"Stranger Danger!" Social Media App Features
Co-designed with Children to Keep Them Safe Online

Karla Badillo-Urquiola
University of Central Florida
Orlando, FL USA
kbadillo@ist.ucf.edu

Diva Smriti, Brenna McNally, Evan Golub, Elizabeth Bonsignore
University of Maryland
College Park, MD USA
{dsmriti, egolub, ebonsign}@umd.edu
brenna.mcnelly@gmail.com

Pamela J. Wisniewski
University of Central Florida
Orlando, FL USA
pamwis@ucf.edu

ABSTRACT
Mobile social media applications (“apps”), such as TikTok (previously Musical.ly), have recently surfaced in news media due to harmful incidents involving young children engaging with strangers through these mobile apps. To better understand children’s awareness of online stranger danger and explore their visions for technologies that can help them manage related online risks (e.g., sexual solicitations and cyberbullying), we held two participatory design sessions with 12 children (ages 8-11 years old). We found that children desired varying levels of agency, depending on the severity of the risk. In most cases, they wanted help resolving the issue themselves instead of relying on their parents to do it for them. Children also believed that social media apps should take on more responsibility in promoting online safety for children. We discuss the children’s desires for agency, privacy, and automated intelligent assistance and provide novel design recommendations inspired by children.

Author Keywords
Stranger Danger; Online Safety; Children; Cooperative Inquiry; Participatory Design; Social Media; Mobile Applications.

ACM Classification Keywords
K.4.1 Public Policy Issues: Human safety, Privacy.

INTRODUCTION
According to a 2018 survey by The Family Online Safety Institute [33], parents indicated that “stranger danger” scenarios are their top online safety concern. For many years, the phrase “stranger danger” has been used to teach children about issues in physical safety by telling them to stay away from people they do not know as these strangers may mean them harm [16]. Recently, news media have reported numerous stranger danger and cyberbullying incidents involving young children using social media apps. For example, an 8-year old girl was asked to send naked pictures of herself to a predator posing as Justin Bieber [29] through the TikTok app (previously Musical.ly) [34]. Another 10-year old girl committed suicide after a video of her getting bullied was uploaded to the social media platform by a bystander [8]. These news reports suggest that young children have accounts and use the service to interact with strangers. An investigation by the Federal Trade Commission found that the app was in violation with the Children’s Online Privacy Protection Act (“COPPA”) which requires all online services and websites to collect parental permission for any user under the age of 13 [12].

As younger children are increasingly using internet-connected devices and smartphones [13], they become increasingly vulnerable to online stranger danger risks. Design research to identify new solutions may shed light on risky behaviors and help safeguard children against potential threats. Further, we argue that working with children as end users to co-create online safety features is beneficial to finding solutions children desire and are less likely to try to circumvent. Therefore, we pose the following research questions:

**RQ1**: How do children think about and understand “stranger danger” in online contexts?

**RQ2**: What solutions do children come up with when asked to design mobile social media app features that can help them cope with different “stranger danger” situations?

**RQ3**: What do children think about design-based solutions for online “stranger danger” designed by other children?

To understand children’s perspectives on online stranger danger and discover solutions that can help them manage these situations, we held two participatory design sessions with seven children in each session, ages 8 to 11 years old. Two of the children attended both sessions for a total of 12 child co-designers contributing to this research.

Overall, the design solutions conceptualized by children revealed that they are fairly well-attuned to dangerous online situations and are aware of potentially risky scenarios in which they would need adult assistance.
Nonetheless, the children desired more personal agency, as opposed to constantly being monitored by their parents. Our findings suggest that social media apps should provide children and parents with opportunities to learn together and facilitate conversations around online safety, rather than supporting unlimited parental surveillance and control. Our study privileges children's insights, knowledge, and design ideas to inform the development of novel, child-centered features for online safety. We make the following contributions to research on child online safety and privacy:

- Provides insights on children's understanding of online stranger danger.
- Acknowledges and incorporates children's design-based perspectives for addressing online stranger danger risks.
- Iterates to have children evaluate design solutions from other children for addressing these risks.
- Makes actionable design recommendations for future online safety solutions that promote varying levels of agency, automated intelligent assistance, and education for helping children manage online risks.

**BACKGROUND**

We situate our study within two main streams of research on children's safety: 1) stranger danger in online contexts and 2) co-designing online safety features with children. We also provide an overview of the TikTok app (formerly Musical.ly), which we used as a design probe during the two participatory design sessions.

**Reconsidering Stranger Danger Online**

Today's technological landscape has rapidly brought issues of stranger danger to online contexts. For example, 37% of children between the ages of 11 and 16 years old have experienced trolling, 18% have viewed aggressive and violent content, 12% were the subject of cyber stalking, and another 12% report receiving unwanted sexual solicitations online [19]. To address these concerns, there has been substantial work conducted in the intersection of children, social media, and online safety within the SIGCHI community (e.g., [17,20,31,32]). Studies have investigated parents' and children's sharing preferences, as well as their perceptions of online threats [17]. Others have specifically examined online risks [26–28], though much of this work has focused on adolescents as opposed to younger children.

For younger children, Zhang-Kennedy et al. [32] found that online security and privacy risks for children aged 7-11 was far greater with known family members and friends than strangers, suggesting that stranger danger may not be as great a risk for young children as it is for teens. This was because parents who allowed their younger children to use social media had their profiles set on the highest privacy settings, allowing only close family members to view and comment on their photos. Similarly, a study by Davis and James that focused on middle school “tweens” (ages 10 to 14 years old) found that this group had a greater level of awareness of, and concerns about, privacy in the context of people they know than with strangers [6]. Multiple works have also found that adolescents are more likely to be victims of online sexual solicitation than children [23,30].

While the literature is fairly clear that teens are generally at higher risk of potentially perilous online interactions with strangers, this trend may be shifting as younger children now often have access to smartphones and social media apps [35]. As researchers begin to investigate this apparent shift within the field of child online safety, evidence suggests most children are unable to determine the age and gender of the people they are talking to online, so they tend to be more easily deceived [14]. This may suggest that stranger danger is be a salient online risk for children and that more research needs to explore this changing landscape. Therefore, our research provides insight into children’s perceptions of online stranger danger risks and goes beyond this prior literature by working with children to design solutions that would help them address the problem of stranger danger in online contexts.

**Designing for Online Safety with Children**

Cooperative Inquiry (CI) is an effective method for designing technology solutions that serve the needs of children [3]. CI is a participatory design approach that places children as full partners with researchers [7]. This technique has been used more recently to study online safety for children. For example, McNally et al. [15] studied children’s perspectives on parental control monitoring technologies by having children complete a survey, redesign a commercially available parental control app, and design new features. Another study by Kumar et al. [13] investigated children’s mental models of privacy and security in online contexts. They found that younger children had knowledge gaps and relied heavily on their parents for help when faced with risky online situations.

Therefore, they recommended that parents scaffold children’s privacy and security education by actively mediating their technology use in the home. We draw from and expand upon these related works by focusing specifically on online stranger danger risks within the context of social media apps that are used by children (as opposed to parental control apps used by parents).

Our study contributes to the Child-Computer Interaction literature by investigating what children mean by “strangers” in online contexts and solutions they would find useful for addressing online stranger danger risks. Further, we go a step beyond co-designing social media app features with children to having children evaluate and build upon design-based solutions created by other children. In the next section, we describe a social media app called Tik Tok, which we used as a design probe in our participatory design sessions with children.

**An Overview of the TikTok App**

TikTok is a popular social media app where users create and share short music videos, send personal messages, and create live broadcasts [34]. The app has over 200 million
registered users and has been among the top 100 apps in the App Store for two consecutive years [21]. In 2018, Musical.ly was merged into the existing international platform named TikTok, in-between our first co-design session and our second co-design session. We selected Musical.ly (and later TikTok) for this project as its features exemplify the types of communication common to social media, which can also cause issues of stranger danger [18,36]. Our team considered the app to be a compelling design probe for our child co-design partners, given children’s familiarity with the app and the high visibility of 2017 and 2018 news articles that described how young children under 13 years old were using it [12,29,36].

METHODS
We held two Institutional Review Board (IRB) approved design sessions with children from the University of Maryland’s KidsTeam program.

Study Overview: Cooperative Inquiry
We selected CI [7] as our methodological approach as one of the goals of our work is to understand children's perspectives on their online interactions and to gauge their level of awareness of online stranger danger. Each design session included a team of seven children and ten adult design partners, and followed a similar structure [11]:

Snack time: A 15-minute transition period to orient the team on design work.

Circle time: The intergenerational team responded to a “Question of the Day” to establish context for the design session and was introduced to relevant background.

Main design activity: The team formed ~3 smaller intergenerational groups to complete the session's design activity - detailed in the following sections.

Presentation and discussion ("Big Ideas"): The children from each small group presented their design ideas, aided by the group’s adult team members and by the visual artifacts, to the entire team.

Next, we describe the design activities for each session.

Session 1: “Big Paper”
Our first session was held in December 2017. In this 90-minute design session, all participants first responded to the Question of the Day: “What does Stranger Danger mean to you?” during circle time. Then, the children were introduced to the app Musical.ly (which had not yet been rebranded as TikTok) by watching the first 30 seconds of a video created by Common Sense Media [37]. During the main design activity, small groups used large easel-size sheets of paper (“Big Paper” technique [9,25]) to draw out and annotate features that could be incorporated in the Musical.ly app to help protect children who encounter two stranger danger scenarios: 1) receiving a direct message from an adult stranger posing as a child, and 2) having embarrassing videos of themselves posted to the app by a stranger. Both scenarios were based on and adapted from actual news events [29,36]. Each group received a packet that contained the scenarios with screenshots from the Musical.ly app (Figure 1), markers, and tape so they could iterate on their designs and organize them on the large sheets of paper (Figure 2).

Session 2: “Layered Elaboration”
To evaluate and expand upon the children’s designs from the first session, we held a second, 120-minute co-design session using a set of paper mockups derived from the ideas presented in session 1. This session was held in August 2018. During this session’s circle time, we presented a research poster of the design ideas from the first co-design session. The main design activity for this session employed Layered Elaboration [24], wherein each small group first received a clipboard with one of the three design mock-ups from session 1 (see figures 4–6). Using sheets of transparency plastic to add representations of their ideas without altering the base drawing, the small groups spent 10-15 minutes evaluating and expanding upon the ideas illustrated in each of the mock-ups. Each of the three design rounds were documented on a new sheet of transparency paper to help preserve the ideas through the iterations (Figure 3). At the end of each round we held a stand-up meeting where small groups showed how they adapted their
mockup. In the following section, we describe how we analyzed these research artifacts.

Data Collection and Analysis

The dataset included observational notes from the adult partners, photographs of the sessions, the children’s design artifacts, audio recordings of design activities, and video recordings of the group presentations. During the group presentations of both sessions, an adult team member noted the “Big Ideas” on the white board and conducted an in-situ thematic analysis [4]. Themes from each session were discussed and refined by all the adult designers in team debrief sessions. Two of the adult partners developed debrief documents after each session that contained reflections, images, and notes. Audio recordings were transcribed for further analysis. For our qualitative analysis, we used Braun and Clark’s [4] six-phase guide to thematic analysis, where each of the design ideas (proposed features by children) from session 1 were coded and grouped conceptually (e.g., by functionality or intended user) by the two first authors. For session 2, the authors used the same process to code the children’s evaluations of the feature mock-ups by coding for positive aspects (what the children ‘liked’), negative aspects (what the children ‘disliked’), and suggested improvements (what the children wanted to change). Finally, the authors also coded for emerging themes (e.g., risk assessment, level of agency). The authors had ongoing discussions during this iterative process to establish the coding scheme, define themes, and form consensus. In this paper, we present our results with illustrative quotes to highlight the children’s perspectives and design solutions. Participant quotes are identified by the child’s pseudonym and age (Table 1).

Participant Profiles

The adult participants included students (undergraduate, masters, and doctoral) and faculty. Most were part of UMD’s KidsTeam program, while the first and last authors were researchers from another institution who were invited to lead these sessions. The children came from a variety of socio-economic backgrounds and educational contexts (e.g., public, private, and home-schooled). For the first session, the children were divided by age and sex to consider possible differences in perspectives. For the second session, we focused on making sure children who had participated in the first session or children who were close friends were not in the same group. Since our design sessions were held across two academic years, we had two returning child team members in the second session and acquired five new ones. Table 1 further details the children’s ages, sex, experience co-designing, experience using Musical.ly/TikTok, and CI session participation.

RESULTS

In this section, we begin by discussing the children’s initial perceptions regarding stranger danger. We then summarize the design solutions from session 1. Finally, we present the children’s evaluations of session 2’s design mock-ups.

How do Children Understand Stranger Danger? (RQ1)

During circle time, we asked the children whether they had learned about stranger danger at home or at school. Most of the children (5/7) answered that they had learned about stranger danger at home. Jessica specifically said that her parents initiated these conversations: “I don’t volunteer to talk about it, they come talking to me.”-Jessica, age 11

Three of the children also mentioned learning about stranger danger at school or boy scouts. However, Matt expressed some concern regarding when and where the subject was taught. For instance, he found it inappropriate to learn about the topic in his math class:

“We had this one substitute in math and he’s talking about don’t do drugs, don’t do that, it doesn’t have anything to do with math...I mean, but in math?”-Matt, age 11

We then asked the children and adults, “What does the phrase “stranger danger” mean to you?” A common theme among children’s responses was that they equated
the phrase “stranger danger” to traditional (offline) stranger danger scenarios, rather than online contexts. A recurring example the children offered involved strangers approaching a child in real life: “Maybe when people say ‘Oh you want candy? Come to my car.’” -Mason, age 8.

“They're coming close to you and it's a signal to other people, stranger danger, so it's like a signal to let people know that somebody can hurt you or something.” -Olivia, age 11.

These comments were made even after the team introduced stranger danger in online contexts as the focus of the session. Only adults offered examples of stranger danger online. Although children did not discuss stranger danger online settings, the children demonstrated a strong understanding of the phrase “stranger danger” offline.

**What Solutions Do Children Have for Addressing Online Stranger Danger Risks? (RQ2)**

In session 1, the children conceptualized design features for Musical.ly to help children deal with stranger danger risks. Table 2 provides a summary of the themes that emerged across all three group’s designs, which we describe in more detail below. At the end of this section, we present our mock-ups that were derived from the children’s designs.

**Personal Privacy Features**

The children brought up various privacy features that would help them resolve stranger danger situations. For instance, the groups wanted to be able to block or ban a person whom they suspected to be an adult or a threat. Group 1 designed a “decline” button (similar to a traditional block option) that a child could press to block a friend request or message. Similarly, Group 3 suggested a general reporting feature that would ban videos or users from the platform. The child could report the situation themselves as could others. Many of the children were well-aware of existing privacy features in Musical.ly and even corrected an error in our storyboard regarding how certain privacy features worked:

“But that doesn't make any sense. You guys messed this up... This is his account and this is a friend request. Well guess what? He's public, if he’s public—if he's public you just click follow and you’re automatically following him. You don't have to go through this, only when you have a private account.” -Jessica, age 11.

A key take-away was that many of the children suggested using privacy features that already existed within the Musical.ly app and explained how one should use these features to protect themselves from unwanted interactions.

**Parental Mediation Features**

The children also designed parental mediation features for monitoring and restricting children’s behavior to prevent stranger danger risks. For example, a linked parental account could be used to alert Musical.ly about dangerous situations (Group 1). Group 2 wanted children’s posts to go directly to their parents because they felt parents should closely supervise children: “I wish every time she did that [sent personal info] her parents saw that.” -Matt, age 11.

On the other hand, the older girls (Group 3) suggested that parents and children should be able to monitor each other’s physical location—that is, a child should be able to share their GPS location with their parent, as well as the parent be able to share their location with their child.

**Asking for Help**

Two of the groups (1 and 3) suggested features to ask for help during a dangerous situation (similar to an “SOS” feature in some parental control apps [26]). For example, they wanted a button to alert their parents if they felt that any online interactions were moving beyond what they could control or understand. The difference between this help feature and the parental mediation features was that the child would initiate the action instead of requiring a parent to monitor the child’s activity. The children also designed a button to alert police (“Po-Po”). The children suggested that police use Musical.ly videos as evidence to catch bullies or predators. Similarly, Group 3 suggested having a “Stranger Danger” button to report these types of situations. In addition, Group 3 suggested having an entirely different app that would help children initiate hard conversations with their parents. Olivia suggested a feature that would facilitate communication between children and parents.

<table>
<thead>
<tr>
<th>Privacy Features (3/3)</th>
<th>Parental Mediation (3/3)</th>
<th>Ask for Help (2/3)</th>
<th>Intelligent Assistance (3/3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Decline/Block Features (3/3)</strong></td>
<td>Parental monitoring (3/3)</td>
<td>Parent Alert button (2/3)</td>
<td>Decoding &amp; Flagging Risks (2/3)</td>
</tr>
<tr>
<td>“Decline-can’t send messages ever again.”</td>
<td>&quot;Parent controls, like anytime you posted a message to anyone it will go right to your parents.&quot;</td>
<td>“A button that will alert your parents to let them know.”</td>
<td>“If someone that you don’t know just asks where you live then it will show an emoji”</td>
</tr>
<tr>
<td>Reporting Features (1/3)</td>
<td>Parental Restriction (1/3)</td>
<td>Police “Popo” button (2/3)</td>
<td>“Keywords are ‘Where is your’ or ‘Where do you live?’”</td>
</tr>
<tr>
<td>“If people message mean things to you, you can report them/block them.”</td>
<td>“The purple [button] is for adults so like if someone sends you something really mean you can take it off there.”</td>
<td>“That’s the Popo… so the police will get them and they will go to arrest them.”</td>
<td>App Helps Identify (2/3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stranger Danger button (1/3)</td>
<td>“If it's not a [real] person then it'll show this [red dot].”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“It’s like a whistle and you click on it and it’d be ‘stranger danger! [scream]’”</td>
<td><strong>Bully the Bully (1/3)</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Parent-child Communication (1/3)</td>
<td>“On the bully’s account… they will go delete likes and things”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“It should be an app to let kids talk to their parents”</td>
<td><strong>Recommending Settings (1/3)</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“Musical.ly assists Kenny making an ultra-unguessable account with the same status”</td>
</tr>
</tbody>
</table>

Table 2. Summary of Children’s Design Solutions from Session 1 with Exemplar Quotation
instead of having the problem handled without their input:

“It should be an app to let kids talk to their parents because, you know, how you just start tingling and you just break down.” -Olivia, age 11.

She felt that it might be easier to talk to a parent over a text message (instead of face-to-face) to prevent parents from overreacting. In general, the older girls thought that going to parents for help was difficult because they feared getting in trouble. Therefore, Jessica noted that she would only go to her parents if she felt a situation was serious:

“I do tell them if I think it's gone too far, but if it's like, you know, so tiny and minor, I'm just like 'No, there's nothing to tell you.'” -Jessica, age 11.

Olivia agreed, but she also described a risky incident she encountered online but chose not to tell her parents:

“She sent a provocative video...she looked like she was in a hotel or something...she was facing the mirror and listening to this weird song. She was like in her teens...I didn't do anything about it because I was just like, 'well, this happened a lot to me before.'” -Olivia, age 11.

Both Olivia and Jessica wanted features that could facilitate difficult conversations with their parents without having to feel awkward or scared about it.

**Automated Intelligent Assistance**

All three groups designed features enabling Musical.ly to automatically detect risky content, whether it be words, phrases, or images. The younger children (Group 1) wanted the app to tell them when a situation was “bad” or “dangerous;” so they would know to be cautious. They suggested an angry faced emoji would appear next to inappropriate words or phrases and prevent a child from responding to such messages. The angry face would change size depending on the riskiness of the situation.

“[The app] will show you an emoji that looks like this [angry face], and it won’t let you respond back.” -Liberty, age 8.

Upon reading the first scenario (Figure 2), many of the older children immediately picked up on the idea that Jessy was an adult posing as a child, even though we never explicitly told them this was the case. The older boys (Group 2) instantly said Jessy (the child impersonator from scenario 2) “was a 32-year old man” based on his “fake” profile picture. Jack explained that the typical background of a child’s Musical.ly profile picture would be of their bedroom, unlike Jessy’s white studio background: “It looks like a studio picture. There’s that white background and lighting from all angles.” -Jack, age 10.

For this reason, the older boys wanted Musical.ly profiles to be automatically flagged based on whether the account was real (e.g., the profile belongs to the person it is describing) or fake (e.g., a person is pretending to be someone else). Children suggested that the application could identify the account as fake based on whether the image was taken from another website. If the account were real, there would be a star next to the person’s profile picture. If the account were fake, then a red dot would appear next to the profile picture. The older girls (Group 3) also concluded that Jessy was an adult, based on his profile picture and the writing style of his in-app text messages:

“Jessy is an adult man... because, I mean, he sounds stalkery... You can find those pictures online. I bet you if I brought my phone with me and search up a picture of a boy I could find that picture.” -Jessica, age 11.

They suggested the app could detect key words that insinuate harm, such as “Where is your...” or “Where do you live...” The children’s responses implied they thought it would be fairly simple for Musical.ly to use these types of cues to categorize users as real or fake (i.e., safe or unsafe).

In response to the second scenario (i.e., embarrassing post), the older boys (Group 2) thought the app should help the child delete their old account and make a new one. The app would give step-by-step suggestions to help the child avoid risky situations like those previously experienced. For example, the app would suggest making a username that does not contain the child’s real name. It would also provide suggestions on choosing an appropriate profile image. However, the boys were also concerned about maintaining the child’s reputation and in-app status (e.g., their number of fans and likes). While deleting an account could help a child dissociate themselves from bullies and “bad people,” the boys were clear that they did not want to punish the victim by removing their status within the app.

Finally, the children’s recommendations were not always focused on helping victims, but on educating would-be perpetrators. The younger children’s group suggested the app intervene in negative situations, such as altering a bully’s “mean videos” to “positive” ones. While the older girls’ and boys’ (Groups 2 & 3) desired incorporating educational features in the app, wherein the app would teach bullies how to be good Samaritans.

**Design Mock-ups of Children's Ideas**

Based on the findings from session 1, we developed three mock-ups to reflect the main features the children conceptualized. However, we chose to focus on their novel ideas, as opposed to existing privacy features, which included: 1) Parental Mediation, 2) Asking for Help, and 3) Automated Intelligent Assistance Features. In the first design (Figure 4), a parental monitoring feature allows parents to create their own TikTok account linked to their child/children’s account(s) to view their children’s conversations with other people. Parents also have the option of blocking a person from contacting their child if they deemed the person a threat to their child’s safety. In the child app, a “parent monitoring” status lets the child know they are being monitored by their parent and a message appears when the parent has blocked a contact.
The second mock-up (Figure 5) illustrates a “Stranger Danger” button. The child can press the button, which provides different options for handling the situation, such as calling a parent or emergency 9-1-1, blocking the person, or choosing “other.” Finally, in the mock-up for automated intelligent assistance (Figure 6), the app detects a risky situation, marks it with an angry face, and warns the child about the potential stranger danger risk.

What Did Children Think of These Designs? (RQ3)

In session 2, we had children evaluate and iterate upon the three design mock-ups described above. Since children in the same group often disagreed, we present our analysis by child instead of by group.

Parental Mediation Feature

Overall, most (5/7) of the children strongly disliked the parental mediation feature due to concerns around transparency, authority given to parents, and violating the child’s privacy. While Group 2 agreed that parents seeing messages their child sends and receives helps keep them safer, they were also resistant to allowing their parents access to “all” of their conversations:

“I kind of don’t like that parents can see what you are texting but like... they should be able to see it because what if something happens and you need help?” -Isa, age 11.

Liberty, who also participated in session 1, appreciated that children were notified that they were being monitored:

“Parent is monitoring. That means the parent is watching... Yeah, I like it.” -Liberty, age 9.

However, Liberty and Isa did not like that parents had the power to block anyone they considered a threat. Instead, they suggested that the child, not just the parent, should have the power to block the person. Similarly, Group 1 felt that parents should not block people without giving proper justification to the child. Sophia suggested that the parent and child should negotiate what should be done instead:

“I feel like the parents should have an explanation... for why they blocked it...parents also need to remember that their kid might want some privacy when they’re doing things with their friends. [The child] can talk their parents into like allowing for some privacy on their phone.” -Sophia, age 11.

She explained that the parental control settings could automatically adapt, based on the age of the child, and become more lenient as they grow older. The parent and child would also discuss the appropriateness of the settings as well to enable mutually agreeable adjustments. The rest of the children felt strongly that parents should not be able to see what they are doing online because they wanted their personal privacy. Jennifer explained that parents may misinterpret their child’s online behavior:

“The parent, what they do sometimes is when they look at stuff, they suspect that it’s something different like say, you’re making a joke to someone and they know it’s a joke but then your parent come and read it and they think you’re being mean or the other person’s being mean... and then they like tell you not to use it or on here, they block.” -Jennifer, age 10.

As an alternative, group 2 suggested that parents be allowed access to conversations only when certain “bad words” or...
“bad questions” were detected, such as “what neighborhood you live in?” Parents could add keyword lists when customizing risk detection settings on the app. Jennifer suggested that the app automatically recognize these words and change them to hashtags, so that children do not see them, but parents do not have to intervene:

“I was thinking that instead of having the parent in control feature, the app could like do some of it. Like, first of all, if something inappropriate is said, it automatically changes to hashtags.” -Jennifer, age 10.

Instead, most of the children preferred having direct access to their parents, should they need it. For instance, group 2 added a “Safe Button” to contact their parents when they thought the other person was sending inappropriate messages. This was more similar to the idea of the “Stranger Danger” button, which we discuss next.

**Stranger Danger Button**

Overall, all the children liked having the choice of contacting different types of trusted adults for help. The children agreed that who a child contact depended on the severity of the situation. Group 2 felt that calling a direct emergency number (e.g., 9-1-1 in the US/Canada) was “too extreme.” Isa said she personally didn’t think she would get into a situation where she needed to call 9-1-1, and then only if the situation was “really, really bad.” Liberty agreed and felt more comfortable calling a parent first to receive “parental guidance” on whether to call 9-1-1. They suggested keeping the 9-1-1 option, but having the app or parent confirm whether the situation warranted that:

“When it says call 9-1-1, I think, like, you should press on it… and then like the app should read the text and tell you if this is a right like situation to call.” -Isa, age 11.

Similarly, Group 3 suggested that the app prompt the child before calling the parent or 9-1-1 to give more details about the situation, so it could generate an “emergency text.” For instance, the prompt could include an option to take snapshots of the conversation as police evidence:

“It can call the police, and then the police will go to his house and arrest him. Before that though, it could ask you what happened... so they can show it to the police.” -Elijah, age 7.

Group 3 suggested having the option of adding other trusted adults under the “other” option. For example, Jennifer suggested adding a teacher for when a child is working on a group project or needs help getting out of a cyberbullying incident that involves classmates or contacting another family member when parents are unavailable: “If my mom is busy, I could tell my other sister, or my dad, aunt.” -Jennifer, age 10.

Jennifer and Isa also suggested including a text or email option in addition to calling for help. Regarding visual design options to make the button as intuitive as possible, all the children suggested changing the megaphone icon to something more representative of an emergency, like a red exclamation point, a person with an X, or the phrase “Stranger Danger!”

**Automated Intelligent Assistance Feature**

While most children (4/7) generally liked this feature, they offered several avenues for improvement. For example, Sophia and Liberty felt the warning prompt placed too much trust and responsibility on the child. They felt the app should recommend specific actions the child should take, like blocking or telling a parent, in addition to identifying the potential risk. Their alternative solution was to merge the stranger danger button with the automated intelligent assistance feature:

“We kind of like it, but, it’s a little bit too much trusting of the kid... we should have the options of the last one say if you can block it or maybe not call 9-1-1 but, block it, tell a parent, and all the other things.” – Sophia, age 11.

Isa, on the other hand, was concerned that without parental input as to what “bad words” should be detected, the app would just flag everything as “bad.”

“We were thinking—could like—the app might just think everything’s bad or like things that are not bad, are bad... Parents, like when you get the app, has to put in bad things that other people say that are not good, for them not to be able to write it.” -Isa, age 11.

Finally, none of the children understood (or liked) the angry emoji. They found it confusing and unnecessary. They said that they would rather have the app mark the message as “not so good” in red letters or be more explicit about the nature of the risk detected.

**DISCUSSION**

We discuss the implications of our findings and conclude with recommendations for the design of new technologies that protect children online, limitations, and future work.

**Children’s Conceptualizations of Stranger Danger**

At the beginning of session 1, the children clearly understood that strangers could harm them offline. Their examples of stranger danger focused on physical threats, such as stalking and abduction, and expressed the necessity of having to defend themselves from these strangers. On the other hand, when children talked about being contacted online by strangers, they did not have the same immediate sense of danger. They expressed that these situations were more common online, and some of the children, such as Olivia, gave examples of stranger danger scenarios they encountered in the past. Similar to past research [14], we found that some of the children had a difficult time identifying online stranger danger as imminently harmful in the same way they viewed stranger danger offline. Children underestimated online risks because the technology mediated these interactions in a way made them feel safe from physical harm. Therefore, future research is needed to determine the best strategy for making children aware of
online stranger danger risks beyond being able to characterize online strangers as “real” or “fake.”

**Children Desire Varying Levels of Agency**

Our results further unpack and somewhat diverge from the previous findings of McNally et al. and Kumar et al. [13,15] regarding children’s desire for parental support versus personal agency. Our study confirmed that children understand the need for parental oversight and want assistance from trusted adults when confronted with a risky online situation. Yet, the children in our study were more adamant about maintaining some degree of personal agency and privacy (rather than parental control and restriction) when using social media apps. Even though children in session 1 designed technology mediation features that included parental control, when other children reflected on this design in session 2, most of them rejected the idea of constant parental surveillance and control. This may be because of the different focus in context (i.e., social media app). The key difference being that our sessions focused more on the types of interactions children want to have, rather than what would be useful for a parental app. Our findings highlight the importance of designing more for children’s social media experiences rather than for parental control as the primary interaction design. We also realized that children in session 1 designed the parental control feature for other children, but when children from session 2 evaluated the usefulness of this feature for themselves, their voices on the matter changed. This raises an important methodological insight about using participatory design with children to develop solutions that reflect their needs and desires. To accurately reflect the voices and viewpoints of children in the design of online safety tools made to protect them, future participatory design studies should make a concerted effort to have children both design and evaluate the solutions, as opposed to assuming that the design ideas presented initially by children are the ones they would design for themselves [1].

Instead of parental control features, most of the children in our study preferred self-regulatory [26] features that would assist them in dealing with stranger danger by themselves, through personal privacy features, or with the help of intelligent assistance. Children as young as 7-9 years old (the youngest in our cohort) were attuned to identifying dangerous situations that may occur using social media apps, such as cyberbullying and people using fake accounts for malicious purposes. Nevertheless, the children also acknowledged their own limitations, showing their awareness that identifying or taking action on risky situations may sometimes be beyond what they could or should do themselves. This concern was reflected in their design of “Ask for Help” features. Yet, different children had differing opinions as to whom they would seek help. In some cases, the children relied solely on themselves to take protective action by leveraging existing privacy features. In other cases, children wanted to jointly address the situation with the help of a parent or other trusted adult. Some features connected them with or notified adults, including trusted family members, law enforcement, or application creators. In this way, children often positioned adults as providers of active support.

Interestingly, the children often wanted the app itself (instead of adults) to provide assistance that safeguarded them from stranger danger risks. Children designed for automated intelligent assistance through features that would detect risky content and people, as well as warn them once risk was detected. Yet, children also wanted the app to guide them on what actions they should take, as opposed to a general warning that a risk was detected. Through the design of these intelligent assistance features, children wanted learning scaffolds [13] and in-the-moment assistance [15] to help them address online stranger danger risk situations. The inclusion of such educational and personal control features, while rare [10,26], has the benefit of supporting children’s development on a customizable continuum, from what they can do with assistance to what they can do on their own [5].

**Trade-offs in Transparency and Privacy**

Children clearly opposed constant parental monitoring; however, many children were aware that they could be monitored at any point (e.g., by software, parental co-use). What they appeared to fear the most was a digital version of Foucault’s ideas on Bentham’s Panopticon [38], wherein their digital activities were always potentially being observed, but that they would not know whether or not that was true at any given moment. Personal privacy regarding what their parents saw and knew was an element of this, but that was not the exclusive concern of the children. The fact that several children indicated willingness towards self-reporting interactions demonstrates that the mindset of children was not one of unconditional desire for secrecy from their parents. Instead, children wanted to know when, why, and to what extent they were being monitored. For instance, the children in session 2 appreciated the red bar across the top of the app (Figure 4) that appeared when a parent was actively monitoring the conversation, and red textual descriptions when a parent intervened. Such transparency between child and parent can be an important element of parents’ scaffolding children toward being more cognizant of protecting their privacy when interacting with the outside world through online tools.

**Getting Help from Automated Intelligent Assistance**

Many of the children liked the idea of the social media app monitoring, detecting, and helping them mitigate online stranger danger risks, rather than their parents. Interestingly, in session 1, the children had an implicit trust that the app could detect risky situations to provide personalized guidance. They only questioned the feasibility of such approach once parents were added into the equation or when evaluating the feature in session 2. Children were mostly worried that innocent conversations or actions could be incorrectly flag as dangerous. There was a strong
concern that the app could misclassify a conversation as risky and notify their parents; therefore, notifying a parent would escalate the situation unnecessarily. Similar to prior research [13,27], our participants feared parents would misinterpret the situation, overreact, and blame them. Yet, children still wanted help; thus, they designed solutions that used automated intelligent assistance to alleviate the fear of unfair punishment. An implication of this finding is that we, as adults, may want to reflect in whom and why children place their trust when they need online guidance and support. Advances in automated risk detection and machine learning provide the potential for real-time filtering and intervention. However, these technologies raise concerns regarding how reliable they are, how they should intervene, and the consequences of errors.

Implications for Design
In addition to the designs conceptualized by children and presented as our results, we identified several design recommendations should any of these features be implemented in social media platforms in the future. First, parental monitoring features should be transparent to the child. While children had some expectation of parental monitoring and restriction, they also indicated such features need to clearly incorporate visual clues and terminology related to the extent of monitoring and control. By having a clear, unambiguous, visual indication of when and why they are being actively monitored or blocked from communicating with someone, the level of anxiety that children have about unseen eyes from parents or guardians can be reduced, opening the way for less stressful conversations. Likewise, social media apps should provide parents and children with opportunities for dialogue. For example, based on the children’s design ideas a social media app could notify a parent about a situation the child is facing and suggest topics of conversation. Alternatively, social media apps can have different types of online safety resources, such as videos, that a parent and child can watch together and discuss [13]. Through such dialogue, adults and children can recreate their understanding of safety and create opportunities of shared responsibility [22].

Next, social media apps should notify children and parents about potential dangers and bad actors [2]. Social media platforms can access and analyze user data at-scale, and to this extent, they have more context about their users than any one user has about another. Therefore, children felt that the Musical.ly/TikTok app had a social responsibility to notify them when another user was unsafe or initiating a risky conversation. Here, utilizing automated tools to flag potentially dangerous situations was important, but it is also important to consider the way in which notifications are conveyed (especially to parents). For instance, pairing a notification with a confidence rating and brief description of why the automated system has flagged a message could serve to avoid the escalation of certain types of flagged activities. Related to this, iconography and text should clearly signal the gravity of potentially serious situations.

The children felt there was a mismatch between playful icons (such as an unhappy emoji or a megaphone) and the fact that they were there to indicate dangerous or potentially dangerous situations in some of the mock-ups.

Finally, social media apps should provide a scaffolding that supports and encourages dialogue and education about how to mitigate online risks. With regard to online security, most parents use passive strategies to monitor online use [13], as opposed to the strategies that children sought new designs for, such as prompting active discussion. Instead, we recommend that such apps actively provide parents and children with opportunities for a dialogue about dangers as a part of sending notifications about them. If properly educated and openly supported, most of our participants were willing to self-report dangerous situations.

LIMITATIONS AND FUTURE WORK
Future work should draw from a larger, more diverse sample of children with a more balanced distribution of gender to validate and expand upon the results and design implications presented in this work. Another limitation of our study was that using Musical.ly/TikTok as a design probe may have constrained some of our insights on the topic of online stranger danger more generally. Also, our low fidelity mock-ups in session 2 tended to sway the children’s focus to addressing aesthetics and design elements instead of functionality. We mitigated this shortcoming by asking probing questions, which redirected the focus away from aesthetics. Lastly, this work investigated two scenarios related to news events about a single social media application. Risk scenarios around the varied functions of different social media applications should be investigated, as such work may lead to different understandings of how children wish to address online social media risk. More broadly, future work should investigate how children classify online risks and their ability—and their perceived ability—to address issues themselves, or with automated intelligent assistance, rather than with the support of adults.

CONCLUSION
Through two co-design sessions with children, our work complements and extends youth online safety research by addressing how younger children (7-11 years old) want to approach stranger danger within social media applications and how technology can provide opportunities in response to its effects. It uncovered that children were attuned to these situations, and they wanted to balance having control over situations they may encounter with guidance and assistance in choosing a course of action. It also found that children want to learn about the potential dangers and how to mitigate their risk or address situations they encounter. This work contributes to ongoing efforts to understand and provide support for children’s online activities.

ACKNOWLEDGMENTS
Dr. Wisniewski’s research on adolescent online safety is supported in part by the William T. Grant Foundation.
identifiable data was removed to protect the children’s privacy. All children were audio and video recorded. Finally, all personally identifiable data was removed to protect the children’s anonymity.

**REFERENCES**


**SELECTION AND PARTICIPATION OF CHILDREN**

We had a total of 12 children participants. The children were selected from a wait list, which is always open to new prospective child members, with a goal of balancing the children’s age ranges (7-11) and sex on the team each academic year. The children were between 7 and 11 years of age and had varying levels of previous experience with co-designing (see Table 1). All participants were protected under the study’s IRB approval. We also reviewed the study goals with parents and reminded them that they could stop participation at any time. All parents sign informed consent forms for design team participation, which included consent to be audio and video recorded. Finally, all personally identifiable data was removed to protect the children’s anonymity.


34. TikTok - video social network. Retrieved December 6, 2018 from https://www.tiktok.com/
