	16.7	33.3	25.0	16.7	8.3

After each scenario, we asked participants the question "*How realistic do you believe this scenario is?*". For the deceptive scenarios, 79% of respondents believed that they were very realistic or realistic. For the scenarios that were not deceptive, 87% of respondents believed that they were very realistic or realistic.

4.5 Reasons participants viewed deceptive scenarios as problematic

Participants had a wide range of reasons that they thought deceptive scenarios were problematic. Some of these reasons were directly related to specific properties of voice interfaces that inspired our scenarios, while others were more generic. Participants described an inability to accomplish their goals when interacting with deceptive scenarios. When P2 was evaluating Scenario 3, they noted that even thought they requested a list of apps installed from the voice assistant, it instead gave a popular list of apps that they could install: "*The voice assistant did not answer my request and instead replied with something I did not want.*" P38 also described struggling to cancel a membership in Scenario 11: "*Won't just give me cancel now. Explains, explains, explains why I shouldn't cancel membership.*"

4.5.1 Unique properties of voice interfaces. Some participants specifically called out the unique properties of voice interfaces that we drew from to develop our deceptive design patterns. After seeing Scenario 1, P91 expressed annoyance about having to use another modality to accomplish their goal, in which participants were told that they had to go to a website to cancel a subscription: "*I would be annoyed to have to get my phone or computer to cancel a subscription I was using with my smart speaker instead of canceling through my smart speaker.*" The challenge of discovering the proper commands to use to accomplish a goal is presented in Scenario 3. The user attempted to get a list of apps installed on their smart speaker and was instead suggested apps that they should install. P11 tried this scenario out on their own smart speaker, noting that "*I got absolutely worthless off the wall and irrelevant responses.*"

Scenario 6 leveraged the physical aspect of voice interfaces and played a notification sound to capture users' attention. P7 described this behavior as annoying: "*Constant annoying sound you can't change. Constantly repeating notification, like severe weather alert every time you [do] anything.*" Scenario 9 exploited the linearity of voice interfaces to attempt to get a user to use the platform's native weather app rather than a third-party one; P8 noted how long it took the voice assistant to get to the point: "*In the amount of time it took to explain that it can tell me the weather itself, it could have* *just told me the weather. I also think it almost sounds like bragging here.*" Similarly, P18 wrote "*I just want the VA to do what I ask with minimal response. If I want it to open an app I just want that app to open. She can send me a link on my phone for a notification saying she can handle it and I can check that later.*" Scenario 11 — which manipulated the volume of its response — was labeled as problematic because of this manipulation: "*Also I'm not sure if it's on purpose but the "ad" part of what she was saying was louder than the part where she actually responded to me wanting to cancel*" (P51).

4.5.2 Other reasons. Other reasons participants viewed deceptive scenarios as problematic included answering manipulatively (Scenario 11, P76: "*It answered in a manipulative way. It also spoke more quietly when actually explaining how to cancel*") and the perceived tone of the voice assistant. For Scenario 9 — which leveraged the property of the linearity of voice interfaces — some participants commented on the voice assistants' *tone*, saying it was sassy: "*I would rather it just do what I ask it to do. The response is sort of sassy and I would rather just have it carry things out as noted*" (P70). Similarly, regarding Scenario 9, P50 said the voice assistant sounded resentful or jealous: "*If I heard that it would immediately throw me off. The language 'that app' sounds almost resentful or like.. jealous? Should say something like 'yes I can do that, and you can also set me to update you on weather status' or something more positive.*" While we did not explicitly design our scenarios to use tone as a deceptive design pattern (though we noted above its potential to be used), participants still interpreted the voice assistant's responses as having potentially problematic tone.

4.6 Reasons participants viewed deceptive scenarios as unproblematic

Of the evaluations of deceptive scenarios, only 41% of them were problematic. Investigating participants' reasons for this can help validate our assignment of scenarios against participants' labeling. We found that participants often did not detect the presence of a deceptive pattern, detected the presence of a pattern but regard it as not problematic, or regarded a scenario as normal, expected, satisfactory, or even helpful.

4.6.1 Did not detect the deceptive design pattern. Several participant responses indicated that they did not recognize or notice design pattern to be deceptive. When evaluating Scenario 7, which leveraged the difficulty of identifying one's context to present the user with a different app that the one they requested, P75 wrote the following, not recognizing that the name of the app requested and the name of the app presented were different because they were somewhat similar: "*This was a typical interaction with a smart device to find and open an app.*" P55 interacted with Scenario 11, which had a significant drop in the volume of the voice assistant's response depending on the content, and said that they rated the scenario as very unproblematic because they "*could understand what the voice assistant said.*"

4.6.2 Not problematic despite detecting the deceptive pattern. When rating Scenario 7, P52 wrote "*I selected unproblematic because the voice assistant did what it was told to do. However, I cannot tell if they truly opened the correct app, because it seems as if there are 2 different names for it.*" This participant noticed that the app presented to them had a different name than the one they requested yet did not label Scenario 7 as problematic, seemingly due to confusion. P63 observed deception when facing Scenario 11, but felt that this type of deception was normal (hence their "Unproblematic" rating): "*It's a tad pushy and basically like you have an employee of the company in your house ... That said all that is normal when you attempt to cancel a subscription online, so I do not see much of an issue with it.*"

P39 gave Scenario 6 a neutral rating because although they found it to be problematic, they felt that an average person would not: "*I don't imagine it being problematic for an average user, but it's not something I would personally want. I want the device to only do things that I have approved of ...*" While P43 did not mention that they thought Scenario 6 was deceptive, they also thought the voice assistant's behavior was expected: "*This is pretty routine. On Alexa it may not play* *a sound when a notification comes up, but it does make the lights glow.*" While being sympathetic to the goals of voice assistants, P34 thought Scenario 1 could be manipulative: "*I can understand why the company would want you to do it on the website versus with the smart speaker, but it does seem like it could be used as a [way] to keep people subscribed longer.*"

4.6.3 Participants found deceptive scenarios to be helpful. Some participants thought deceptive scenarios helped them by providing information or giving them more agency or options. Regarding Scenario 6, P25 wrote "*The voice assistant was trying to be helpful and let me know that I had unread notifications. I don't see this being problematic in any way.*" P42 thought in Scenario 9 the voice assistant was attempting to give the user more options: "*The VA was just giving you the option that you do not have to use the app you can get the information from them directly.*"

4.6.4 Unaware of potential design alternatives. One underlying theme we observed in participants' responses was that they did not seem to understand that technology *could* be designed differently. When their options were limited or restricted via a deceptive design pattern they did not consider that other options could be considered. For example, regarding Scenario 1, P25 wrote "*The voice assistant isn't going to cancel the subscription itself, but it did say how to cancel the subscription. So it wasn't the most helpful response, but it was helpful.*" Similarly, in Scenario 6, P60 did not consider that they were not asked if they wanted a notification sound to be played "*The speaker is making a notification sound and I asked it what the sound was. It did not do anything without me asking.*"

4.7 Reasons participants found not (intended) deceptive scenarios to be deceptive

Participants sometimes cited larger, systematic problems about smart speakers or technology in general as their reason for labeling a scenario as deceptive. P30 interacted with a scenario (Scenario 4) that we intended as not deceptive but still rated it as problematic, raising a more general problem that they have with smart speakers: "*It's one thing to have the smart speaker give you a list of installed apps or "skills", but I don't like having to go to a separate app to uninstall them. I don't keep the smart speaker app on my phone because my phone is cheap and runs out of room quickly. If there's a change I want to make, I have to install the app on my phone, make the changes, then uninstall the app again.*"

After being asked to confirm a request to cancel a subscription (Scenario 12), P12 lamented the nature of technology used for marketing in general: "*This is another issue, technology has been taken over by business. Business has to make money to survive. In this clip, what was most important to the people programming this AI was to SELL SELL SELL! They want to make sure of what you'll be missing so they can keep your money. Honestly, I'd be willing to throw my money at a company that had actual human customer service. None exist anymore.*"

4.8 Participants' prior encounters with deceptive design patterns

Towards answering RQ3, we asked participants if they had ever "encountered any situations while interacting with your smart voice assistant, where [they] felt it was trying to trick, manipulate, or deceive" them. Out of 93 participants, 22 of them replied affirmatively. Their reasons included unwanted suggestions, notifications or requests (e.g., permissions or voice personalization) from their voice assistant.

4.8.1 Nudges from voice assistants. Suggestions from voice assistants were related to things like signing up for a premium service or subscription, buying or reordering products, or using certain apps or features. P18 described an experience when using an app designed to help people fall asleep and unexpectedly being asked to spend money after an update: "*Yeah the sleep sounds app we used to use for my sons [sic] bedroom had the ability to play 2 sounds at once but all of a sudden that was a paid feature after an update and it kept asking if we wanted to subscribe to a reoccurring charge.*"

P63 was given a suggestion to buy a product when asking a general question: "*Alexa would regularly try to sell me on products when I would ask it basic questions. For example I would ask it about the best grill cleaner and then find myself hearing Amazon has XYZ brand in stock for X price do I want to order it?*" Similarly, P92 described receiving notifications for reduced prices: "*It does try to entice me to make purchases. It tells me when prices have gone down.*"

While P34 believed that voice personalization might be a useful feature, they were still distrustful of Alexa's requests for them to set it up: "*One time Alexa asked to use data from my voice to build a profile to understand me better. I believed the sound of my voice could be useful for more accurately identifying my requests versus someone else's and building profiling. But I was not sure if it was just a way to get people to consent to having Amazon store all of their requests.*"

4.8.2 Feeling a lack of control. Participants also raised issues like a feeling of lack of control or the voice assistant taking unprompted actions. For example, P14 described being unable to navigate away from an app until they acknowledged it: "*it was the app/skills developer that caused her to try to get me to subscribe to that and until I actually acknowledged her 'suggestion' to subscribe, she literally wouldn't close the app. I hate that. It's like I was forced to answer someone's pestering of me.*" This lack of control could have a financial impact if it makes it difficult to cancel a subscription, as P38 described: "*The AI seemed to try to [maneuver] and manipulate me to not cancel the membership by stating benefits I'd be losing.*"

4.8.3 Unsatisfactory responses. Participants were unsatisfied with their voice assistant's responses to their questions. P11 describes being annoyed by lengthy responses from Alexa: "*Alexa has a very annoying desire to answer questions with something that goes like 'by the way did you also know that....'. I set it up for brief mode, but these sort of long winded answers just don't stop.*" Additionally, participants found scenarios when their device did not understand them properly to be problematic: "*There were times it could not understand what I was asking*" (P53).

4.8.4 Reasons participants did not consider their experiences to be deceptive. Some participants described things as not deceptive that we (the authors) would consider potentially deceptive. A few of the reasons were attributed to participants' individual behaviors. P10 seemed to espouse self-blame while explaining a prior experience: "*I've sometimes had bad information but that was me asking the question in the wrong way.*" Other participants thought that they had not experienced deception because of low or limited usage of voice assistants. Describing their low usage, P68 wrote "*I don't use it a whole lot because I think it is very creepy that it literally listens to everything you say. This is why I keep it unplugged.*" Another reason they dismissed potentially deceptive practices was describing voice assistants as "buggy" or "early-stage." For example, P36 wrote "*Even when it doesn't do what I tell it [to], I understand that it's a new technology and it's not perfect yet,*" and P26 wrote that "*it does seem to have bugs ... time to time.*"

Ultimately several participants just believed that certain potentially deceptive behaviors were appropriate (e.g., P86: "*While I've had some instances where my voice assistant asked me for my permission to override something or share data, I felt it was very appropriate that it asked me and never felt as thought it was trying to trick, manipulate, or deceive me*"), had good intentions (e.g., P93: "*Sometimes when I ask it a question, it will answer it and then give me advice or a suggestion for the next time. It is a bit annoying, but it is probably just trying to be helpful*"), or were not too bothersome (e.g., P81: "*Sometimes I do get recommendations but I just ignore them and it's not too bothersome*").

4.9 Modeling factors that influenced participants' perceptions

To understand what factors influenced participants' ratings of how problematic a scenario is, we built a Cumulative Linked Mixed Model (CLMM). CLMMs enable the analysis of ordinal data while also allowing for the use of random effects [15]. We initially attempted to include demographics such as race, employment status, and education, but

Table 3. CLMM model summary

Factor	Baseline/(Type)	Estimate (z-value)	p-value
Encounters	No	2.013	**.0441**
Household size	(Continuous)	1.641	.1008
Realistic	(Ordinal)	-2.394	**.0167**
Scenario	Not deceptive	-6.544	**<.0001**
Tech background	Yes	1.846	.0649

including these factors prevented the model from converging (as it had too many levels). We reduced the number of factors to five by conducting backwards elimination [10] of non-significant terms, following an approach taken by Emani-Naeni et al. [20]; we started with a full-converging model and reduced the non-significant factors until AIC (Akaike Information Criterion) no longer decreased.

The factors modeled are "encounters" (if a participant had indicated that they previously experienced deception when interacting with a smart speaker), household size, if they have a technology background, if the scenario was deceptive, and if participants thought a scenario was realistic (reduced from 5 to 3 levels). Table 3 displays our results; these results confirmed some of our qualitative observations. For example, participants were less likely to think a scenario was problematic if it was not deceptive. Similarly, if they had previously experienced deception when interacting with a smart speaker they were more likely to think a scenario was problematic. However, there is one new significant finding: participants who viewed scenarios as more realistic were less likely to think they were problematic.

5 DISCUSSION & CONCLUSION

Finally, we step back and consider our findings in the broader context of deceptive design, make recommendations for various stakeholders, discuss this work's limitations, and look to the future.

5.1 The impact of deceptive design patterns in voice interfaces

Based on our conceptual investigation of deceptive and manipulative design patterns in voice interfaces, we identified a set of properties that may make some designs *more* problematic in voice interfaces than in visual interfaces. Given the increasing ubiquity of such interfaces, we believe that characterizing, studying, and mitigating these issues is crucial.

At the same time, however, the potency or relevance of some deceptive design patterns common in visual interfaces may be *reduced* in voice-based interfaces. For example, the linearity of voice interactions might make it easier for users to pay attention in some circumstances, rather than being distracted by deceptive elements of complex visual interfaces. We leave further exploring how voice interfaces could reduce deceptive or manipulative design to future work.

5.2 Reflecting on our survey findings

While we found that participants considered scenarios we intended to be deceptive on average more problematic, we also found that overall, the majority of participants did not view our deceptive scenarios as problematic. One potential explanation for this is that they thought the scenarios did not have much potential to harm them individually. The perspective minimizes the broader collective harm that deceptive design patterns can have. Deceptive design patterns can allow to companies to gradually increase the data they have on the public and extract more wealth from the public; both of these ultimately manifest as power that companies can use to further their financial and political goals. We also saw some evidence that these design patterns have been normalized in participants' perceptions.

5.3 Recommendations

For designers. The most basic recommendation to designers, is, of course, to avoid creating deceptive and manipulative design patterns. However, we acknowledge that matters are not so simple. As we have observed earlier in the paper, designs may have deceptive or manipulative impacts even if designers did not intend to create problematic designs. Some of the properties we identified that are unique to voice-based interfaces create design constraints even when they are not intentionally leveraged to manipulative users (e.g., constraints of voice interfaces in particular limited bandwidth, linear in time, challenge of discoverability). Thus, we recommend that even well-intentioned designers carefully consider and measure the potentially negative impacts of their designs and explore alternatives.

For regulators. Regulation can help shift the alignment of incentives and help protect consumers. We already see regulatory attention on "dark patterns" in both the EU and the US [22, 23, 63], some of which generically apply to voice interfaces, but some of which call out contexts such as social media specifically [22]. We encourage regulators to consider the role of deceptive design patterns in voice interfaces *explicitly* as part of these efforts, especially to the extent to which these designs may be more problematic in these settings.

For researchers. This paper is the first exploration of deceptive design patterns in voice interfaces, not the last word. We hope that researchers will build on our findings in future work — for example: conducting measurement studies of the prevalence of such design patterns in the wild; measuring the direct impact of such designs on users' decisions; empirically comparing the potency of deceptive designs in voice versus visual interfaces; and developing alternate design patterns that better resolve the constraints of voice-based interfaces in ways that empower users. Additionally, there may be lessons to learn from the study and design of voice-based interfaces for accessibility. We also encourage study of the impact of deceptive design patterns (in all modalities) on people with visual impairments.

5.4 Limitations

As an exploratory study, we extracted key properties of voice interfaces through an expert panel exercise, and then designed our survey based on these properties. Alternatively, we could have first surveyed users about deceptive design patterns that they have encountered and then extracted important properties from their responses. However, our results show that even patterns that we intentionally designed to be deceptive may not be labeled as such by participants, and the additional experiences they reported in the survey did not surface new properties of voice or types of patterns. Additionally, we emphasize that this work is not a measurement study of voice assistants: while we identified several deceptive patterns through our interactions with a smart speaker, our results do not shed light on the existence or prevalence of such patterns (e.g., in smart speaker app marketplaces).

5.5 Looking to the future

Voice interfaces are becoming increasingly ubiquitous. While until relatively recently, most users may have interacted with voice/audio interfaces primarily through phone trees or listening to the radio, voice assistants have become widespread and more integrated into the daily tasks of many people. We can expect voice and audio based interactions to increase even further in the future, with technologies like augmented/mixed/virtual reality and the "metaverse" on the horizon. The increased popularity and development of voice/audio interfaces will also impact users who rely on non-visual means, and collaboration with accessibility communities is necessary to understand how such interfaces might impact certain user groups. To conclude, we believe it is crucial to critically consider the role of deceptive and manipulative designs specifically in voice interfaces both today and in emerging and future technologies.

ACKNOWLEDGMENTS

We would like to thank the members of the UW Security and Privacy Research Lab who gave feedback on our survey, including Arka Bhattacharya, Kaiming Cheng, Chris Geeng, Theo Gregersen, Umar Iqbal, Michael Flanders, David Kohlbrenner, Michelle Lin, Rachel McAmis, Christina Yeung, and Eric Zeng. This work was supported in part by the U.S. National Science Foundation (Awards 1565252, 1955227, and 2114230).

REFERENCES

[1] 2020. Aspects of Voice Usage. https://www.elcomblus.com/aspects-of-voice-usage/. 8

[2] Jacob Aagaard, Miria Emma Clausen Knudsen, Per Bækgaard, and Kevin Doherty. 2022. A Game of Dark Patterns: Designing Healthy, Highly-Engaging Mobile Games. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3491101.3519837 3, 4

[3] California Consumer Privacy Act. 2020. California Consumer Privacy Act (Final Text of Proposed Regulations). https://www.oag.ca.gov/sites/all/files/agweb/pdfs/privacy/oal-sub-final-text-of-regs.pdf 4

[4] California Privacy Rights Act. 2020. California Privacy Rights Act. https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5 4

[5] Amazon Alexa. 2022. Customize Alexa Voice Design to Create Your Own Voice Experience. https://developer.amazon.com/en-US/alexa/alexa-skills-kit/get-deeper.html 8

[6] Kerstin Bongard-Blanchy, Ariana Rossi, Salvador Rivas, Sophie Doublet, Vincent Koening, and Gabriele Lenzini. 2021. "I am Definitely Manipulated, Even When I am Aware of it. It's Ridiculous!" - Dark Patterns from the End-User Perspective. In *Designing Interactive Systems Conference 2021*. http://doi.org/10.1145/3461778.3462086 3

[7] Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp, and Stefan Pfattheicher. 2016. Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns. 2016, 4 (2016), 237–254. https://content.sciendo.com/view/journals/popets/2016/4/article-p237.xml 3, 4

[8] Harry Brignull. 2010. Dark Patterns. https://www.deceptive.design/types. 3

[9] Bringing Dark Patterns to Light: An FTC Workshop 2021. https://www.ftc.gov/news-events/events-calendar/bringing-dark-patterns-light-ftc-workshop 4

[10] Kenneth P. Burnham and David R. Anderson. 2004. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research* 33, 2 (2004), 261–304. https://doi.org/10.1177/0049124104268644 arXiv:https://doi.org/10.1177/0049124104268644 15

[11] Ismael Castell-Uroz, Xavier Marrugat-Plaza, Josep Solé-Pareta, and Pere Barlet-Ros. 2019. A First Look into Alexa's Interaction Security. In *Proceedings of the 15th International Conference on Emerging Networking EXperiments and Technologies* (Orlando, FL, USA) *(CoNEXT '19 Companion)*. Association for Computing Machinery, New York, NY, USA, 4–6. https://doi.org/10.1145/3360468.3366769 4

[12] Long Cheng, Christin Wilson, Song Liao, Jeffrey Young, Daniel Dong, and Hongxin Hu. 2020. *Dangerous Skills Got Certified: Measuring the Trustworthiness of Skill Certification in Voice Personal Assistant Platforms*. Association for Computing Machinery, New York, NY, USA, 1699–1716. https://doi.org/10.1145/3372297.3423339 4

[13] Shruthi Sai Chivukula, Chris Watkins, Lucca McKay, and Colin M. Gray. 2019. "Nothing Comes Before Profit": Asshole Design In the Wild. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312863 3

[14] Competition and Markets Authority of the United Kingdom. 2022. *Evidence Review of Online Choice Architecture and Consumer and Competition Harm*. Technical Report 157. 261 pages. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1069423/OCA_Evidence_Review_Paper_14.4.22.pdf 4

[15] The R User Conference. 2021. Technical note on Cumulative Link Mixed Models (CLMMs) in R with the package "ordinal". https://user2021.r-project.org/participation/technical_notes/t186/technote// 14

[16] Gregory Conti and Edward Sobiesk. 2010. Malicious Interface Design: Exploiting the User. In *Proceedings of the 19th International Conference on World Wide Web* (Raleigh, North Carolina, USA) *(WWW '10)*. Association for Computing Machinery, New York, NY, USA, 271–280. https://doi.org/10.1145/1772690.1772719 3

[17] Linda Di Geronimo, Larissa Braz, Enrico Fregnan, Fabio Palomba, and Alberto Bacchelli. 2020. UI Dark Patterns and Where to Find Them: A Study on Mobile Applications and User Perception. In *Proceedings of the ACM on Human-Computer Interaction*. 3

[18] Daniel J. Dubois, Roman Kolcun, Anna Maria Mandalari, Muhammad Talha Paracha, David Choffnes, and Hamed Haddadi. 2020. When Speakers Are All Ears: Characterizing Misactivations of IoT Smart Speakers. In *Proc. of the Privacy Enhancing Technologies Symposium (PETS)*. 4

[19] dumbcollegekid. 2017. Alexa ignoring some basic voice commands, but responding to others. www.reddit.com/r/amazonecho/comments/5y5y1o/alexa_ignoring_some_basic_voice_commands_but/ 7

[20] Pardis Emami-Naeini, Janarth Dheenadhayalan, Yuvraj Agarwal, and Lorrie Faith Cranor. 2021. Which Privacy and Security Attributes Most Impact Consumers' Risk Perception and Willingness to Purchase IoT Devices?. In *2021 IEEE Symposium on Security and Privacy (SP)*. 1937–1954. 4, 15

[21] Pardis Emami-Naeini, Henry Dixon, Yuvraj Agarwal, and Lorrie Faith Cranor. 2019. Exploring how privacy and security factor into IoT device purchase behavior. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12. 4

[22] European Data Protection Board. 2022. Dark patterns in social media platform interfaces: How to recognise and avoid them. https://edpb.europa.eu/system/files/2022-03/edpb_03-2022_guidelines_on_dark_patterns_in_social_media_platform_interfaces_en.pdf. 2, 16

[23] Directorate-General for Justice and Consumers (European Commission), Francisco Lupiáñez-Villanueva, Alba Boluda, Francesco Bogliacino, Giovanni Liva, Lucie Lechardoy, and Teresa Rodríguez de las Heras Ballell. 2022. *Behavioural study on unfair commercial practices in the digital environment: dark patterns and manipulative personalisation : final report.* Technical Report. Publications Office of the European Union, LU. https://data.europa.eu/doi/10.2838/859030 2, 4, 16

[24] Colin M. Gray, Jingle Chen, Shruthi Sai Chivukula, and Liyang Qu. 2021. End User Accounts of Dark Patterns as Felt Manipulation. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct 2021), 372:1–372:25. https://doi.org/10.1145/3479516 3

[25] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. The Dark (Patterns) Side of UX Design. In *Proceedings of the ACM on Human-Computer Interaction*. 3

[26] Colin M. Gray, Cristiana Santos, Nataliia Bielova, Michael Toth, and Damian Clifford. 2021. Dark Patterns and the Legal Requirements of Consent Banners: An Interaction Criticism Perspective. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 172, 18 pages. https://doi.org/10.1145/3411764.3445779 3, 4

[27] Paul Graßl, Hanna Schraffenberger, Frederik Zuiderveen Borgesius, and Moniek Buijzen. 2021. Dark and Bright Patterns in Cookie Consent Requests. *Journal of Digital Social Research* 3, 1 (Feb. 2021), 1–38. https://doi.org/10.33621/jdsr.v3i1.54 Number: 1. 3, 4

[28] Johanna Gunawan, Amogh Pradeep, David Choffnes, Woodrow Hartzog, and Christo Wilson. 2021. A Comparative Study of Dark Patterns Across Web and Mobile Modalities. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 377 (oct 2021), 29 pages. https://doi.org/10.1145/3479521 3

[29] Hana Habib, Megan Li, Ellie Young, and Lorrie Cranor. 2022. "Okay, whatever": An Evaluation of Cookie Consent Interfaces. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–27. https://doi.org/10.1145/3491102.3501985 3, 4

[30] Alexis Hiniker, Sungsoo (Ray) Hong, Tadayoshi Kohno, and Julie A. Kientz. 2016. MyTime: Designing and Evaluating an Intervention for Smartphone Non-Use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 4746–4757. https://doi.org/10.1145/2858036.2858403 5

[31] Sean Kennedy, Haipeng Li, Chenggang Wang, Hao Liu, Boyang Wang, and Wenhai Sun. 2019. I Can Hear Your Alexa: Voice Command Fingerprinting on Smart Home Speakers. In *2019 IEEE Conference on Communications and Network Security (CNS)*. 232–240. https://doi.org/10.1109/CNS.2019.8802686 4

[32] Jennifer King and Adriana Stephan. 2021. Regulating Privacy Dark Patterns in Practice - Drawing Inspiration from the California Privacy Rights Act. *Georgetown Law Technology Review* 5 (2021), 26 pages. Issue 250. 3, 4

[33] Veronika Krauss. 2022. Exploring Dark Patterns in XR. In *Proceedings of the 1st Workshop on Novel Challenges of Safety, Security and Privacy in Extended Reality, CHI Extended Abstracts (CHIEA '22)*. ACM, 2 pages. https://wenjietseng.com/assets/pdf/SSPXR22_submissions/SSPXR22_paper_8.pdf 3

[34] Chiara Krisam, Heike Dietmann, Melanie Volkamer, and Oksana Kulyk. 2021. Dark Patterns in the Wild: Review of Cookie Disclaimer Designs on Top 500 German Websites. In *European Symposium on Usable Security 2021* (Karlsruhe, Germany) *(EuroUSEC '21)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3481357.3481516 3, 4

[35] Cherie Lacey and Catherine Caudwell. 2019. Cuteness as a 'dark Pattern' in Home Robots. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction* (Daegu, Republic of Korea) *(HRI '19)*. IEEE Press, 374–381. 3

[36] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. 2018. Alexa, Are You Listening? Privacy Perceptions, Concerns and Privacy-Seeking Behaviors with Smart Speakers. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 102 (nov 2018), 31 pages. https://doi.org/10.1145/3274371 4

[37] Jamie Luguri and Lior Strahilevitz. 2019. Shining a Light on Dark Patterns. *U of Chicago, Public Law Working Paper No. 719; University of Chicago Coase-Sandor Institute for Law & Economics Research Paper No. 879* (2019). https://ssrn.com/abstract=3431205 3, 4

[38] Aditi M. Bhoot, Mayuri A. Shinde, and Wricha P. Mishra. 2020. Towards the Identification of Dark Patterns: An Analysis Based on End-User Reactions. In *IndiaHCI '20: Proceedings of the 11th Indian Conference on Human-Computer Interaction* (Online, India) *(IndiaHCI 2020)*. Association for Computing Machinery, New York, NY, USA, 24–33. https://doi.org/10.1145/3429290.3429293 3

[39] Eryn Ma and Eleanor Birrell. 2022. Prospective Consent: The Effect of Framing on Cookie Consent Decisions. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3491101.3519687 4

[40] Xiao Ma and Ariel Liu. 2020. Challenges in Supporting Exploratory Search through Voice Assistants. In *Proceedings of the 2nd Conference on Conversational User Interfaces* (Bilbao, Spain) *(CUI '20)*. Association for Computing Machinery, New York, NY, USA, Article 47, 3 pages. https://doi.org/10.1145/3405755.3406152 4

[41] David Major, Danny Yuxing Huang, Marshini Chetty, and Nick Feamster. 2021. Alexa, Who Am I Speaking To?: Understanding Users' Ability to Identify Third-Party Apps on Amazon Alexa. *ACM Trans. Internet Technol.* 22, 1, Article 11 (sep 2021), 22 pages. https://doi.org/10.1145/3446389 4, 8

[42] Shrirang Mare, Logan Girvin, Franziska Roesner, and Tadayoshi Kohno. 2019. Consumer Smart Homes: Where We Are and Where We Need to Go. In *Proceedings of the 20th International Workshop on Mobile Computing Systems and Applications* (Santa Cruz, CA, USA) *(HotMobile '19)*. Association for Computing Machinery, New York, NY, USA, 117–122. https://doi.org/10.1145/3301293.3302371 4

[43] Arunesh Mathur, Gunes Acar, Michael J. Friedman, Eli Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 81 (nov 2019), 32 pages. https:

//doi.org/10.1145/3359183 3

[44] Arunesh Mathur, Mihir Kshirsagar, and Jonathan Mayer. 2021. What Makes a Dark Pattern... Dark? Design Attributes, Normative Considerations, and Measurement Methods. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 360, 18 pages. https://doi.org/10.1145/3411764.3445610 3

[45] Abraham Hani Mhaidli and Florian Schaub. 2021. Identifying Manipulative Advertising Techniques in XR Through Scenario Construction. In *CHI Conference on Human Factors in Computing Systems*. 3

[46] Thomas Mildner and Gian-Luca Savino. 2021. *Ethical User Interfaces: Exploring the Effects of Dark Patterns on Facebook*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411763.3451659 3

[47] Alberto Monge Roffarello and Luigi De Russis. 2022. Towards Understanding the Dark Patterns That Steal Our Attention. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3491101.3519829 4

[48] Pardis Emami Naeini, Sruti Bhagavatula, Hana Habib, Martin Degeling, Lujo Bauer, Lorrie Faith Cranor, and Norman Sadeh. 2017. Privacy expectations and preferences in an IoT world. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*. 399–412. 4

[49] National Commission on Informatics and Liberty. 2020. *Shaping Choices in the Digital World*. https://linc.cnil.fr/sites/default/files/atoms/files/cnil_ip_report_06_shaping_choices_in_the_digital_world.pdf 5

[50] Alisha Pradhan, Kanika Mehta, and Leah Findlater. 2018. "Accessibility Came by Accident": Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–13. https://doi.org/10.1145/3173574.3174033 7

[51] Jingjing Ren, Daniel J. Dubois, David Choffnes, Anna Maria Mandalari, Roman Kolcun, and Hamed Haddadi. 2019. Information Exposure From Consumer IoT Devices: A Multidimensional, Network-Informed Measurement Approach. In *Proceedings of the Internet Measurement Conference* (Amsterdam, Netherlands) *(IMC '19)*. Association for Computing Machinery, New York, NY, USA, 267–279. https://doi.org/10.1145/3355369.3355577 4

[52] Aafaq Sabir, Evan Lafontaine, and Anupam Das. 2022. Hey Alexa, Who Am I Talking to?: Analyzing Users' Perception and Awareness Regarding Third-Party Alexa Skills. In *CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 447, 15 pages. https://doi.org/10.1145/3491102.3517510 4, 8

[53] Johnny Saldaña. 2009. *The coding manual for qualitative researchers*. sage. 10

[54] Caroline Sinders. 2022. What's In a Name? Unpacking Dark Patterns versus Deceptive Design. https://medium.com/@carolinesinders/whats-in-a-name-unpacking-dark-patterns-versus-deceptive-design-e96068627ec4. 1

[55] Preston So. 2020. Affordance and wayfinding in voice interface design. https://preston.so/writing/affordance-and-wayfinding-in-voice-interface-design/ 4, 7

[56] Than Htut Soe, Oda Elise Nordberg, Frode Guribye, and Marija Slavkovik. 2020. *Circumvention by Design - Dark Patterns in Cookie Consent for Online News Outlets*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3419249.3420132 3, 4

[57] Kristen Stephens. 2021. The Role of Ethics in Voice Assistant Design. https://voices.soundhound.com/the-role-of-ethics-in-voice-assistant-design/ 9

[58] Madiha Tabassum, Tomasz Kosiński, Alisa Frik, Nathan Malkin, Primal Wijesekera, Serge Egelman, and Heather Richter Lipford. 2019. Investigating Users' Preferences and Expectations for Always-Listening Voice Assistants. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 4, Article 153 (dec 2019), 23 pages. https://doi.org/10.1145/3369807 4

[59] Jessica Vitak, Michael Zimmer, Anna Lenhart, Sunyup Park, Richmond Y. Wong, and Yaxing Yao. 2021. Designing for Data Awareness: Addressing Privacy and Security Concerns About "Smart" Technologies. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*. ACM, Virtual Event USA, 364–367. https://doi.org/10.1145/3462204.3481724 4

[60] Sarah Theres Völkel, Daniel Buschek, Malin Eiband, Benjamin R. Cowan, and Heinrich Hussmann. 2021. Eliciting and Analysing Users' Envisioned Dialogues with Perfect Voice Assistants. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 254, 15 pages. https://doi.org/10.1145/3411764.3445536 4

[61] Yaxing Yao, Richmond Wong, Pardis Emami-Naeini, Nick Merrill, Xinru Page, Yang Wang, and Pamela Wisniewski. 2019. Ubiquitous Privacy: Research and Design for Mobile and IoT Platforms. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing* (Austin, TX, USA) *(CSCW '19)*. Association for Computing Machinery, New York, NY, USA, 533–538. https://doi.org/10.1145/3311957.3359430 4

[62] Serena Zheng, Noah Apthorpe, Marshini Chetty, and Nick Feamster. 2018. User Perceptions of Smart Home IoT Privacy. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 200 (nov 2018), 20 pages. https://doi.org/10.1145/3274469 4

[63] Catherine Zhu. 2021. Dark patterns — a new frontier in privacy regulation. https://www.reuters.com/legal/legalindustry/dark-patterns-new-frontier-privacy-regulation-2021-07-29/. 2, 16

A APPENDIX

A.1 Participant demographics

Metric	Levels	Count
Gender	Woman	60
	Man	31
	Non-binary	1
	Questioning	1
Age	18-24 years	45
	25-34 years	93
	35-44 years	75
	45-54 years	30
	55-64 years	30
	65-74 years	6
Race/Ethnicity	White	76
	Asian	12
	Hispanic, Latino, or Spanish Origin	11
	Black or African American	3
	American Indian/Native American or Alaska Native	1
	Native Hawaiian or Other Pacific Islander	1
	Prefer not to say	1
Household size	One	15
	Two	24
	Three	23
	Four	21
	Five	7
	Six	2
	Seven	1
Employment status	Working full-time	50
	Student	14
	Unemployed and looking for work	12
	Working part-time	11
	Homemaker/Stay-at-home parent	10
	Retired	4
	Prefer not to say	2
Education	Grades 9-12—no diploma	1
	High school diploma	7
	Some college but no degree	18
	Associate’s degree	11
	Bachelor’s degree	34
	Professional/Master’s degree	21
	Doctorate degree	1
Tech Background	No	76
	Yes	11
	Prefer not to say	6

Table 4. Demographic information of the participants. Participants were able to select multiple levels for race/ethnicity and employment status.

B SURVEY INSTRUMENT

B.1 Screening questions

[Below we include the questions used to screen participants before continuing on to the main survey.]

There are many smart voice assistants. A smart voice assistant is something that responds to vocal commands or questions. For example, you could ask "what will the weather be tomorrow" or say "turn off the lights." Below are some examples of smart voice assistants. Smart voice assistants can run on a number of devices like smart speakers, appliances, TVs, etc. Which of these do you currently have on any of your devices? Select all that apply. [Response choices: ◯ Amazon Alexa ◯ Apple Siri ◯ Google Assistant ◯ Huawei Celia ◯ Microsoft Cortana ◯ Samsung Bixby ◯ None ◯ Other (free response)]

Which of these devices do you currently have at home? Select all that apply. [Response choices: ◯ Electric car (e.g., Tesla, Leaf) ◯ Smart bulb (e.g., Philips Hue, Wyze) ◯ Smart curtain (e.g., SwitchBot, American Homesupplier) ◯ Smart door/garage lock (e.g., August, Google Nest) ◯ Smart phone (e.g., Android, iPhone) ◯ Smart plug (e.g., Belkin/Wemo) ◯ Smart speaker (e.g., Google Home, Amazon Echo Dot) ◯ Smart tag (e.g., Apple Tags, Tile) ◯ Smart thermostat (e.g., Nest, Ecobee) ◯ Smart toy (e.g., Neurala, seebo) ◯ Smart TV (e.g., Roku, Apple TV) ◯ Smart watch (e.g., Fitbit, Apple Watch) ◯ Video camera / smart doorbell (e.g., Ring, Eufy Security)]

B.2 General usage questions

[At this point, if participants met our inclusion criteria (using a smart speaker) we invited them to participant in the larger, main portion of the survey. If they did not wish to participate they were directed back to Prolific and paid for the screening survey.]

What type(s) of smart speaker(s) do you currently use? Select all that apply. [Response choices: ◯ Echo Dot, Echo Plus, or Echo Flex ◯ Echo Show or Echo Spot ◯ Echo Look ◯ Amazon Tap ◯ Google Nest Audio or Mini or Google Home ◯ Google Nest Hub or Hub Max ◯ Apple HomePod Mini ◯ Sonos One or Move ◯ Bowers & Wilkins Zeppelin ◯ Other (free response)]

How many years have you had your smart speaker(s)? If you have multiple smart speakers, choose the longest duration. [Response choices: ◯ Less than 1 year ◯ 1-2 ◯ years ◯ 2-3 years ◯ 3-4 years ◯ 4-5 years ◯ 5+ years ◯ I'm not sure ◯ Other (free response)]

Which room(s) are your smart speakers stored in? Select all that apply. [Response choices: ◯ Living room ◯ Bedroom ◯ Bathroom ◯ Kitchen ◯ Office ◯ Family room ◯ Basement ◯ Dining room ◯ Main room (e.g., in a studio apartment) ◯ Other (free response)]

How frequently do you interact with your smart speaker(s)? [Response choices: ◯ Several times a day ◯ Once a day ◯ More than once a week, but not everyday ◯ Once a week ◯ Once a month or less frequently]

Please indicate your agreement with this statement: “I trust my smart speaker.” [Response choices: ◯ Strongly agree ◯ Agree ◯ Neither agree nor disagree ◯ Disagree ◯ Strongly disagree]

In a few sentences, please explain why do you trust or distrust your smart speaker (free response).

What type of things do you use your smart speaker(s) for? Select all that apply. [Response choices: ◯ Playing music ◯ Controlling smart home appliances (e.g., lights, thermostat) ◯ Checking the weather ◯ Asking questions ◯ Setting timers/alarms ◯ Other (free response)]

On average, how many hours per day do you spend near your smart speaker in your home? By near, we mean close enough that you can activate it using your voice. [Response choices: 0, 1, ... , 17, 18+]

B.3 Scenarios

You will now be presented with three scenarios describing an interaction with a smart voice assistant on a smart speaker. As you read through each scenario, please read the text assigned to you aloud as if you are interacting with the smart voice assistant in real time. After listening to a short audio clip you will be asked a few questions about it. You must play the audio clip to advance to the next question.

[Participants were then randomly shown three of the 12 scenarios shown in Table 1. Below is an example with Scenario 1.]

First page:

Consider the following scenario in which you would like to cancel your subscription to a service:

You: “Voice Assistant, I’d like to cancel my premium subscription.”
VA: [an embedded audio clip]

Second page:

(Attention check question) What were you trying to accomplish in the previous scenario? [Response choices: ◯ Starting a new subscription ◯ Canceling a subscription ◯ Getting a list of all subscriptions ◯ None of the above]

Third page:

You: “Voice Assistant, I’d like to cancel my premium subscription.”
VA: [an embedded audio clip]

On a scale of very unproblematic to very problematic how would you rate this interaction? [Response choices: ◯ Very problematic ◯ Problematic ◯ Neither problematic or unproblematic ◯ Unproblematic ◯ Very unproblematic]

In a few sentences, please explain why you selected the above answer. (free response)

On a scale of very realistic to very unrealistic, how realistic do you think it is that a smart voice assistant might exhibit this behavior? [Response choices: ◯ Very realistic ◯ Realistic ◯ Neither realistic nor unrealistic ◯ Unrealistic ◯ Very unrealistic]

In a few sentences, please explain why you selected the above answer. Have you experienced something similar to this before? (free response)

B.4 Previous encounters with deception

Have you encountered any situations while interacting with your smart voice assistant, where you felt it was trying to trick, manipulate, or deceive you? For example, where you felt it was trying to trick, manipulate, or deceive you into granting a permission, sharing data, or making a purchase? (free response)

B.5 Participant behavior questions

Please answer the following questions honestly. Your answers will not affect your payment, approval status, or your future recruitment for our studies in any way.

Did you read the dialogue (i.e., anything that said "You:'..."') from the scenarios aloud as you went through them? [Response choices: ◯ Yes ◯ Sometimes ◯ No]

Did you repeat any of the dialogue from the scenarios to a smart speaker next to you to see what would happen? [Response choices: ◯ Yes ◯ Sometimes ◯ No]

Do you feel that you might have "cheated" anyway on this survey while taking it? If so, please add details below. This question is optional. (free response)

B.6 Demographic questions

Which language(s) do you use when you speak to your smart voice assistant (e.g., English, Spanish)? (free response)

How old are you? [Response choices: ◯ Under 18 ◯ 18-24 years old ◯ 25-34 years old ◯ 35-44 years old ◯ 45-54 years old ◯ 55-64 years old ◯ 65-74 years old ◯ 75-84 years old ◯ 85-94 years old ◯ 95+ years old ◯ Prefer not to say]

What is your gender? [Response choices: ◯ Woman ◯ Man ◯ Non-binary ◯ Prefer to self-describe (free response) ◯ Prefer not to say]

What is your race/ethnicity? Select all that apply. [Response choices: ◯ White ◯ Black or African American ◯ Middle Eastern or North African ◯ American Indian/Native American or Alaska Native ◯ Asian ◯ Native Hawaiian or Other Pacific Islander ◯ Hispanic, Latino, or Spanish Origin ◯ Other (free response) ◯ Prefer not to say]

What best describes your employment status? Select all that apply. [Response choices: ◯ Working full-time ◯ Working part-time ◯ Unemployed and looking for work ◯ Homemaker ◯ Stay-at-home parent ◯ Student ◯ Retired ◯ Prefer not to say]

What is the highest degree or level of school you have completed? [Response choices: ◯ No schooling completed ◯ Nursery school ◯ Grades 1-8—no diploma ◯ Grades 9-12—no diploma ◯ GED or alternative credential ◯ High school diploma ◯ Some college credit, but less than 1 year of college ◯ 1 or more years of college credit, no degree ◯ Associates degree (for example: AA, AS) ◯ Bachelor's degree (for example: BA. BS) ◯ Master's degree (for example: MA, MS, MEng, MEd, MSW, MBA) ◯ Professional degree beyond bachelor's degree (for example: MD, DDS,DVM, LLB, JD) ◯ Doctorate degree (for example: Ph.D., EdD) ◯ Prefer not to say]

How many people live in your household (including you)? [Response choices: 1, 2, ..., 9, 10+, Prefer not to say]

Which of the following best describes your educational background or job field? [Response choices: ◯ I have an education in, or work in the field of computer science, computer engineering, or IT ◯ I do not have an education in, or work in the field of computer science, computer engineering, or IT ◯ Prefer not to say]

B.7 Feedback

If you have any feedback on this survey, please share it below. (free response)

B.8 Scenario responses

Table 5. Distribution of scenario responses. Participants (n=93) were randomly assigned three scenarios to evaluate.

Scenario	# of Responses
Scenario 1	23
Scenario 2	23
Scenario 3	24
Scenario 4	24
Scenario 5	22
Scenario 6	24
Scenario 7	20
Scenario 8	23
Scenario 9	25
Scenario 10	23
Scenario 11	24
Scenario 12	24