Are You Going to Answer That? Measuring User Responses to Anti-Robocall Application Indicators

Imani N. Sherman, Jasmine D. Bowers, Keith McNamara Jr., Juan E. Gilbert, Jaime Ruiz, Patrick Traynor
shermani, jdbowers, kmcnamara1, juan, jaime.ruiz, traynor @ufl.edu
University of Florida

***Abstract*—Robocalls are inundating phone users. These automated calls allow for attackers to reach massive audiences with scams ranging from credential hijacking to unnecessary IT support in a largely untraceable fashion. In response, many applications have been developed to alert mobile phone users of incoming robocalls. However, how well these applications communicate risk with their users is not well understood. In this paper, we identify common real-time security indicators used in the most popular anti-robocall applications. Using focus groups and user testing, we first identify which of these indicators most effectively alert users of danger. We then demonstrate that the most powerful indicators can reduce the likelihood that users will answer such calls by as much as 43%. Unfortunately, our evaluation also shows that attackers can eliminate the gains provided by such indicators using a small amount of target-specific information (e.g., a known phone number). In so doing, we demonstrate that anti-robocall indicators could benefit from significantly increased attention from the research community.**

I. Introduction

Robocalls are overwhelming phone users. Long existent but relatively rare, the combination of deregulation, interconnectivity of telephony networks and a lack of end-to-end authentication has recently made it simple and inexpensive to send such calls untraceably at very large scale. This problem continues to grow almost entirely unabated, and it is believed that nearly 50% of all calls in 2019 will be robocalls [1].

While regulatory mechanisms to combat robocalls exist [2], the volume of such calls has only continued to grow rapidly. The technical community has responded by creating a number of anti-robocall applications for mobile devices. These applications generally rely on centrally manicured blacklists, and have collectively been downloaded by over tens of millions of users. However, the efficacy of these applications and, specifically, how well users understand and respond to the indicators they present for incoming calls is currently unknown. As such, it is clear users want to avoid spam calls but unclear if anti-robocall applications are an effective means of alerting users and modifying their behavior.

In this paper, we evaluate the impact of interface design elements on user decision-making for anti-robocall applications. Our goal is not only to determine if anti-robocall applications can reduce the number of malicious/nuisance calls answered by users, *but also whether the answering rate can be improved for legitimate calls given additional guarantees about a caller's identity*. This distinction is critical, as the sheer volume of such calls is forcing many to simply stop answering calls or abandon the platform for voice communications altogether [3].

Our work begins by examining the ten most popular of anti-robocall applications for Android and identifying the prominent visual elements used to alert users of the nature of an incoming call. We then assemble focus groups to discuss how users currently attempt to screen robocalls, to identify which elements in current applications most unambiguously alert them to the presence of such calls, and then ask them for the features they wish to see in such applications. Finally, we develop five applications, a control, two based on the strongest indicators identified in current applications and two more based on the requested indicators, and demonstrate that both applications lower the percentage of robocalls from unknown numbers answered by users. Unfortunately, we then show that adversaries can erase the gains provided by these applications by spoofing meaningful numbers (which are easily gathered from sources such as social media or sold by advertisers), unless Authenticated Caller ID is in place.

This paper makes the following contributions:

- **Survey of Current Applications:** The anti-robocall application space has grown rapidly and with little coordination, leading to a wide array of indicators (ranging from emojis such as a 'thumbs down' and a cartoon octopus to backgrounds spanning the range of the color spectrum). To our knowledge, we are the first to systematically characterize these indicators.
- **Identify Strongest Current Indicators:** We recruit six focus groups and three interviewees to provide feedback on current robocall indicators. These groups noted significant confusion regarding lock symbols, but identified the international prohibition sign and checkmark symbol favorably. Moreover, the focus groups also indicated that they would like to see warnings accompanied by an alerting color covering the entire screen (but not red).
- **Moderate Positive Behavioral Change:** We conducted an interactive study to evaluate the effectiveness of warnings for robocalls. We show a 43% decrease in answered calls when a spam call warning is present. However, we then demonstrate that attackers can largely eliminate these positive behavioral changes by calling from a number known to the callee (e.g.,

Network and Distributed Systems Security (NDSS) Symposium 2020
23-26 February 2020, San Diego, CA, USA
ISBN 1-891562-61-4
https://dx.doi.org/10.14722/ndss.2020.24286
www.ndss-symposium.org

a family member, their bank, etc), unless Authenticated Caller ID is in place. Finally, we show that using Authenticated Caller ID increases the number of users that answer non-malicious calls from unknown numbers by 15%.

This work does not attempt to judge the quality of the blacklists powering anti-robocall applications, nor is it a total ranking of specific apps. Instead, we seek to characterize how users respond to the robocall alerts that accompany them. As such, we believe that our approach attempts to capture a "best-case" approximation of interface effectiveness. Identifying effective indicators has proven extremely challenging historically. The browser community, for instance, spent well over two decades refactoring their indicators to best alert users of danger [4], [5], [6], [7], [8], [9], [10]. Anti-robocall interfaces introduce new challenges. Specifically, decisions to answer phone calls are real-time (unlike visiting a website, where a user could potentially take an arbitrarily long time before proceeding), and because most carriers lack the ability to strongly authenticate Caller ID. Accordingly, determining whether or not an incoming call is a robocall is a related, but new challenge in secure interface design.

The remainder of this paper is organized as follows: Section II provides critical background information about robocalls and Caller ID spoofing; Section III describes our assumptions about the adversaries; Section IV evaluates the current state of the art in robocall detection app interfaces; Section V details the feedback from our focus groups to identify the most critical elements of alerting interfaces; Section VI shows how users reacted to robocalls using three different interfaces; Section VII provides additional discussion and insight; Section VIII details related work; and Section IX provides concluding remarks.

II. Background

The global telephony infrastructure includes cellular networks, Voice over Internet Protocol (VoIP), and the Public Switched Telephone Network (PSTN) (Figure 1). These networks are connected via gateways, which allow calls made in one network to reach endpoints in other networks. Each technology generates its own associated metadata; however, we cannot guarantee that any of this data can be delivered end-to-end except voice and Caller ID, neither of which is authenticated.

Even though some devices authenticate directly to their provider network, the ability to confirm identity beyond one's provider does not exist. Robocallers and telemarketers take advantage of the ability to call anyone while claiming an arbitrary identity. Traditionally, telemarketing companies have depended on a collection of numbers to deliver unsolicited information through the PSTN. However, solely using the PSTN to make multiple calls can be difficult and costly. Telemarketers and robocallers now largely use autodialers and VoIP services to inject calls. VoIP generally provides the cheapest means of making calls and a simpler way to spoof Caller ID. In many of these systems, a user can choose the name and number shown on the Caller ID since that information originates from the client side in that protocol. The ability to change Caller ID information allows robocalls to appear as a familiar or trusted contact.

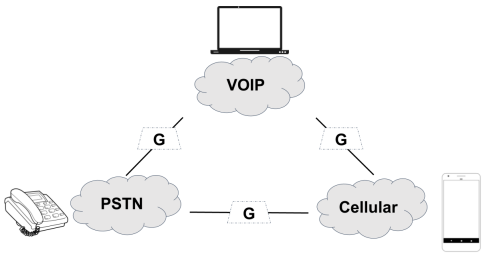


Fig. 1. A high-level overview of the global telecommunications infrastructure. The PSTN, VoIP, and Cellular Network make up the infrastructure. Different devices rely on each network but can communicate with each other through gateways. Robocalls take advantage of the lack of end-to-end authentication and low cost to flood this infrastructure with calls.

Both the increase of robocalls and limited robocall prevention have prompted research to understand the threats [11], [12], [13] and solutions from heuristics to cryptography [14]. Solutions include using Caller ID (assuming no spoofing), black or whitelisting [15], call back verification [16], content and audio analysis [17], [18], chatbots [19], provider-based solutions (e.g., SHAKEN/STIR [20], Authenticated Caller ID [21], [22]), end-to-end solutions (e.g., AuthentiCall [23], [24]), and mobile applications that implement some of these solutions. This work investigates the use of mobile applications and evaluates the warning designs being used to alert users of incoming robocalls.

Since some end-to-end solutions include Authenticated Caller ID, this feature was also included in the designs tested. End-to-End Authenticated Caller ID stems from the work of Reaves et. al, [23], [24] where an application can verify that a caller is who they claim to be by cryptographically authenticating both parties. For this work, Authenticated Caller ID is the presence of Caller ID information that has been verified.

III. Threat Model

We assume an adversary as similar as possible to real-world robocallers.

Robocallers are able to place a large number of low-cost or free phone calls. This adversary does not have special access to a provider core network; rather, they rely on either disposable phone numbers [25], a simbox [26], or alter call meta-data to "spoof" the source of the call (i.e., Caller ID spoofing). Attackers select their targets via multiple strategies including enumerating the address space, web scraping, purchasing contact information from advertising networks, or directly from social media.

With the above-described abilities, an adversary could decide to call from either a random number, a number with a small edit-distance from their target (e.g., Caller: 999-555-1234; Receiver: 999-555-1235), or even from a trusted institution (e.g., the Internal Revenue Service, a financial institution) or someone within the target's social network (e.g., a parent or grandparent).

We assume that call blacklists accurately identify malicious calls. We aim to test whether or not currently deployed user interface elements effectively alert users.

IV. Survey of Anti-robocall Applications

We begin with a study of the state-of-the-art in anti-robocall applications. Because these efforts have been without any central planning or standardization, it is critical that we characterize the wide array of techniques already in place.

A. App Selection

Figure 2 shows screenshots of the applications evaluated in our study. All of the anti-robocall apps selected from the Google Play store 1) appeared as a search result for *spam call blocker* in October 2018, 2) were free to download, 3) had an average rating of at least four stars, 4) had at least one million downloads, and 5) were not designed by a telephone carrier.[1] Based on the order Google Play presented the search results, the first 10 applications were chosen. The privacy policy, website and Google Play page of each app were analyzed to determine how the robocall apps identify spam calls and alert users of spam calls.

B. Results

The apps that met our requirements in October of 2018 are shown in Figure 3, which also includes the icons, star rating, number of installs, and the abbreviations used to help differentiate them in this paper.

Of these ten apps, A2 is the only application that is solely focused on call blocking and does not provide a warning for incoming calls. Accordingly, this application will not be discussed further.

1) Robocall Identification Method: In addition to using lists, many applications rely on Caller ID, publicly available lists, and community comments to identify robocalls and phone spam. A1's privacy policy and website do not fully detail how blocked numbers are handled. However, it does mention that their global blacklist is comprised of data from sixty sources. A3 identifies spam using complaint information provided by FCC, FTC, IRS, State of Indiana and their community of users. The remaining applications do not specify exactly where their database information comes from, but they do state that they build their database based on the spam calls detected by their community of users. Finally, A4 and A9 state in their privacy policy that users' contacts could potentially be added to the organization's database. We mention robocall identification methods in this design paper for completion. Further investigation of app accuracy and reliability is left for future work.

2) Warning Design: The warning designs used for each application were analyzed using Wogalter's warning design guidelines [27]. We focus on wording, layout, placement, and pictorial symbols. Salience and personal factors are a part of the guidelines but were not considered because they require feedback from users.

[1]This information is based on the result from 2018. Since then, *Should I Answer?* has been updated and uses a new icon and interface. Also, Hiya and Mr.Number have a similar warning design because they were created by the same company. However, both applications were popular thus landing them on our list.

Wording. Wogalter states that a warning should consist of four components - signal word, identification of hazard, explanation of consequences, and directions for avoiding the hazard. Each app's warning met those requirements by including words that "attract attention" and identify the hazard. Words such as *robocall, spam* or *fraud* help bring the user's attention to the type of call they are receiving. Users download the apps because they know the hazards of answering spam calls and are using the app to avoid them. Therefore, installing and using the app addresses the hazard consequence and avoidance criteria. Also, community-based apps A4, A5, and A9 often present the hazard's consequence within the wording of the warning. For example, an app alert might be, "Previously reported financial scam," which would let the user know that if they answer and comply with the call, they could become a victim of a financial scam. When the spam alert provides these details, additional information may not be necessary to communicate the possible consequence.

Layout & Placement. Each app's warning design is included in Figure 2. The most popular design includes a rectangle in the middle of the screen that overlays the incoming call screen. The majority of the warnings include white text over a red background and a symbol to indicate an alert is being expressed.

Pictorial Symbols. Pictorial Symbols are used in each app to convey the warning message. A6, A7, and A10 use warning messages that include symbols that are often displayed to demonstrate or bring attention to an issue. A8 uses its company logo, a green octopus with a red background, which is shown in Figure 2h, as the spam call warning symbol. When using A8, the user will need to read the fourth line of the warning to see that the call is categorized as a spam call if they misunderstand the meaning behind the octopus.

C. Discussion

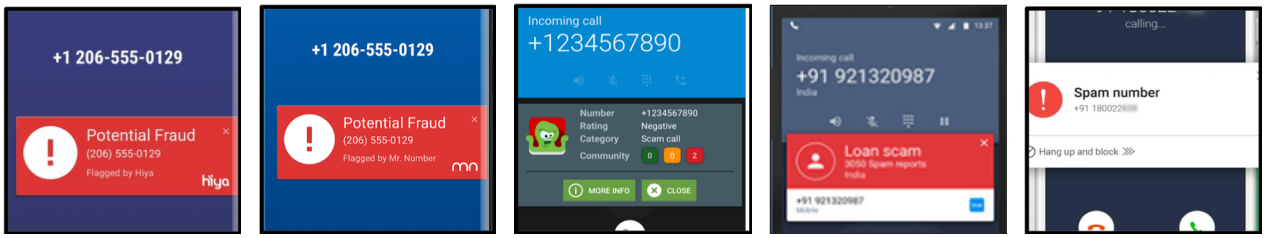
The results show that the majority of the apps use the color red in a rectangular warning screen overlay and place their warning in the middle of the screen. A8 displays a lot of information closely together, and providing too much or cluttered information can effect warning detection [28]. The user would need to read the four lines to find the call's category, possibly making it difficult for the user to easily and quickly interpret the warning. The inclusion of the octopus logo as a symbol in A8 is unique and might become helpful over time as the user has more experience with the app, but an octopus is not among the symbols often used and most recognized by users as an indicator of an alert or warning [29], [30]. All of the apps except A8 meet the warning design guidelines by using clear wording, symbols, and placing warnings where the user can see them.

However, meeting Wogalter's criteria is only the first step in creating an effective warning. As mentioned in previous work [31], [32], [7], warnings should be tested and adjusted for the specific danger it is being created to alert.

V. User Experience Collection

After identifying the spam call warning design elements used in anti-robocall applications, we conducted focus groups to understand user experiences with robocalls and identify

(a) CallApp: Caller ID (A1) (b) Call Blocker (A2) (c) Call Control (A3) (d) Caller ID and Call Blocker - DU Caller (A4) (e) Clever Dialer (A5)



(f) Hiya (A6) (g) Mr.Number (A7) (h) Should I Answer? (A8) (i) Truecaller (A9) (j) Who's Calling (A10)

Fig. 2. Each application, as shown above, displayed warning design examples on their respective Google Play store pages. Most (80%) of the apps use the color red to indicate an incoming spam call. A6 and A7 use similar warning designs because they were created by the same company. A2 only blocks calls and does not show a spam warning, so its blocking options are shown in this Figure instead.

Name:	Call App (A1)	Call Blocker (A2)	Call Control (A3)	Caller ID & Call Blocker (A4)	Clever Dialer (A5)
Stars:	4.6	4.6	4.4	4.6	4.6
Installs:	100M+	10M+	5M+	5M+	1M+

Name:	hiya (A6)	Mr.Number (A7)	Should I Answer? (A8)	Truecaller (A9)	Who's Calling (A10)
Stars:	4.5	4.2	4.7	4.5	4.4
Installs:	100M+	10M+	1M+	100M+	10M+

Fig. 3. The number of installations, rating and icon for each application that was reviewed. The majority of the applications incorporate the phone symbol in their icon and the word *call* in their name.

(a) Spam Call (b) Authenticated Call

Fig. 4. At the end of each focus group, the participants were shown five call notification designs that were randomly selected from a group of 54. These two designs showcase the red background that was disliked and the blue background that was liked by participants. The other notification designs can be seen in Appendix E.

the warning design elements users preferred. The following research questions will be answered in this section:

- RQ1: How do users determine if they will accept a phone call?
- RQ2: How do users detect and stop spam calls?
- RQ3: What are users' notification preferences? How do they receive and react to various visual cues?
- RQ4: What characteristics and features would compel users to download an anti-robocall app?

A. Methodology

Focus groups and interviews were conducted using semi-structured questioning in a 60-minute conversation.[2] Each participant was asked to recall their experiences with spam calls, advise on how to handle spam calls and discuss the features and abilities of an app they would use to handle spam calls. In addition, they were asked to provide their opinion on five of the 54 available warnings and notice design probes, examples of which are shown in Figure 4.[3] For the context of this study, we used the phrase "spam call" interchangeably with "robocall" to refer to all types of unwanted calls. Subjects were recruited using an online research administration system at a University. Our protocol was approved by the local Institutional Review Board. Some participants were volunteers and others received extra credit for their participation. Our study was one of many extra credit opportunities offered to those who participated for extra credit.

After each focus group and interview, the resulting audio recording was transcribed. Then, open coding was performed

[2] Although we only wanted to hold focus groups, scheduling conflicts led to the interviews. We decided to hold interviews to make sure that everyone had an opportunity to participate regardless of their personal circumstance.

[3] All of which are in Figure 8 in Appendix E.

by the research team. A codebook was created after reading through the transcriptions and each response was coded by the researchers. All coding disagreements were discussed until an agreement was reached. Axial coding was done to align responses with categories or themes which are present below.

B. Participants

A total of 18 people participated in either a focus group or an interview, similar to the number of participants in analogous studies on diabetes [33], Tor browser [34], and password storage [35]. A total of six focus groups (15 people) and three interviews (3 people) were conducted. Participants were between the ages of 21 and 32 ($\bar{x} = 23.056, \sigma_x = 2.859$). Most (13 participants) of the participants were male, and the remaining were female. The participants identified themselves as being a part of five ethnic groups: African American (2 participants), Caucasian (4 participants), Latino/Hispanic (5 participants), South Asian (3 participants), and East Asian (4 participants). Eight participants currently use their phone provider's app service to manage spam calls or third-party apps, and the remaining participants use other mechanisms (e.g., blocking or ignoring unknown calls).

C. Results

As previously mentioned in Section 4.1, the focus group participants discussed how they handled spam calls and warning design needs. The resulting responses have been categorized and are used to answer our research questions in the following subsections.

1) ***RQ1: How do users determine if they will accept a phone call?****:* All of the participants mentioned that they look at the Caller ID, area code, and the time to determine if they will accept a call. They use Caller ID to determine if the call is from a known or unknown number. They use the area code to determine if the call is from a location they would expect to receive a call from. Participant T06 said, "If it's a number from my hometown, I never take it because my family doesn't live there anymore." They looked at the time to see if they had time available to talk or to see if it matched the time in which they were expecting a call. If busy, they were more likely to ignore a call from an unknown number. Also, they refer to their personal call expectations to determine if they are expecting a call from an unknown number during that time in their life. As many participants mentioned, they are more likely to answer a call from an unknown number if actively on the job market.

2) ***RQ2: How do users detect and stop spam calls?****:* Participants were asked to advise on how to detect and stop incoming spam calls, and the majority of them reference using Caller ID in some capacity. The full list of responses can be found in Appendix A. The top five responses to this question (in order of frequency) are as follows:

1) Look at the area code and determine if the call is coming from an area in which you would expect to receive a call at that time.
2) Block the number of a known spam caller.
3) Do not answer calls from an unrecognizable number, but you should check to see if the caller left a voicemail.
4) Do not answer calls from numbers you do not recognize in general.
5) If the first six to nine digits of the number match your own, it is likely a spam call and you should not answer.

3) ***RQ3: What are users' notification preferences? How do they receive and react to various visual cues?****:* Answers to this question were gathered by showing a random set of five designs from the 54 total call notification designs in Appendix E. These were inspired by designs seen in anti-robocall apps and the idea of Authenticated Caller ID. Each group saw a different set of notifications, but their responses were similar and resulted in three takeaways.

1) **Background Color:** A background color is more noticeable when it differs from the original incoming call screen and fills the entire screen. Although participants noticed the color when it only filled a portion of the screen, they stated that when a color fills the entire screen the warning was more noticeable. However, a full red screen was noticeable but undesirable. In reference to Figure 4a, participant T08 said, "looks like I have a virus on my phone. If I saw this I would uninstall the app immediately." The other participants in T08's focus group and other groups agreed that the color was too alarming. When participants described their ideal app, they mentioned seeing the color green or blue for authenticated calls and red, black, yellow, gray, or orange for spam calls. Although the red background color was undesirable, participants still suggested the use of that color for the background or icons used within their ideal app.
2) **Icons:** The check and "X" mark conveyed the same message to all participants that saw those symbols. The use of the locks confused participants. Participant T05 stated "the open and closed lock both mean secure to me". Because locks convey a mixed messaged in this context, they should not be used for this purpose. This is not surprising since browser warning researchers have also come to this conclusion [31], [32], [7]. The designs that used emojis (emotionally expressive faces) were rejected immediately by the participants that saw them. Although they were able to come to similar conclusions on the meaning of the emojis, they did not believe those symbols were serious enough for indicating call type. Participant T17 stated, "Seeing a sad face makes me think it's saying the [call] signal itself is bad." In addition, when a few participants saw a sad face, thumbs down, or the international prohibition sign along with the word "Mom", they interpreted it to mean that their relationship with their mom was not good so they should not answer that call.
3) **Caller ID:** Participants trust Caller ID and want authenticated Caller ID. When asked if they would answer a call that the app declared as spam but the Caller ID declared "Mom" is calling (Figure 4a), many participants chose to answer the call. Some said they would be willing to take the chance but would more than likely not answer if the app's warning was always correct in the past. Seventeen of the

eighteen participants saw the notification warning them that the call from "Mom" was possibly spam. Of those, nine participants said they would take that call. Participant T06 even said, "I mean it's my mom. I'll always take her call."

*4) **RQ4: What characteristics and features would compel users to download an anti-spam call app?**:* During the focus groups, participants were asked to express how a spam management app would look and behave if they used one through writing, illustration or both. The participant responses were sorted into six categories: call blocking, call log, number database, app design, call options, and pre- and post-spam needs.

Call Blocking: Fifteen participants wanted the ability to select which numbers to block automatically, either during app installation or after receiving certain calls. Participants addressed the need to be able to have some calls blocked automatically, such as calls from out of state numbers, and then block some calls individually, e.g., those from someone they would no longer wish to receive calls from.

Incoming Calls: Ten participants discussed the handling of incoming calls. During one focus group, the participants all agreed that one way to handle spam calls was to simply let the phone ring. However, some participants complained about being unable to use their phones while waiting. All participants in one focus group agreed that the app needed to be able to send calls to voicemail without making the phone inoperable and without letting the caller know the call was forwarded. Participant T16 stated that their ideal mobile app would "let the phone ring in back so I can still use the phone. But let the caller think it's still ringing." Participants also discussed wanting an app that would warn them that a call was spam while the phone was ringing, the importance of including Caller ID, and the ability to support international calls.

Pre/Post Spam Needs: Eight participants discussed features that could assist users before or after receiving a spam call. They expressed interest in receiving tips and hints about spam calls and knowing when to give personal information or their phone number. If they miss a spam call, and the spammer leaves a message, they would like the app to delete that message. If they do receive a spam call, they would like for the app to put them on the blacklist for that specific number. So instead of having to answer the call and select the appropriate button to be added to the blacklist for that particular spam call, they wanted an app to do it for them. Unfortunately, users could be putting themselves at risk by adding their number to a list owned by an unknown entity. If they are continuously bothered by a specific caller, they would like the ability to report the spammer to the proper authorities using the app.

Call Logs: Seven participants expressed interest in being able to filter and monitor spam calls through the call log. They want to be able to filter this list and see a list of spam calls they have received and how many times a particular number has called them. Participant T15's call log would be able to tell them "where the call was from, like by area code and maybe even... if possible, tell me what kind of spam it was. Like was it insurance, banking or whatever." [4]

Number Database: Five participants discussed having an app that would be able to detail information about a particular phone number by using crowd-sourcing. The app would use feedback about phone numbers from various app owners and use that to relay information to other users. Participant T12 said, "It [the app] should say 'so many people reported this to be a spam number' or something like that." Participants also expressed interest in being able to search for an unknown phone number within the app and retrieve information about it.

App Design: Three participants commented on app design. Participant T06 suggested the app design be as simple as possible. The other two participants discussed the app's relationship with the native app. Participant T10 believed that the app should work with the native phone calling app, while Participant T03 believed the app should replace the native phone calling app.

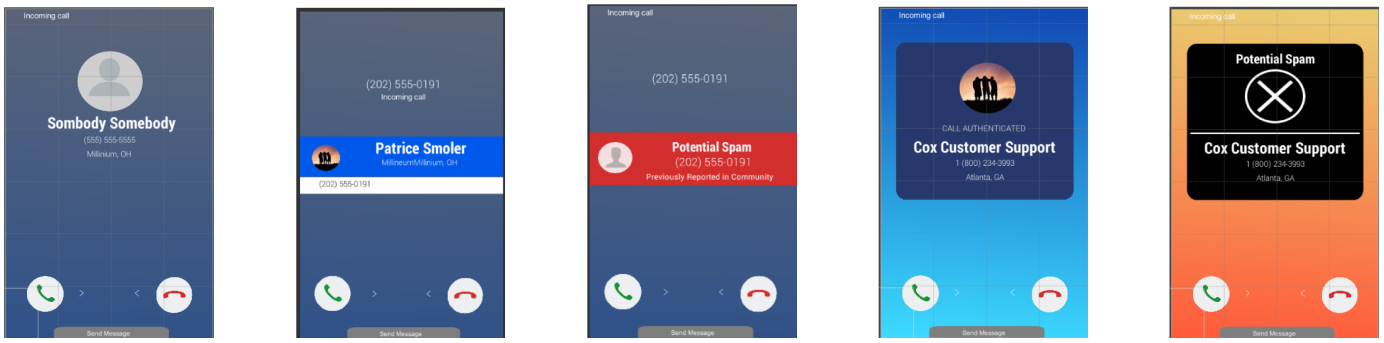
D. Discussion

The focus group results show that cell phone users have adopted common and unique practices to manage spam calls. Only eight participants were using or had used a third-party app or an app provided by their telecommunications provider to manage spam calls. Based on the responses, some participants did not know third-party apps were an option, or they believed that their method worked well enough without the additional application. However, everyone mentioned using Caller ID to make call response choices, whether they used an app to manage their calls or not. This is problematic because Caller ID information is never authenticated. Currently, spam is handled using blacklists which rely on Caller ID. Caller ID can be spoofed, which makes blacklists and other Caller ID based methods unreliable. Although most participants discussed receiving calls from familiar numbers that were actually spam callers, they still relied on Caller ID.

The results also suggest that people will react and adapt differently to warnings. Therefore, designers should choose characteristics, like background color, that will effectively communicate the message and should constantly update the design as phones update. Users should not need to adapt or be trained on what a warning means. They should be able to see it and use context clues to understand it.

Finally, participants disliked the use of padlocks, emojis, and a completely red background but liked the use of a completely blue background, checkmark, "X" mark, and Authenticated Caller ID in incoming call notices. However, besides Authenticated Caller ID, these or similar attributes

[4]The focus group participants suggested various app features that are not currently or widely available. Although these features could be helpful, it is out of scope for this paper to investigate whether or not these app features would be used if provided. We leave this task for future research. Our focus is on the suggested design elements and the presentation of features for incoming call announcements that are currently being implemented and further developed like Authenticated CallerID.



(a) Control: Imitates incoming call on Samsung Galaxy S7

(b) Avail-CID: Inspired by anti-robocall apps

(c) Avail-Spam: Inspired by anti-robocall apps

(d) Focus-AID: Inspired by focus group results

(e) Focus-Spam: Inspired by focus group results

Fig. 5. Participants were shown five designs during the interactive survey. The Control design was created to imitate the incoming call screen on a Samsung Galaxy S7 device. The *Available* (Avail-CID and Avail-Spam) and *Focus* (Focus-Auth and Focus-Spam) Category designs were inspired by results of the app analysis and focus groups, respectively.

can be seen in existing robocall apps on the market like apps A5, A7, and A10 reviewed in Section IV. Although we did not investigate built-in robocall apps, we believe that these warning design elements could work for both third-party and native apps.

VI. Warning Design User Study

The app analysis results show that several anti-robocall applications follow Wogalter's warning design guidelines and use a similar warning layout. The focus group results suggest that users desired a warning with easy-to-interpret icons and a noticeable background color that fills the entire screen. The results also show that users relied heavily on Caller ID. As a follow-up, we test the identified warning design elements that the users desired and current applications use against reliance on Caller ID. This section discusses those results by answering the following five research questions:

- RQ5: Do robocall warnings impact users' reaction time to *incoming calls*?
- RQ6: Do robocall warnings affect users' responses to incoming calls from *unknown numbers*?
- RQ7: Do robocall warnings affect users' responses to incoming calls from *known numbers*?
- RQ8: Will the "Available" (Avail-CID, Avail-Spam) and "Focus" (Focus-AID, Focus-Spam) designs have a significantly different effect on user response or reaction time?
- RQ9: How will participants rank the various designs shown?

A. Methodology

We developed a mobile application to 1) display mock phone calls (screenshots) and 2) capture the participants' responses to the mock calls. Five warning designs were used in the survey, as shown in Figure 5. The *Control* Design (**Control**) is an imitation of what users of Android-based Samsung Galaxy S7 devices would see if they received a call without a robocall detection application.

Available Category: The *Available* category (Avail) was inspired by the app analysis results in Section IV, which is based on the top ten currently available apps in the Google Play Store. The majority of the selected apps used a red or blue bar in the middle of the screen for their warning. **Avail-CID**, the non-spam notification, incorporated the blue bar and Caller ID information. The *Authenticated Call* label was not added because this was not incorporated in the apps we reviewed. **Avail-Spam**, the spam warning, incorporated the red bar and some of the Caller ID information. The name was removed from the Caller ID because most of the screenshots from the apps reviewed also removed the name, as shown in Figure 2.

Focus Category: The *Focus* category (Focus) was inspired by the focus group discussions that covered both participants' ideal app and feedback on suggested designs. Participants mentioned the color red when describing their ideal app. But when later shown warning suggestions, most noted that the color red was too alarming, especially when it was the background color, for the spam warning. These results lead us to minimize the color red in the background and use the other colors (yellow and orange), color scheme suggestions of the participants in **Focus-Spam**. We chose the yellow-to-red gradient because participants mentioned that the colors yellow and orange would also be alarming. We chose the black box with white text because the design needed a clear contrast between elements on the screen. **Focus-AID**, the authenticated warning, includes a blue gradient background and a blue notification box at the top of the screen. This design choice was made based on focus group results which showed that users like the blue background color.

TABLE I. A list of the independent variables, and their levels, for this experiment.

Variables:	Levels
Response	Accept, Did Not Accept
Round	R1, R2 ,R3
Warning Design	Focus-AID, Focus-Spam, Control, Avail-CID, Avail-Spam
Number	N1, N2, N3, N4, N5, N6

Pilot Study: Before we began the user study, we held a pilot study with five participants to test the Focus designs to make sure they were acceptable. In particular, we wanted

TABLE II. DESCRIPTION AND BREAKDOWN OF EACH WARNING DESIGN

Category	Warning Design	Description
Control	Control	Imitates the typical incoming call screen on the Samsung Galaxy S7
Available	Avail-CID	Mimics the non-spam warning design of the top ten anti-robocall apps which includes Caller ID
	Avail-Spam	Mimics the spam warning design of the top ten anti-robocall apps
Focus	Focus-AID	Includes the non-spam warning design elements preferred by the focus group participants which includes Authenticated Caller ID
	Focus-Spam	Includes the spam warning design elements preferred by the focus group participants

to make sure that the **Focus-Spam** design was alerting even without the color red. Each pilot study participant was asked specifically about the designs. Two participants requested a smaller spam warning icon and an increase in text size. One participant requested to change the background color of the authenticated call notice to green. The color scheme for the Focus-Spam designs were not changed for several reasons. First, no one in the pilot study had an issue with the current spam color scheme. Second, group participants stated the colors yellow and orange would also be alerting, and finally, warning research suggests that these colors can be used to express different hazard levels [36]. We chose not to change the background of Focus-AID because the majority of the focus group participants (16) approved the blue color scheme in the examples presented. Additionally, research shows that blue motivates people to "behave in a more explorative, risky manner [37], [38], [39]", which would be beneficial in this context.

Setup: Each design was shown with an incoming call from six unique numbers. The numbers chosen were based on the types of spam calls experienced by the focus group participants:

- N1, N2: Two known numbers entered by the participant (N1, N2)
- N3: An unknown number where the contact name is a city and state (N3)
- N4: "Harold Rogers" whose number includes the same first 9 digits as the participant's number (N4)
- N5: "Veranda Gardens" which appears to be located in the same area as the participant (N5)
- N6: "Ashford Loans", a loan organization with an area code different from the participant (N6)

Before completing any tasks, participants were told they were testing out potential app alert designs for an upcoming robocall application. They were asked to respond to each incoming call as they would in real life. Participants then provided two known numbers (N1, N2). Each participant entered the contact information of two individuals whom they regularly communicate with (just as it is in their personal devices) on their assigned mobile device. They were then shown various mock calls that displayed until the fifth ring of a monophonic or polyphonic ringtone ($\sim$23 sec). If the participant did not respond to the mock call within the time allotted, the next call would appear. The participants saw every possible combination of numbers and designs six times across three rounds in random order. During each round, three practice mock calls were initiated first followed by the 30 experimental mock calls that were randomly displayed twice. After each round of 63 mock calls, the users were then allotted a 5-minute break. The participants were shown each design multiple times to get their true response to the call since their initial response may not be their true response. After the study, each participant was debriefed on the true purpose of the study, asked a few questions about their experience, and given a follow-up survey. A total of 34 participants responded to 30 mock calls shown six times in random order. This lead to the collection of 6,120 data points. All independent variables are listed in Table I and each warning design is described in Table II. This experiment is set up to detect cause and effect, thus having high internal validity. We wanted participants to focus on the warning design and therefore provided a best-case scenario that has limited external influences [40].

B. Participants

A total of 34 participants were recruited through a participatory system at a University for this study, a participant total and composition similar to analogous studies [41], [42], [43]. Some participants were volunteers and others received extra credit for their participation. Our study was one of many extra credit opportunities offered to those who participated for extra credit. The participants spent 30 minutes participating in the study on average and were between the age of 20 and 32 ($\bar{x} = 24.5, \sigma_x = 3.369$), where half of the participants were female. The racial and ethnic backgrounds of the participants include East Asian (15%), Caucasian (26%), African American (18%), South Asian (26%), Latinx/Hispanic (6%), Middle Eastern (6%), and Caribbean (3%). There was no overlap in participants between the focus groups and this study. Participants had to be 18 years of age or older and had to have experience with spam calls to participate. We did this to capture experiences from those who have and have not used robocall applications.

C. Analysis

For each mock call shown, we recorded the time-lapse as participants determined if they would or would not answer (*Reaction Time*) and the final decision for each mock call (*Response*). *Reaction Time* was measured from the time the mock call was shown until the participant pressed the button to accept or decline a call. *Response* is measured as the action participants chose to take when the mock call is received. This is measured on a dichotomous scale where participants either accept or reject a call. First, we reviewed participants' responses to make sure no one responded in a pattern to all calls, especially N1, N2, and N6. We found nothing out of the ordinary. Then, Shapiro-Wilkes and Anderson-Darling tests were run on the results using the statistical computation system R [44]. The first 5,000 data points were tested for normality. The resulting p-values were less than .001, indicating non-parametric data, which was also confirmed with a histogram. The Aligned Rank Transform (ART) [45] was used to transform the data and was followed by a Repeated Measures Analysis of Variance test (RM ANOVA).

The RM ANOVA is used to calculate the significant difference for *Reaction Time* within the independent variables *Warning Design*, *Number*, and *Round*. All of the main effects, except *Number*, and interactive effects on *Reaction Time* were statistically significant ($\alpha = .05, p<.05$ in all cases). The

TABLE III. REPEATED MEASURES ANOVA RESULTS

Independent Variable		**Response**		**Reaction Time**	
	Df	F-value	*p*	F-value	*p*
Warning Design	4,132	62.085	$< .001$	5.013	$< .001$
Number	5,165	51.49	$< .001$	1.055	.192
Warning Design: Number	20,660	22.361	$< .001$	7.962	$< .001$
Round	2,66	–	–	177.262	$< .001$
Warning Design: Round	8,264	–	–	5.202	$< .001$
Number: Round	10,330	–	–	1.8232	.017
Warning Design:Number: Round	40,1320	–	–	2.887	$< .001$

TABLE IV. PERCENT OF ACCEPTED CALLS FOR EACH WARNING DESIGN AND PAIRWISE COMPARISONS RESULTS FOR KNOWN AND UNKNOWN NUMBERS (RESPONSE)

	%	All Numbers	Known #s (N1, N2)	Unknown #s (N3, N4,N5,N6)
	Control	56.4%	100%	35%
	Focus-AID	61%	100%	42%
	Focus-Spam	25%	65%	5%
	Avail-CID	55%	95%	34%
	Avail-Spam	13%	34%	3%
	p-value			
Focus-AID vs	Control	ns	ns	ns
	Focus-Spam	$< .001$	$< .001$	$< .001$
	Avail-CID	ns	ns	ns
	Avail-Spam	$< .001$	$< .001$	$< .001$
Focus-Spam vs.	Control	$< .001$	$< .001$	$< .001$
	Avail-CID	$< .001$	$< .001$	$< .001$
	Avail-Spam	.03	$< .001$	ns
Avail-CID vs.	Control	ns	ns	ns
	Avail-Spam	$< .001$	$< .001$	$< .001$
Avail-Spam vs.	Control	$< .001$	$< .001$	$< .001$

TABLE V. PERCENT OF CALLS NUDGED IN INTENDED DIRECTION FOR EACH WARNING DESIGN AND PAIRWISE COMPARISONS RESULTS FOR KNOWN AND UNKNOWN NUMBERS FOR NUDGE RESPONSE

	%	All Numbers	Known #s (N1, N2)	Unknown #s (N3, N4,N5,N6)
	Control	77%	100%	65%
	Focus-AID	61%	99%	42%
	Focus-Spam	75%	35%	95%
	Avail-CID	55%	95%	34%
	Avail-Spam	87%	66%	97%
	p-value			
Focus-AID vs	Control	$< .001$	ns	$< .001$
	Focus-Spam	.002	$< .001$	$< .001$
	Avail-CID	ns	ns	ns
	Avail-Spam	$< .001$	$< .001$	$< .001$
Focus-Spam vs.	Control	ns	$< .001$	$< .001$
	Avail-CID	$< .001$	$< .001$	$< .001$
	Avail-Spam	.04	$< .001$	ns
Avail-CID vs.	Control	$< .001$	ns	$< .001$
	Avail-Spam	$< .001$	$< .001$	$< .001$
Avail-Spam vs.	Control	ns	$< .001$	$< .001$

Wilcoxon test with Holm's sequential Bonferroni correction was used to determine where significance occurred for *Reaction Time* during post-hoc analysis. These tests showed that on average, participants responded to mock calls faster in Round Three than in Round Two and One ($p< .05$ in all cases).

The RM ANOVA is also used to calculate the significant difference for *Response* within the independent variables *Warning Design* and *Number*. This was done twice. In the first case, *Response* is the percentage of accepted calls for every possible combination of *Warning Design* and *Number* over all *Rounds*. In the second case, we calculate *Nudge Response* which is the percentage of calls that were responded to in the way in which participants were nudged for every possible combination of *Warning Design* and *Number* over all *Rounds*. Participants were nudged to answer calls from N1 and N2 with the Control design and all calls shown with Avail-CID and Focus-AID. Participants were nudged to decline all other calls. We report the first case in Figure III since we viewed no changes in p-values when evaluating Nudge Response. The RM ANOVA results showed that all main effects, and interaction effects on *Response* and *Nudge Response* were statistically significant ($\alpha = .05, p<.05$ in all cases). Both tests were followed by pairwise comparison.

The posthoc analysis using Bonferonni correction shows a significant difference between all spam designs (Focus-Spam and Avail-Spam) and non-spam designs (Control, Focus-AID, and Avail-CID), thus showing that participants declined more calls when they saw a spam alert in comparison to when there was no spam alert present ($p < .05$ in all cases). Specifically, the percent of accepted calls decreased by 43% when accompanied by the Avail-Spam warning and 31% for Focus-Spam. The analysis also indicated a significant difference in *Response* between all known numbers (N1 and N2) and unknown numbers (N3, N4, N5, and N6), thus showing that participants were more likely to accept calls from known numbers than unknown numbers ($\alpha = .05, p<.05$ for all comparisons).

D. Results

1) ***RQ5: Do robocall warnings impact users' reaction time to incoming calls?****:* Focus Group participants expressed that the amount of time available to answer a call determined if a call gets answered. Warnings should not increase the amount of time users spend on answering calls and the results suggest that they would, as shown in Figure 6. The main effect of *Round*, *Warning Design* and all interaction effects were statistically significant ($p < .05$). This was because participants responded faster to calls during Round Three and slower to calls that had Avail-CID warnings. On average, participants responded faster to mock calls in round three ($\mu = 1.483$ sec, $\sigma = 1.306$ sec) than in round two ($\mu = 1.774$ sec, $\sigma = 1.430$ sec) and round one ($\mu = 2.477$ sec, $\sigma = 1.975$ sec), which means participants began to respond faster to the mock calls as they progressed in the study ($p < .05$ for all comparisons). In addition, on average, participants responded

TABLE VI. PERCENT OF ACCEPTED CALLS FOR EACH WARNING DESIGN AND PAIRWISE COMPARISONS RESULTS FOR RESPONSE

	%	N1	N2	N3	N4	N5	N6
	Control	100%	99%	29%	42%	44%	25%
	Focus-AID	100%	99%	38%	50%	54%	27%
	Focus-Spam	65%	66%	3%	11%	5%	2%
	Avail-CID	94%	97%	29%	42%	43%	24%
	Avail-Spam	35%	34%	2%	2%	2%	3%
	p-value						
Focus-AID vs.	Control	ns	ns	ns	ns	ns	ns
	Focus-Spam	.002	.03	.002	$< .001$	$< .001$	ns
	Avail-CID	ns	ns	ns	ns	ns	ns
	Avail-Spam	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$	ns
Focus-Spam vs.	Control	.001	.018	ns	ns	.001	ns
	Avail-CID	ns	ns	.ns	ns	$< .001$	ns
	Avail-Spam	ns	.012	ns	ns	ns	ns
Avail-CID vs.	Control	ns	ns	ns	ns	ns	ns
	Avail-Spam	$< .001$	$< .001$	ns	$< .001$	$< .001$	ns
Avail-Spam vs.	Control	$< .001$	$< .001$	ns	$< .001$	$< .001$	ns

TABLE VII. PERCENT OF CALLS NUDGED IN INTENDED DIRECTION FOR EACH WARNING DESIGN AND PAIRWISE COMPARISONS RESULTS FOR NUDGE RESPONSE

	%	N1	N2	N3	N4	N5	N6
	Control	100%	99%	71%	58%	56%	75%
	Focus-AID	100%	99%	38%	50%	54%	27%
	Focus-Spam	35%	34%	97%	89%	95%	98%
	Avail-CID	94%	97%	29%	42%	43%	24%
	Avail-Spam	65%	66%	98%	98%	98%	97%
	p-value						
Focus-AID vs.	Control	ns	ns	.01	ns	ns	$< .001$
	Focus-Spam	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$
	Avail-CID	ns	ns	ns	ns	ns	ns
	Avail-Spam	.002	ns	$< .001$	$< .001$	$< .001$	$< .001$
Focus-Spam vs.	Control	$< .001$	$< .001$	ns	ns	.002	ns
	Avail-CID	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$
	Avail-Spam	ns	.03	ns	ns	ns	ns
Avail-CID vs.	Control	ns	ns	$< .001$	ns	ns	$< .001$
	Avail-Spam	ns	ns	$< .001$	$< .001$	$< .001$	$< .001$
Avail-Spam vs.	Control	$< .001$	.031	ns	$< .001$	$< .001$	ns

slower to Avail-CID ($\mu = 2.024$ sec, $\sigma = 1.75$ sec), than any of the other *Warning Designs* (Avail-Spam ($\mu = 1.811$ sec, $\sigma = 1.545$ sec), Control ($\mu = 1.845$ sec, $\sigma = 1.461$ sec), Focus-AID ($\mu = 1.9355$ sec, $\sigma = 1.656$ sec), Focus-Spam ($\mu = 1.941$ sec, $\sigma = 1.807$ sec)). *Reaction Time* of Avail-CID was statistically different than Avail-Spam, Control, and Focus-Spam ($p< .05$ in all cases). However, the difference between these times is milliseconds, which would likely not amount to a noticeable difference for users. In addition, none of the participants mentioned spending additional time to answer calls under specific *Warning Designs* during the debrief.

*2) **RQ6: Do robocall warnings affect users' response to incoming calls from unknown numbers?**:* The consequences of answering robocalls can have a negative effect on how telephone users respond to legitimate unexpected calls [46]. Warnings should help users distinguish between a robocall and a call from a legitimate entity. The results show that the presence of an authenticated call notice did increase the percentage of accepted calls. Participants accepted 42% of calls from unknown numbers under the Focus-AID, 35% under the Control design and 34% under the Avail-CID design. The number of participants that answered calls from unknown numbers increased by 10% under Focus-AID, when compared to the Control design. When nudged, participants were more likely to respond as intended to unknown numbers under Focus-AID and Avail-CID than Control ($p< .05$) as shown in Table V. When receiving an incoming call from an unknown number, robocall warnings significantly decreased the number of calls answered when compared to the Control design. Participants answered 35% of calls from unknown numbers under the Control design, which is significantly more than Focus-Spam design (5%), and Avail-Spam design (3%) ($p< .001$ for all comparisons with Control), as shown in Table IV.

*3) **RQ7: Do robocall warnings affect users' response to incoming calls from known numbers?**:* Research shows that email spam from known places or people is more likely to successfully trick users. This technique is also used for spam calls, which is why people fall victim to spam calls from the IRS or banking institutions. The results of this study show that spam warnings could potentially help to solve this problem.

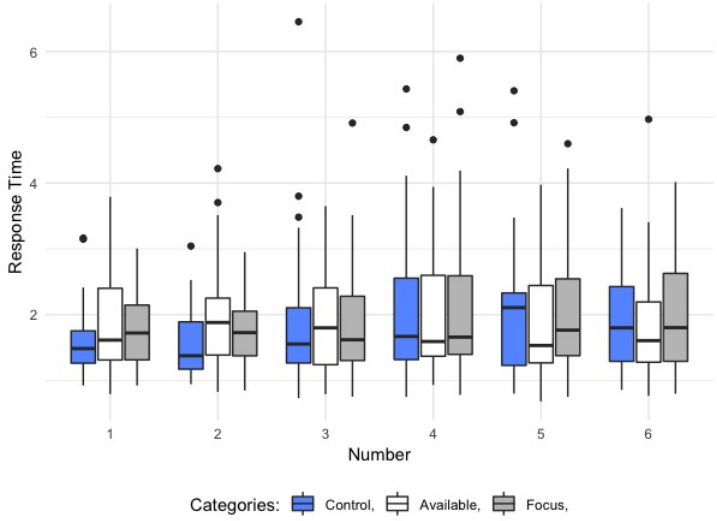


Fig. 6. This box plot shows the *Reaction Time* for each *Number* under the three *Warning Design* categories. It illustrates the mean *Reaction Time* for each and, in most cases, that most participants spent more time reacting to calls with a warning present.

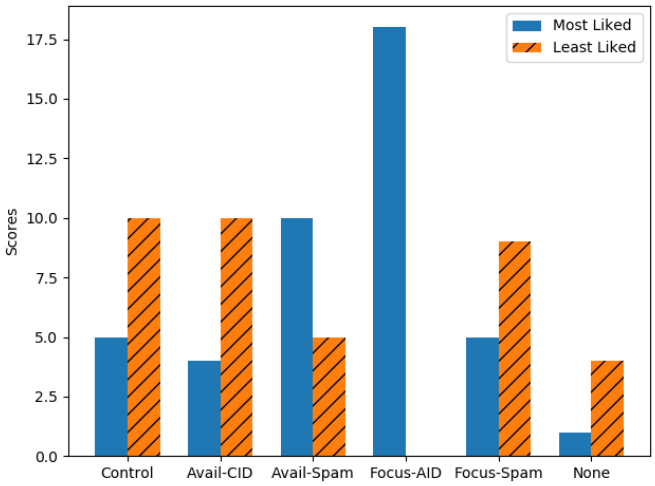


Fig. 7. At the end of each user study, the participants were asked which design they liked the most. Some participants chose more than one design, and some had no preference. This bar graph shows that Focus-AID was liked the most and Avail-Auth and Control were like the least.

Even though participants provided the known numbers shown in the study, they answered significantly fewer calls from known numbers when presented with a spam warning call design compared to calls from those same numbers without a warning. Participants were more likely to answer calls from known numbers when no warning was present (100%) compared to when Avail-Spam (34%) or Focus-Spam (65%) was shown ($p< .001$ for all comparisons), which is shown in Table VI and VII. The presence of Avail-Spam also decreased the average number of participants that answered spoofed calls by 13% when compared to the number of participants that answered those calls under the Control design.

4) ***RQ8: Will the Available and Focus design have significantly different effects on user response?****:* As previously discussed, the Focus and Available design are inspired by the design elements desired by our participants and used by anti-robocall apps, respectively. The results suggest that the Focus and Available design did have a significantly different effect on participant *Response* and *Reaction Time*. Participants also accepted significantly more calls from unknown numbers with Focus-AID than Control ($p < .05$) and Avail-CID than Control ($p < .05$). There was no significant difference between the number of unknown calls accepted when comparing Focus-AID and Avail-CID. However, participants were more likely to answer spam calls from known numbers when Focus-Spam (65%) was shown compared to Avail-Spam (34%) ($p<$ .05). Avail-Spam was more effective at getting participants to decline calls, whereas Focus-AID was more effective at getting participants to answer calls from unknown numbers.

5) ***RQ9: How will participants compare the various designs shown?****:* The Focus-Spam and Avail-Spam differ in warning placement, screen color, warning label color, and icon used. Although we cannot conclusively point to the effectiveness of each element, the qualitative results show that screen and warning label color played a crucial role in decision making and likability. At the end of each study, the participants were asked to indicate which design they preferred the most and which they liked the least. As shown in Figure 7, some participants picked more than one design and some liked or disliked the designs equally and decided not to make a choice. The Focus-AID was *liked the most* by the majority of participants (53%), and no participants mentioned it when discussing the design they *liked the least*. Users reported that the color blue was "pretty assuring" and made them feel "safe", thus answering calls they may not have answered otherwise. They also favored the *Authenticated Call* label, which Avail-CID did not have, thus leaving participants with the additional task of interpreting the blue bar whenever Avail-CID was shown.

Focus-Spam was the third least liked design. For some, the color was not alerting enough. Participants expected to see the color red more and indicated that the "black box is off putting". Almost half (48.4%) of the participants suggested that Focus-Spam could be improved if the black box in the layout or background color was changed to red. When asked what stood out in the study experience, half of the participants (50%) mentioned Avail-Spam's red bar and alert message. Since all of the participants understood that the Focus spam design was indicating that the call was spam and seventy-one percent (71%) of participants were able to correctly describe a spoofed call, we conclude that the design elements in the Focus-Spam design encouraged participants to ultimately not follow the warning.

E. Warning Design Discussion

Wogalter's work on warning design criteria [47] and interpretation [48] will be used to discuss the designs in Figure 5. Since the designs used in the experiment used similar signal words, hazard statements, conciseness, and clear instructions, these will not be mentioned below. However, this section will discuss areas in which the designs differ, such as comprehension, notice of consequence, hazard matching, durability, arousal strength, and noticeability.

1) Comprehension: A comprehensible warning is designed so that the user can easily understand and interpret the

message. Although every participant was able to correctly interpret the purpose and meaning of each warning during the debrief, this understanding did not translate into their response. Participants were not confused by the components of the Avail-Spam and the Focus-AID design. However, the Focus-Spam and the Avail-CID designs included an interpretation challenge. Although some participants (15%) liked the Focus-Spam design the most, others were confused about the alert icon used. Participant P03 mentioned that "it ['X' mark] either means that the call is spam or this person is missing a photo," which confused some participants. The Avail-CID design did not offer a new message to participants. Participant P25 mentioned they were " still using the number to make a choice." The blue color is used to indicate a call that is not identified as spam. However, this caused participants to treat unknown calls under Avail-CID and Control design the same way ($p = 1$).

2) Durability and Arousal Strength: Durability originally refers to a warning's ability to withstand wear and tear. In this context, we redefine durability as the ability of the warning to withstand natural human response. Since Authenticated Caller ID is not available for the everyday telephone user, it would be natural to answer spoofed calls from entities that would likely contact you. A warning should change or interrupt this behavior. *Arousal Strength* is the sense of urgency received by a warning and the ability of a warning to motivate a user to take an action [49]. Participants declined significantly more calls from known numbers under Avail-Spam and Focus-Spam. ($p< .001$ in all cases). Participants were able to bypass their normal behavior and answer calls from unsaved numbers or refuse calls from known numbers due to the warning design.

3) Notice of Consequence and Hazard Matching: Hazard Matching is accurately expressing risk using an appropriate warning message. This includes a notice of consequence or adequately expressing the consequence for a specific action. The negative effects of receiving or answering spam calls often push users to download robocall detection applications. Due to this, it may not be completely necessary to include the consequence in the warning. In this study, some participants (17%) mentioned that they liked the fact that the Avail-Spam warning informed them about why a call was being flagged and believed it would have been beneficial for the Focus-Spam warning.

4) Noticeable: At the conclusion of the study, participants were asked to recall what they saw in the study and discuss designs that caught their attention. Every participant mentioned the color red and blue. The *call authenticated label* was noted as a positive characteristic of the Focus-AID warning. However, the color scheme was most noticeable and thus the most favored design in the study. Every participant that stated they liked that design the most because of the color scheme. The Avail-Spam was noticeable and easy to interrupt. Participant P12 said the "red banner across the screen... stood out" which made the decision making process "straight forward." Participants understood these warnings and did not identify any components that were confusing. The Focus-Spam warning stood out to 27% of participants but 47% of participants agreed that although they understood what was being displayed it was not alerting. Participant P31 said the design was a "clear indication of spam," but the "background doesn't scream warning." They wanted a red background instead of the yellow to red gradient. This result contradicts the Focus Group results and is likely due to the amount of time each group of participants had to look at the design. In the Focus Groups and Pilot Testing, participants had more time to look at each design and see what was being presented. During that time, red may have been more alerting alongside the other elements used to warn the viewer of a spam call. However, in quick 23 second intervals, the red is likely more helpful in that it can quickly provide the user with the intended message.

Participants also discussed layout and warning elements that grabbed their attention. The Avail-Spam warning's *Previously Reported in Community* message and the contrast between the background and foreground for the *Potential Spam* label were mentioned as positive elements. However, some participants noted that the warning placement was a bit low on the screen. Participant P27 responded to the placement saying, "it reminds me of an ad" and "it doesn't feel natural." The Avail-CID warning was viewed as similar to Control Design, and therefore the majority of users noticed there was no label to verify what the warning was trying to indicate. On the contrary, the check mark used in the Focus-AID warning, and the *Authenticated Call* label placements were mentioned as positive elements of this design. However, some participants wanted the design to be bigger, saying that the label was not noticeable enough. Similarly, in the Focus-Spam warning, the size and placement of the alert icon were mentioned as positive design attributes and participants wanted the overall warning block to be bigger.

VII. Discussion

As indicated in similar studies [10], [50], [8], warning design can affect user decision making. The results of this study show that the same is true for robocall warnings. Warnings used in this study were sometimes able to effectively change the user's original decision to answer or decline a call. Focus-AID increased the number of calls that were answered through the use of the *Authenticated Call* label. Anti-robocall apps available today do not determine if Caller ID information is valid. However, the results of this study suggest that users want that capability and would go so far as to trust the notice, answering calls they would usually ignore.

Robocalls from Spoofed Known Numbers: Avail-Spam and Focus-Spam decreased the number of spoofed calls from *known* numbers that were answered. However, the impact of each was drastically different. Spoofed calls from known numbers were always answered by 50% of the participants when Focus-Spam was shown. This decreases to 15% of participants when Avail-Spam is shown, which is less than the 53% of participants that said they would answer calls from known numbers regardless of the warning during the focus groups. Also, more calls were declined under the Avail-Spam than Focus-Spam. Many participants stated that the presence of the color red made the difference. Since participants were given 23 seconds to respond to a call, the color red stood out, making it easier for participants to decide. However, it is also possible that the data omitted made a difference as well. The Avail-Spam omitted the name of the caller, whereas Focus-Spam displayed all of the Caller ID information. In the end, only 6% of participants mentioned that the name of the

caller was missing on the warning during the study debrief. Prior work shows that users tend to rely on technology to recall details [51], including phone numbers [52], [53], and many of our participants had to review numbers (N1, N2) in their phone's contact list before saving them in the research phone. Removing the names could have pushed users to rely on numbers, which they may not have memorized, causing them to trust the warning. The removal of the name could have also made the call feel less personal, since the number may not have been recognizable to the participant.

Limitations: Except for the Avail-CID call notice, every user understood what each design was aiming to communicate. However, the presence of the Caller ID and the short amount of time allotted may have affected the success of each warning. In particular, the Focus-Spam design was inspired by the focus groups but was disliked by the interactive survey participants. We believe this is because the two groups had different viewing experiences. The focus group participants looked at their designs and the design probes for at least three minutes each. In the interactive survey, participants saw each call notification for approximately 6 of the 23 seconds given. It is possible that a design element may be more noticeable the longer a user sees it, which may have caused the focus group participants to suggest elements that interactive survey participants rejected. Focus-Spam also included all of the Caller ID information, which gave participant's the option to respond to the design, Caller ID or both. Additionally, the lack of consequences for answering a spoofed call could have also affected the users' response to spoofed calls over time. Future work in this area should investigate the effect time has on user response to warning elements and the effects of Caller ID and consequences on user response to spoofed calls. We show the possibility of multiple visual design factors affecting the participant's response to spam calls. However, due to the limitations of this study, we are unable to pinpoint the design element with the strongest effect but we know that the Focus-Spam design does not work. Additional studies in this area should work to identify which element has the greatest effect. In doing this, researchers should consider the multimodal warning design which includes investigations of Caller ID, ringtones, and vibrations.

Authenticated Caller ID: As of the writing of this paper, Authenticated Caller ID has not been deployed beyond very small test scenarios. Accordingly, it is not clear that all users have an understanding of precisely what this mechanism would provide them. However, the creation of provider-centric (e.g., SHAKEN/STIR) and end-to-end (e.g., AuthentiCall [23], [24]) mean that consumers are likely to soon see such solutions. In fact, SHAKEN/STIR, which provides Authenticated Caller ID for VOIP, will soon be used by all carriers and is currently used by T-Mobile. Whether or not greater awareness will *further* improve the success of these approaches (in particular, related to allowing users to confidently answer calls that are strongly authenticated) remains to be seen.

Lab vs Field Testing: The experiments discussed in this paper were designed to tightly isolate the security indicators of anti-robocall applications. This was critical, as it removed issues related to blacklist quality and environmental stress (e.g., noise, receiving calls while driving, etc.), both of which would likely have a significant impact on the evaluation. As such, the work here can be seen as an early approximation of "best-case" performance of current mechanisms. Similar to other studies [54], [31], the user study setup has high internal validity and thus lower ecological validity [40]. This work investigates the impact of design and limited external factors to maintain this focus. Requesting participants to respond to mock calls in a lab setting allows participants to focus on the task and provide immediate introspective feedback. This allows us to measure cause and effect directly.

TABLE VIII. MEAN REACTION TIME FOR EACH ROUND

Rounds	Accepted Calls %	Mean Reaction Time
R1	41%	2.478
R2	42%	1.774
R3	43%	1.483

TABLE IX. SPOOFED CALL ACCEPTANCE FOR KNOWN NUMBERS OVER ROUNDS

Variables	Accepted Calls		
	R1	R2	R3
Focus-Spam+N1	63%	67%	65%
Focus-Spam+N2	63%	65%	69%
Avail-Spam+N1	25%	35%	44%
Avail-Spam+N2	29%	31%	41%

Future warning designs should be tested in the field to capture these and other factors. For instance, this study was limited by the absence of consequences for answering spoofed calls. As seen in Table VIII, the time that participants used to make decisions decreased the more they saw the designs. However, because participants typically answer calls from people they knew based on Caller ID information, they began to adopt that behavior over time. Participant P14 stated, "At first I declined every call that was labeled as spam. Once I realized some of those calls were from my friends, I started answering them, even when it said spam." This was also true for other participants, which is shown in Table IX. This is problematic as targeted spoofed calls can be easily done. The malicious actor would only need to look at the target's phone number, Facebook or Twitter account to determine what entity they should pretend to be. Participants would need to experience the consequences of answering these spoofed calls to observe their true response over time. Instead, because participants did not experience the consequences of answering a spoofed call from a known number, they began to answer more spoofed calls over time. However, similar to what was found in Tu et. al study [55], spoofed Caller ID affects how participants respond to calls and should be investigated further.

VIII. RELATED WORK

Previous research has investigated the design and use of warnings in other areas.

Felt et al. tested the effectiveness of Android permission warnings [56] and found that the majority of their participants did not pay attention to permission warnings during installation and had low permission comprehension scores. To improve, the researchers suggest conveying risk and other information related to permissions more clearly.

Egelman et al. [31] tested the effectiveness of active and passive phishing warnings on browsers. They found that active

warnings were more effective than passive warnings in preventing users from accessing phishing websites. However, their participants were "highly susceptible" to the spear-phishing attack the researchers deployed. This was due to the lack of dynamic warnings, clear choices, and habituation. *Habituation* is "the extent to which individuals will continue to pay attention to a warning after seeing it multiple times [31]." Majority of participants were exposed to similar warnings in their everyday life and considered them not serious, ultimately ignoring the warning.

To prevent habituation and protect users from dangerous behaviors, Bravo et al. [57] tested the use of inhibitive attractors as a solution. These interface modifications are designed to draw attention to an area and prevent users from choosing until a specific amount of time has passed or a specific action has occurred. Their results showed that, although users disliked experiencing a delay, this method was effective in "reducing the likelihood" that participants would complete insecure actions.

Other research investigates SSL warnings [58], [50], [8], software download warnings [57], warning fatigue [59], indicators [9], browser warnings [10], and malware warnings [60]. From these studies, we can conclude that warnings are an important aspect of user security. Based on these findings, warnings should follow the criteria from Wogalter's research and should support users in reaching their primary security goals [61].

Users want to avoid spam calls, just like they want to avoid phishing and malware, and multiple apps assist users in doing so. To our knowledge, there has been no publicly available study on the effectiveness of indicators for spam call warnings. Warning in the context of spam or robocalls creates a different challenge than those of previously researched area. In other areas, users are often not restricted to a specific time limit in which they need to respond before a decision is made for them. However, telephone users have a limited amount of time to determine if they will answer an incoming call, creating a unique challenge for warning designs. This research addresses this challenge and provides insight for future work in this area.

IX. Conclusion

Cell phone users are interrupted by robocalls daily. Telephone providers and third-party organizations have developed applications to solve this problem by detecting and blocking robocalls. This research surveyed the top third-party anti-robocall apps, identified strong spam call warning indicators for users, and determined the effectiveness of warnings for spam calls. We reviewed the top ten robocall apps and found that 1) they all use blacklists to detect robocalls calls and 2) the majority of spam call warnings used in these apps placed a red bar in the middle of the screen. We then held focus groups which found that all of our participants 1) relied on Caller ID, and 2) desired a spam call warning that uses a checkmark and prohibition sign, along with an alerting background color that fills the entire screen. We applied these design elements to the Available and Focus categories respectively, and compared their effect on users to each other and to the Control category design which had no warning. The results show that warning designs can change user behavior and reliance on Caller ID. In particular, the number of answered spoofed calls decreased by 43% when the spam warning removed the name of the caller and had a red background in the middle of the screen. We believe future research should further investigate the removal of Caller ID, use of multimodal warnings that include visual, audible and physical cues, and how robocall warning indicators perform in the wild.

Acknowledgment

This work was supported in part by the National Science Foundation under grant number CNS-1617474. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] First Orion, "Nearly 50% of u.s. mobile traffic will be scam calls by 2019," 2018. [Online]. Available: https://firstorion.com/nearly-50-of-u-s-mobile-traffic-will-be-scam-calls-by-2019/

[2] Federal Communications Commission, "Stop unwanted robocalls and texts," 2018. [Online]. Available: https://www.fcc.gov/consumers/guides/stop-unwanted-robocalls-and-texts

[3] A. Meek, "The spam call epidemic is stopping people from answering their phones," BGR.com – https://bgr.com/2019/01/29/smartphone-usage-statistics-new-data/, 2019.

[4] R. Dhamija, J. D. Tygar, and M. Hearst, "Why phishing works," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2006, pp. 581–590.

[5] S. E. Schechter, R. Dhamija, A. Ozment, and I. Fischer, "The emperor's new security indicators," in *IEEE Symposium on Security and Privacy*. ieeexplore.ieee.org, 2007, pp. 51–65.

[6] M. Wu, R. C. Miller, and S. L. Garfinkel, "Do security toolbars actually prevent phishing attacks?" in *SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2006, pp. 601–610.

[7] C. Amrutkar, P. Traynor, and P. C. van Oorschot, "Measuring SSL indicators on mobile browsers: Extended life, or end of the road?" ser. Lecture Notes in Computer Science, D. Gollmann and F. C. Freiling, Eds. Springer Berlin Heidelberg, Sep. 2012, pp. 86–103.

[8] D. Akhawe and A. P. Felt, "Alice in warningland: A large-scale field study of browser security warning effectiveness." in *USENIX Security Symposium*, 2013.

[9] A. P. Felt, R. W. Reeder, A. Ainslie, H. Harris, M. Walker, C. Thompson, M. E. Acer, E. Morant, and S. Consolvo, "Rethinking connection security indicators," in *SOUPS*, 2016, pp. 1–14.

[10] R. W. Reeder, A. P. Felt, S. Consolvo, N. Malkin, C. Thompson, and S. Egelman, "An experience sampling study of user reactions to browser warnings in the field," in *ACM CHI*, 2018, p. 512.

[11] P. Gupta, B. Srinivasan, V. Balasubramaniyan, and M. Ahamad, "Phoneypot: Data-driven understanding of telephony threats." in *NDSS*, 2015.

[12] M. Sahin, A. Francillon, P. Gupta, and M. Ahamad, "Sok: Fraud in telephony networks," in *IEEE EuroS&P*, 2017, pp. 235–250.

[13] N. Miramirkhani, O. Starov, and N. Nikiforakis, "Dial one for scam: A large-scale analysis of technical support scams," *arXiv preprint arXiv:1607.06891*, 2016.

[14] H. Tu, A. Doupé, Z. Zhao, and G.-J. Ahn, "Sok: Everyone hates robocalls: A survey of techniques against telephone spam," in *IEEE S&P*, 2016, pp. 320–338.

[15] S. Pandit, R. Perdisci, M. Ahamad, and P. Gupta, "Towards measuring the effectiveness of telephony blacklists." in *NDSS*, 2018.

[16] H. Mustafa, W. Xu, A. R. Sadeghi, and S. Schulz, "You can call but you can't hide: Detecting caller id spoofing attacks," in *44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*, June 2014, pp. 168–179.

[17] V. A. Balasubramaniyan, A. Poonawalla, M. Ahamad, M. T. Hunter, and P. Traynor, "Pindr0p: using single-ended audio features to determine call provenance," in *ACM SIGCSAC CCS*, 2010, pp. 109–120.

[18] F. Maggi, "Are the con artists back? a preliminary analysis of modern phone frauds," in *10th IEEE CIT*, 2010, pp. 824–831.

[19] M. Sahin, M. Relieu, and A. Francillon, "Using chatbots against voice spam: Analyzing lenny's effectiveness," in *SOUPS*, 2017, pp. 319–337.

[20] Alliance for Telecommunications Industry Solutions, "Signature-based handling of asserted information using tokens (shaken): Governance model and certificate management," 2017. [Online]. Available: http://www.atis.org/sti-ga/resources/docs/ATIS-1000080.pdf

[21] H. Tu, A. Doupé, Z. Zhao, and G.-J. Ahn, "Toward authenticated caller id transmission: The need for a standardized authentication scheme in q. 731.3 calling line identification presentation," in *ITU Kaleidoscope: ICTs for a Sustainable World (ITU WT)*. IEEE, 2016, pp. 1–8.

[22] A. Sheoran, S. Fahmy, C. Peng, and N. Modi, "Nascent: Tackling caller-id spoofing in 4g networks via efficient network-assisted validation," in *IEEE INFOCOM*, 2019, pp. 676–684.

[23] B. Reaves, L. Blue, H. Abdullah, L. Vargas, P. Traynor, and T. Shrimpton, "Authenticall: Efficient identity and content authentication for phone calls," in *26th USENIX Security Symposium*, Vancouver, BC, 2017, pp. 575–592.

[24] B. Reaves, L. Blue, and P. Traynor, "AuthLoop: End-to-End Cryptographic Authentication for Telephony over Voice Channels," in *Proceedings of the USENIX Security Symposium (SECURITY)*, 2016.

[25] B. Reaves, N. Scaife, D. Tian, L. Blue, P. Traynor, and K. Butler, "Sending out an SMS: Characterizing the Security of the SMS Ecosystem with Public Gateways," in *IEEE S&P*, 2016.

[26] B. Reaves, E. Shernan, A. Bates, H. Carter, and P. Traynor, "Boxed Out: Blocking Cellular Interconnect Bypass Fraud at the Network Edge," in *Proceedings of the USENIX Security Symposium (SECURITY)*, 2015.

[27] M. S. Wogalter, V. C. Conzola, and T. L. Smith-Jackson, "Research-based guidelines for warning design and evaluation," *Applied ergonomics*, vol. 33, no. 3, pp. 219–230, 2002.

[28] M. S. Wogalter, M. J. Kalsher, and B. M. Racicot, "Behavioral compliance with warnings: Effects of voice, context, and location," *Safety Science*, vol. 16, no. 5-6, pp. 637–654, 1993.

[29] M. S. Wogalter, N. C. Silver, S. D. Leonard, and H. Zaikina, "Warning symbols," *Handbook of warnings*, pp. 159–176, 2006.

[30] T. L. Smith-Jackson and M. S. Wogalter, "Users' hazard perceptions of warning components: An examination of colors and symbols," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. SAGE Publications Sage CA: Los Angeles, CA, 2000, pp. 6–55.

[31] S. Egelman, L. F. Cranor, and J. Hong, "You've been warned: an empirical study of the effectiveness of web browser phishing warnings," in *ACM SIGCHI CHI*, 2008, pp. 1065–1074.

[32] C. Grier, S. Tang, and S. T. King, "Secure web browsing with the op web browser," in *IEEE S&P*, 2008, pp. 402–416.

[33] P. M. Desai, M. E. Levine, D. J. Albers, and L. Mamykina, "Pictures worth a thousand words: reflections on visualizing personal blood glucose forecasts for individuals with type 2 diabetes," in *ACM CHI*, 2018, p. 538.

[34] K. Gallagher, S. Patil, B. Dolan-Gavitt, D. McCoy, and N. Memon, "Peeling the onion's user experience layer: Examining naturalistic use of the tor browser," in *ACM SIGSAC CCS*, 2018, pp. 1290–1305.

[35] A. Naiakshina, A. Danilova, C. Tiefenau, M. Herzog, S. Dechand, and M. Smith, "Why do developers get password storage wrong?: A qualitative usability study," in *ACM SIGSAC CCS*, 2017, pp. 311–328.

[36] K. R. Laughery and M. S. Wogalter, "Designing effective warnings," *Reviews of human factors and ergonomics*, vol. 2, no. 1, pp. 241–271, 2006.

[37] R. Mehta and R. J. Zhu, "Blue or red? exploring the effect of color on cognitive task performances," *Science*, vol. 323, no. 5918, pp. 1226–1229, 2009.

[38] R. S. Friedman and J. Förster, "The influence of approach and avoidance motor actions on creative cognition," *Journal of Experimental Social Psychology*, vol. 38, no. 1, pp. 41–55, 2002.

[39] R. S. Friedman and J. Forster, "Effects of motivational cues on perceptual asymmetry: Implications for creativity and analytical problem solving." *Journal of personality and social psychology*, vol. 88, no. 2, p. 263, 2005.

[40] J. Lazar, J. H. Feng, and H. Hochheiser, *Research methods in human-computer interaction*. Morgan Kaufmann, 2017, pp. 63.

[41] I. Arawjo, C.-Y. Wang, A. C. Myers, E. Andersen, and F. Guimbretière, "Teaching programming with gamified semantics," in *ACM CHI*, 2017, pp. 4911–4923.

[42] A. N. Antle, E.-S. McLaren, H. Fiedler, and N. Johnson, "Evaluating the impact of a mobile neurofeedback app for young children at school and home," in *Conference on Human Factors in Computing Systems*. ACM, 2019, p. 36.

[43] H.-K. Kong, W. Zhu, Z. Liu, and K. Karahalios, "Understanding visual cues in visualizations accompanied by audio narrations," 2019.

[44] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. [Online]. Available: https://www.R-project.org

[45] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins, "The aligned rank transform for nonparametric factorial analyses using only anova procedures," in *ACM SIGCHI CHI*, 2011, pp. 143–146.

[46] S. Flagging and C. Blocking, "its impact on survey research," *American Association for Public Opinion Research*.

[47] M. S. Wogalter, S. S. Godfrey, G. A. Fontenelle, D. R. Desaulniers, P. R. Rothstein, and K. R. Laughery, "Effectiveness of warnings," *Human Factors*, vol. 29, no. 5, pp. 599–612, 1987.

[48] M. S. Wogalter, "Communication-human information processing (c-hip) model," *Handbook of warnings*, pp. 51–61, 2006.

[49] E. Hellier, D. B. Wright, J. Edworthy, and S. Newstead, "On the stability of the arousal strength of warning signal words," *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, vol. 14, no. 6, pp. 577–592, 2000.

[50] A. P. Felt, A. Ainslie, R. W. Reeder, S. Consolvo, S. Thyagaraja, A. Bettes, H. Harris, and J. Grimes, "Improving ssl warnings: Comprehension and adherence," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015, pp. 2893–2902.

[51] C. Thompson, *Smarter than you think: How technology is changing our minds for the better*. Penguin, 2013.

[52] S. M. Nir, "Dumbed-down dialing," 2010. [Online]. Available: https://www.nytimes.com/2010/08/29/fashion/29Noticed.html

[53] K. Lab, "How to survive in the digital amnesia? world," 2010. [Online]. Available: https://usa.kaspersky.com/blog/digital-amnesia-survival/5548/

[54] R. Dhamija, J. D. Tygar, and M. Hearst, "Why phishing works," in *ACM SIGCHI CHI*, 2006, pp. 581–590.

[55] H. Tu, A. Doupé, Z. Zhao, and G.-J. Ahn, "Users really do answer telephone scams," in *28th {USENIX} Security Symposium*, 2019, pp. 1327–1340.

[56] A. P. Felt, E. Ha, S. Egelman, A. Haney, E. Chin, and D. Wagner, "Android permissions: User attention, comprehension, and behavior," in *SOUPS*. ACM, 2012, p. 3.

[57] C. Bravo-Lillo, S. Komanduri, L. F. Cranor, R. W. Reeder, M. Sleeper, J. Downs, and S. Schechter, "Your attention please: Designing security-decision uis to make genuine risks harder to ignore," in *SOUPS*. ACM, 2013, p. 6.

[58] J. Sunshine, S. Egelman, H. Almuhimedi, N. Atri, and L. Cranor, "Usenix security symposium," *Crying Wolf: An Empirical Study of SSL Warning Effectiveness*, pp. 399–416, 2009.

[59] C. Bravo-Lillo, L. Cranor, S. Komanduri, S. Schechter, and M. Sleeper, "Harder to ignore, revisiting pop-up fatigue and approaches to prevent it," *USENIX Association*, pp. 105–111, 2014.

[60] H. Almuhimedi, A. P. Felt, R. W. Reeder, and S. Consolvo, "Your reputation precedes you: History, reputation, and the chrome malware warning," in *SOUPS*, 2014.

[61] A. Sasse, "Scaring and bullying people into security won," *IEEE S&P*, no. 3, pp. 80–83, 2015.

APPENDIX

A. Repeated Advice from Participants on How to Handle Spam Calls

Below you will find the responses received by at least 2 participants on handling spam calls.

1) Check the first 6 digits of the incoming number. If it matches yours, its more than likely a spam call
2) Check the area code and determine if the call is coming from a location you'd expect it to
3) Block numbers that you know are spam callers
4) Send suspicious numbers to voicemail and check the voicemail later
5) If you receive a call from a company, it's okay to hang up. If they really want to contact you, they'll send you a letter in the mail or email you
6) Save number's from companies or people you trust so you'll be sure to answer when they call.
7) If they're asking you for money of the phone its a scam
8) Sign-up for the national do not call registry
9) Report spam callers to the proper authorities or see if your provider can help you
10) Download an app to help you
11) Don't answer calls from unknown numbers
12) Asked to be removed from the spam callers list

B. Focus Group Questions

1) Please state your name, major, and type of phone you have
2) Please walk me through your thought process when you receive a phone call
3) Take a moment and think about all of the spam calls you've answered. Please recall your most memorable experience. If you don't have a memorable experience, please discuss something about spam calls that has stuck with you.
4) You've recently met someone who is having a hard time with spam calls. They ask you 2 questions. Please state your response to each
    a) How do you know if an incoming call is a spam call?
    b) What should I do about spam calls? How do I solve this problem?
5) We will now be taking about 10 minutes to brainstorm independently. I will be asking you two questions and you will need to write and draw your answer. You won't have to show the group, but we will be collecting the papers for our research.
   If you were to use an app to help with spam calls, please illustrate and write what you think it would look like and how you'd expect it to behave
6) I will now be showing you 5 photos. If you saw this on your phone, what would you think it means?

C. Think Aloud Questions

1) You have just downloaded this app, Authenticall, to protect you from spam calls. Without knowing how the app works, please tell me what you think you should be able to do with this application
2) You received a call and see the following on your phone (show device), what do you do? This will be asked multiple times.
3) If you were asked to be a consultant for the AuthentiCall company, what changes would you make to the app?

D. Interactive Study Survey Debrief Questions

1) Can you tell me about what you saw on the cell phone?
2) How did feel while taking the survey?
3) What do you think about the various alerts you saw?
4) Did any of the alerts or things you saw stand out to you? Anything stuck in your memory?
5) What did you do when you got a call from Harold Rogers or Veranda Gardens?
6) What are your thoughts about design 1?
7) What are your thoughts about design 2?
8) What are your thoughts about design 3?
9) Which design did you like the most?
10) Which design did you like the least?
11) Was anything missing from the designs?

E. Photos Shown In Focus Group (Figure 8 and 9)

Fig. 8. Displays what an incoming call looks like without the notifications

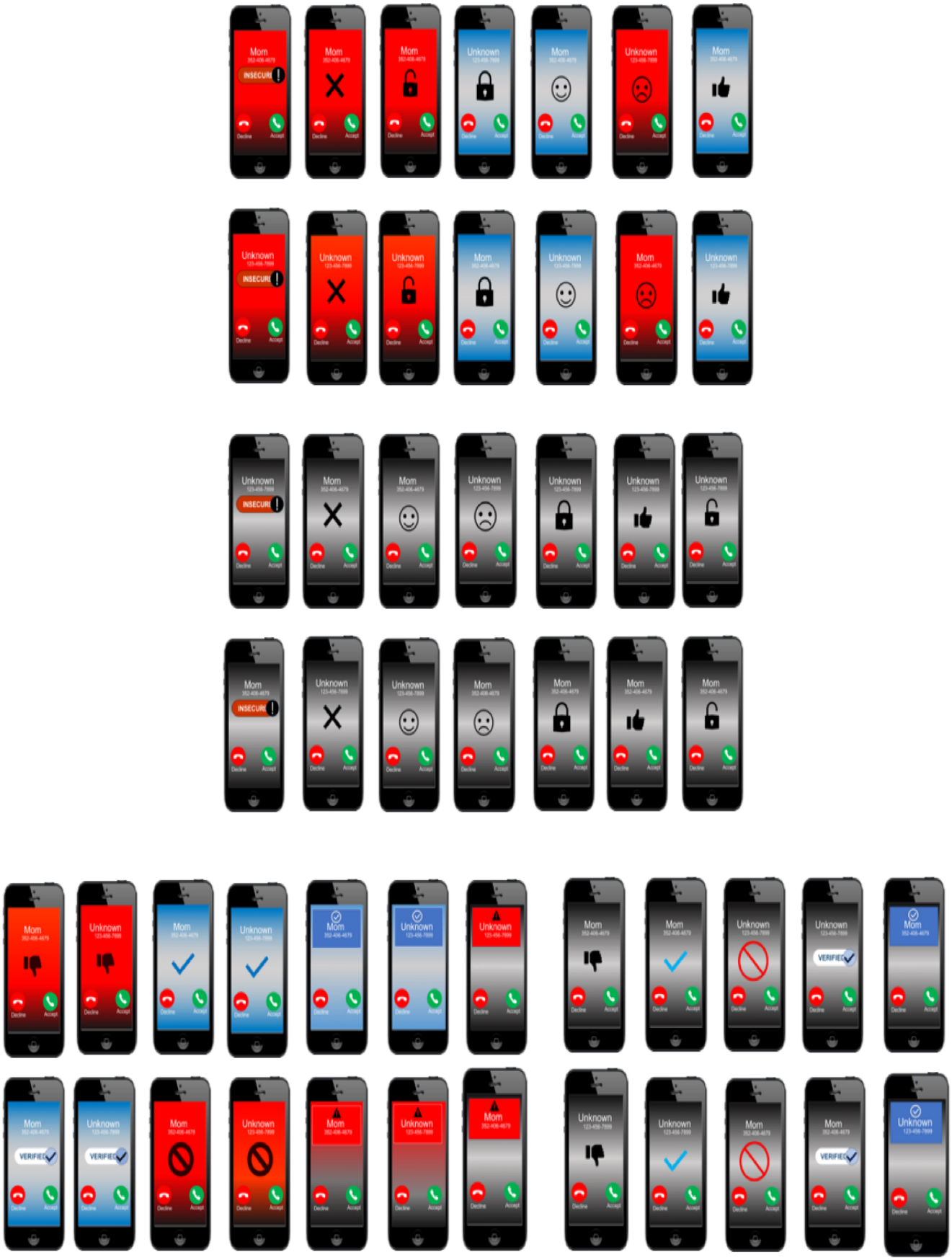


Fig. 9. All of the call notification designs available for use in the focus groups. Each notification was inspired by designs seen in the wild. Five of the 54 designs were shown to each focus group at random.