Designing Cyberbullying Mitigation and Prevention Solutions through Participatory Design With Teenagers

Zahra Ashktorab
College of Information Studies
University of Maryland, College Park
parnia@umd.edu

Jessica Vitak
College of Information Studies
University of Maryland, College Park
jvitak@umd.edu

ABSTRACT

While social media platforms enable individuals to easily communicate and share experiences, they have also emerged as a tool for cyberbullying. Teenagers represent an especially vulnerable population for negative emotional responses to cyberbullying. At the same time, attempts to mitigate or prevent cyberbullying from occurring in these networked spaces have largely failed because of the complexity and nuance with which young people bully others online. To address challenges related to designing for cyberbullying intervention and mitigation, we detail findings from participatory design work with two groups of high school students in spring 2015. Over the course of five design sessions spanning five weeks, participants shared their experiences with cyberbullying and iteratively designed potential solutions. We provide an in-depth discussion of the range of cyberbullying mitigation solutions participants designed. We focus on challenges participants' identified in designing for cyberbullying support and prevention and present a set of five potential cyberbullying mitigation solutions based on the results of the design sessions.

Author Keywords

Cyberbullying; Social Media; Participatory Design; Adolescents

INTRODUCTION

With the rise of mobile phones and social media, teenagers have become a driving force in the development and design of new tools for interaction with their friends, and are ardent adopters of sites including Facebook, Ask.fm, and Twitter [26]. More than 90% of all U.S. teens 13-17 own a cell phone, while nearly three-quarters (73%) own a smartphone [24]. In recent years, researchers have documented shifts in social media use as teenagers seek "parent free" spaces; while Facebook no longer holds the same appeal for this younger demographic, Snapchat, Twitter, and Instagram are seeing burgeoning growth [23, 24]. On these platforms, teenagers can continue to engage in innocuous correspondences and exchanges. At the same time, many of these spaces have become grounds for a range of cyberbullying activities, which can have serious negative effects on young people's mental health [2].

Research is mixed regarding potential outcomes of social media use by adolescents. Several studies have highlighted both positive and negative correlations between specific behaviors (e.g., frequency of use, engagement in social compensation) and adolescents self-esteem [3, 39, 40]. These spaces also enable posting of mean, unflattering, and bullying content. In a large study of adolescents by Patchin and Hinduja [20], they found that 20% of youth in reporting school districts reported being victims of cyberbullying, and 20% reported engaging in cyberbullying at some point in their lives. In a separate study, the Pew Internet Project [25] found that nearly all teens had witnessed someone being cruel online, while 15 percent said they had been the recipient of mean comments in the last year. Victims of cyberbullying may experience significant emotional problems, including anxiety and depression [19, 21] and, in extreme situations, they may commit suicide.

Designing for teenagers is challenging, as their motivations for using these spaces and the emergent norms that develop among teenagers are often significantly different from their adult counterparts [16]. Because of this discrepancy in social media motivations, designers may overlook factors and features that facilitate cyberbullying or harassment. In recent years, the few attempts of designing for cyberbullying prevention have not included the perspective of those most affected: young people [10].

Participatory design represents a potential solution to this gap in understanding and may help designers build novel tools [44]. In this study, we extend existing participatory design techniques with adolescents [12, 13, 14, 18, 44] to the cyberbullying mitigation domain. We share our methodology and findings from participatory design sessions conducted with ninth and twelfth graders at a private high school in a large metropolitan region in the U.S. In working with these two groups of teenagers, we systematically sequenced a range of design techniques over five sessions with the goal of prompting participants to both explore potential solutions for the larger problem at hand (cyberbullying across all social platforms) and smaller sub-problems (specific types of cyberbullying across specific platforms). These findings are synthesized in the discussion section, with a focus on

challenges to designing for support and potential avenues for future design work in this area.

RELATED WORK

Increasingly, social media plays an important role in the social lives of children [12, 32]. While interacting with peers contributes to a child's growth, development, and well-being [40], directed malice and cyberbullying are stark realities of using social networks. Children who are experiencing and engaging in cyberbullying can be viewed as domain experts of cyberbullying. Currently however, there has been little published research that involves children and adolescents, those most affected by the malice of cyberbullying, in designing and building technology to mitigate the effects of cyberbullying.

Participatory Design and Youth

Partnering with children as design partners has led to technologies being better suited for the needs of children. There have been quite a few participatory design techniques for children implemented and evaluated to design new technologies. Druin et al. introduced cooperative inquiry by interpreting participatory design contextual inquiry methods for children as partners in the design process [13] The Mixing Ideas technique is used to encourage collaboration during the design process [18]. Comics and roleplaying through techniques like KidReporter demonstrate diverse ways to elicit information from children [6].

With the burgeoning use of social media platforms by teens, the importance of incorporating teens in the design process is being realized. Fitton et al. [16] included several ongoing studies that are involving teenagers in their design processes in order to yield better interactive products. A recent study on Snapchat highlighted the importance of including teenagers in the conversation since they are domain experts on some of these social media platforms [36]. Similarly, another study found it beneficial to involve teens in the design process for developing applications that aim to prevent unhealthy habits [31].

Designing Cyberbullying Mitigation Tools

Until this point, there has been very little work on the design of technological solutions for cyberbullying. For instance, there has been work on cyberbullying that focuses on community involvement and parental responsibility to address the problem (i.e., education) [9]. Dinakar et al. [10] introduced "reflexive interface" prototypes as a means to prevent cyberbullying across a limited range of subjects, including appearance, intelligence, racial and ethnic slurs, social acceptance, and rejection. The reflective interface encourages positive digital behavioral norms and consists of the following interactions in order to deter malicious behavior: notifications, action delays, displaying hidden consequences, system-suggested flagging, and interactive education. The reflective interfaces to mitigate cyberbullying did not involve youth in the design or evaluation processes.

User-Centered Design and Cyberbullying

Traditionally, users have been placed in a reactionary role in the design and development process of technologies [30]. The pitfall of such an approach is that users are only reacting to what designers are creating and are not offering their own designs or solutions. User-centered design seeks to introduce users to the design process in the earlier stages so that they can influence the design of the tool. Many methodologies bring users of technology into the design and development process. Users have been involved in all stages of development including co-designers, testers and subjects.

Children provide valuable insight in the design process. Increasingly over the past decade, researchers have included children more actively in the design process, successfully demonstrating that children offer fresh perspectives that lead to innovative designs in technology [14, 35]. However, in the realm of cyberbullying, very few design innovations have been introduced through participatory design. Bowler et al. conducted a narrative-inquiry based study in their participatory design sessions [7]. The narrative inquiry was the only participatory design methodology employed to develop their seven emergent themes for cyberbullying mitigation: design for reflection, design for consequence, design for empathy, design for personal empowerment, design for fear, design for attention, and design for control and suppression. While this methodology is valuable, it is only an initial step in the right direction. To move this area of research forward, we conducted five participatory design sessions with ninth and twelfth graders. We employed multiple participatory design techniques in our participatory design sessions including: Focus Groups, "Bags of Stuff" and [13, 18, 44]. A narrative-inquiry based study limits participants to a singular narrative and thus confines their capacity for introducing novel solutions. Our diverse techniques allowed our teen co-designers to consider all types of cyberbullying, and how the various types are enacted the context of different social network platforms.

Existing Cyberbullying Mitigation Tools

Existing literature reflects that there have not been many formal studies focusing on the design of cyberbullying mitigation tools in the realm of research. However, many independent developers have created tools to help promote well-being of social media users. For example, the application "You're Valued" searches Twitter for tweets that say "nobody likes me" and then sends a response tweet with messages like "I like you", "You're valued", or "You matter" [42]. Another application, Honestly, looks to combat cyberbullying by asking friends of a particular user questions about a person like "Can I sing well?". In an attempt to boost the self-esteem of the user, only the positive responses are shared with the user [38] to boost confidence and self esteem of the recipient. While the intent of all of these applications is to mitigate low self-esteem and low confidence (one of the effects of cyberbullying), none have included children and adolescents in the design to process to gauge the potential impact of such interventions..

METHOD

To enable teenagers to become a part of the design process, the research team collaborated with a local private high school during the 2014-2015 academic year. The research team worked closely with school administrators to develop a survey instrument and identify classes that would be able to participate in multiple sessions during the spring semester. The team chose to work with high school age youth (14-17 years old) because they are legally allowed to use social media applications [33] and are more likely to own smartphones. In addition, 9th and 12th graders are at different life stages and likely have different world views toward technology.

After receiving informed consent from parents and child assent forms, we began running design sessions during participants' "open period" averaging one session per week. We worked with fourteen 9th graders and seven 12th graders over the course of five weeks, using a range of participatory design techniques, such as "Bags of Stuff" and Mixing Ideas [13, 18, 44]. We purposefully chose a diverse set of PD activities to encourage the participants to consider multiple aspects of cyberbullying and mitigating solutions.

Below, we provide additional details on the descriptive data collected from the participants through a survey, as well as a breakdown of each of the sessions.

Survey Details

Before the PD sessions began, we surveyed our 21 prospective participants to understand how they interact with social media and different technologies, and to provide contextual information to tailor our design sessions to their personal experiences.

The participants were heavy technology users. Everyone owned a cell phone and all but one (a 9th grader) owned a smartphone, and they spent a large portion of their day using their phones–eight participants said they spend more than two hours a day using their cell phone. Regarding other technologies, they spent, on average, about one hour per day using a computer or laptop and an additional 1-2 hours per day using a tablet (which they used in school). Only six participants reported that they played games regularly. Looking at their use of various social media platforms, Snapchat (80%) and Instagram (76%) were the most popular by far among overall, while less than half said they used Facebook, Google Plus, Tumblr, Vine, or Twitter. Everyone in the 9th grade group used Snapchat (compared to just three of the 12th graders); on the other hand, 12th graders were significantly more likely to use Facebook than 9th graders (71% vs. 36%).

Regarding past experiences with bullying and cyberbullying, five of the 21 participants (24%) said they had personally experienced cyberbullying, while eight participants (38%) said they had at least one friend who had been cyberbullied. In total, just under half (48%) of participants had personal experience with cyberbullying. For the five students who had personally experienced cyberbullying, one said there had only been one incident, while the other four participants said

Item	Mean	SD
Many people in my school engage in cyberbullying.	2.33	1.49
Many people in my school have been victims of cyberbullying.	2.95	1.77
People in my school don't step in when someone is being cyberbullied.	3.19	1.87
When a cyberbully has been caught, he/she was punished by the school.	4.35	1.60
Most of the cyberbullying I've seen occurs anonymously.	4.10	1.22
People in my school are quick to report cyberbullying to parents, teachers, coaches, etc.	4.24	1.55

Table 1. Participants' attitudes toward cyberbullying at their high school (Response scale: 1=Strongly Disagree, 5=Strongly Agree).

they had been cyberbullied a couple times. Cyberbullying incidents reported by these participants occurred most frequently through text messages (three participants reporting this), with Twitter, YouTube, Ask.fm, Instagram, Kik, and online games each being selected by one participant. In dealing with these incidents, the participants said they either did nothing or turned to a friend for support. Below, we show a sampling of comments participants provided when asked to describe a time when they or their friend had been cyberbullied.

- *"People took pictures and made fun of my friend. Also bullied her on Twitter."*
- *"Someone called my friend fat, and mean words."*
- *"I was online playing Minecraft and this kid just starting cursing at me because i won the game and he lost and now i go on 'no cursing' servers on Minecraft."*
- *"I was at a different school and people were making fun of me on twitter and with their other friends. People took pictures of me and made fun of me."*

In general, the participants who participated in this study reported that cyberbullying was not a problem at their school, but when it did occur, participants reported the incident and the school administration was likely to step in. Table 1 includes responses to five Likert-type items about their perceptions of cyberbullying at their school (Scale: 1=Strongly Disagree, 5=Strongly Agree). There were no significant differences between responses from the two groups.

PD Sessions: Our Design Partners

We conducted a total of ten sessions, five each with two classes at a local high school. The ninth graders were ages 14 and 15, while the twelfth graders were ages 17 and 18. Sessions were constrained to the times during which participants had "free periods," which typically lasted 45 minutes. Since cyberbullying is common in high school settings [36], we chose freshmen and seniors as are co-designers; this allowed us to gain insights into how perspectives vary between younger and older adolescents, who have different degrees of access to technology and social media [27]. Our ninth grade sessions consisted of 14 participants, 10 of whom were female and four of whom were male. Our twelfth grade class consisted of seven participants, five of whom were female, two of whom were male. All of the participants reported through our surveys that they had been active on some social media platform. We applied the

same structure to both sets of participants across the five sessions. Three adult researchers were present at each session to facilitate discussion and collaborate with the participants in creating new design ideas.

The Design Activities
Each of the five sessions conducted with the participants focused on specific aspects of participatory design. These were: (1) Focus Groups, (2) Scenario Centers, (3) Bags of Stuff, (4) Mixing Ideas, and (5) Evaluating Prototypes.

Session I: Focus Group
In our initial session, we held a focus group to familiarize ourselves with the participants' environment and to familiarize them with the researchers. In these sessions, we stressed that the participants would not be graded on their performance in the sessions. Additionally, we explained that we were studying online harassment and we wanted to work with the participants for the best solutions to fix the problem of online harassment. Our goal was to understand how the participants interacted with social media platforms and how these platforms might be being used for cyberbullying. Participants sat in a large circle and the moderators of the sessions asked questions about social media practices.

Session II: Scenario Centers
In the second session, the researchers developed a set of "Scenario Centers" to help participants begin to think about the specific scenarios they would be designing for. The concept of "Scenario Centers" stems from the childhood experience of *center time*, in which participants learn and engage in different experiences in "centers" [18]. Participants in each grade were presented with scenarios from the different social media platforms based on the themes that emerged during the focus group session. Participants got into groups to discuss each scenario, then were asked to think critically about possible technological and non-technological solutions to mitigate the negative behaviors. The goal of this question was to prompt them to begin thinking about technological mechanisms (both existing and non-existing) on the social media platforms that would aid in helping the victim or prevent the bully in the situations presented them in the centers.

Our scenarios included diverse social media platforms and representative of different types of cyberbullying including: Flaming, Harassment, Cyberstalking, Denigration, Outing and Trickery, and Exclusion [43] and to include the social media platforms on which participants indicated they were most active. Below we describe a subset of the scenarios.

1. **Facebook (Denigration)**: Sara is a new girl at school who dresses differently than the other kids. She is quiet and introverted so she has had trouble finding friends. Another kid at school starts taking pictures of Sara and posting them on a group on Facebook, "Sara's Weird Outfits." The page has over 1000 likes and people start to comment on the strange clothing Sara wears to school. The comments seem to keep getting more malicious and personal. Recently, someone wrote: "She's so ugly and her style sucks. She needs to die."

2. **Snapchat (Flaming)**: Kyle keeps receiving repeated snapchats from Tom and Jake calling her names. They update their public stories which video messages of themselves saying "Kyle is ugly" or "Kyle needs to die." These videos are sometimes coupled with captions. Tom and Jake also send direct snapchats to Kyle.

3. **Ask.fm (Harassment/Cyberstalking)**: Jenna keeps receiving repeated anonymous messages on her ask.fm account: "Go kill yourself" and "No one likes you." She responds to these messages to show that they aren't affecting her. Because the messages are anonymous, she doesn't know if they are coming from multiple people or just one person.

4. **Instagram (Exclusion)**: Jenny, Kayla, Sara and Felicia are all very good friends and have lunch together at school daily. Recently, however, Jenny has been avoiding Felicia. Jenny begins posting multiple pictures on Instagram in which she crops out Felicia and only tags Kayla and Sara. Felicia is feeling sad about how Jenny is excluding her and does not know how to react. She does not know if it was something that she did to make Jenny feel and act this way.

   Tom and Jake had an argument recently and things have escalated online. Jake keeps tweeting threatening posts saying how he's going to "beat the crap" out of Tom. Things get even worse when Jake starts tweeting where Tom normally is during certain times of day. These messages are repetitive and Tom is afraid he could be ambushed at any time.

5. **YouTube (Outing and Trickery)**: Frank secretly records video of Sara and Ben getting intimate at a party, then posts it to YouTube the next day. The video goes viral within the school and is shared on all the social networks. The situation escalates when people from school begin commenting "sl*t"and "wh**e" on the video.

Session III: Bags of Stuff/Low-Fidelity Prototyping
In our third session, we employed a "Bags of Stuff" [15] design method to allow participants to design low-fidelity prototypes to address a specific cyberbullying issue. The researchers presented each group with a broad description of a cyberbullying even and instructed participants to utilize the provided art supplies (markers, white paper, construction paper, pipe cleaners, and stickers) to design a solution. The goal of "Bags of Stuff" is to allow children to feel that *anything* is possible in design. In the previous sessions, where participants considered different scenarios, it became clear that participants were limiting their discussions of potential solutions to tools that already existed. Thus, the research team that "Bags of Stuff" would encourage the participants to stretch their thinking and be creative in their solutions.

The scenario presented to both groups of participants included the range of cyberbullying activities outlined in previous research [43] and incorporated the various social media platforms identified by the participants in previous sessions.

"Bags of Stuff" Scenario

John is a victim of various types of cyberbullying on all of the social media platform of which he is a part of. He keeps receiving repeated vulgar, offensive, or insulting messages/comments/snaps/posts on Snapchat, Facebook, Twitter, Instagram (respectively). John feels sad and depressed from all of the negative messages that are being received. John fears for his safety because he is repeatedly receiving threats online.

People are also posting cruel gossip or rumors about John to tarnish or damage his reputation. Students in his school have created a group on Facebook, a standalone website, or page dedicated to insulting him. Students in his school have also hacked his social media (Twitter, Facebook, Instagram) and are posting as him in an attempt to get John into trouble or make him look bad. John has also been tricked into giving personal or embarrassing information and pictures to Mike who shared it with everyone on Instagram, Facebook and all other social media. John is intentionally being excluded from his friends' online chat group on Facebook.

Participants were instructed to create a tool or application with their available materials that would address one or more of the following: 1) Prevent the cyberbullying behavior from happening, 2) Ease the emotional pain for the cyberbullying victim, 3) Stop the cyberbullying behavior from happening again (once it has already happened), and 4) "Solve" this problem in another way. They were told that the tool they created could be a separate application/website not associated with any particular social media platform AND/OR it could be incorporated within any or all of the social media platforms previously discussed.

Session IV: Mixing Ideas
In the fourth session, we reviewed the prototypes that participants had created and used the design methodology of "Mixing Ideas" to create new solutions. "Mixing Ideas" further fosters collaboration between participants by encouraging them to think about common themes between their solutions to create better solutions and prototypes [18]. With the researchers acting as discussion moderators, the classes then identified a set of prominent themes across groups, then got back into groups and spent the remainder of the session to further refine their prototypes.

Session V: Evaluating Prototypes
In the final session, the participants discussed the feasibility of each of the designs they had created by addressing strengths and limitations to implementation. While discussing each of the prototypes, the research team posed the following questions: "How would this work in real life?", "Is this implementable and useable?", and "Is this solution ethical?" While the participants were encouraged in the previous sessions to think "outside of the box with an open mind", the goal of this last session was for the designers to think about the *implications* of their designs.

RESULTS
Across our two groups of co-designers, differences in social media platform use affected the perceived "coolness factor" of various platforms. In sessions, participants

unanimously echoed their classmates (and the results of the preliminary survey) by focusing the most attention on Instagram and Snapchat. The participants responded that Facebook and LinkedIn were the most "uncool" social media platforms. The appeal or "coolness" of a particular application is relevant when designing applications because of the bystander role in a bullying scenario could be played by an adult. One student said, "I go on Facebook because these old people relatives don' have any other social media. So I go to wish them happy birthday on there." Another student said, "Google Plus–who even uses that anymore?" The co-designers' responses to the different types of social media demonstrated that our design focuses for cyberbullying mitigation should focus on the affordances of social media platforms that are more widely used by demographics who are most affected by cyberbullying.

Negative Experiences on Social Media Among Design Partners
One of the main takeaways from the focus group sessions was understanding the co-designers familiarity with various social media platforms, the norms of interacting in these spaces, and what they perceived to be bad behavior. When asked about negative experiences on Snapchat, one student expressed that it is concerning when the recipient of a "snap" screenshots the conversation because Snapchat, unlike many other social media sites, affords ephemeral interactions. When someone takes a screenshot of a post, they violate the sender's trust by breaking the "unwritten rules of Snapchat." For Instagram, one student said users will create "hate pages" by posting screenshots of someone's social media posts and "making fun of the person." With regards to malicious behavior, the participants mentioned that the silencing of dissenting opinions on Tumblr could occasionally "get out of hand." One student reported that some trending topics like "#stopblackpeople" or "#stopwhitepeople" start out as a joke and then just turn into racist posts as they spiral out of control. Regarding Tumblr, one of the participants said, "I hate all the racism and stereotyping that happens on there."

Design Applications
Below, describe the nine design solutions created by participants as a result of the prototyping sessions. The first set of seven were developed during Session III. Figure 1 includes visualizations of four of the prototypes.

1. **SMILE: Social Media Informative Life Empowerment**: SMILE is a third-party application that addresses the lack of control social media users have over what content *other users* can post to their pages. Users first create a "buzzword" list containing potential sources of bullying or harassment. When a user receives a comment/post on their profile that includes one of their identified buzzwords, they receive a notification and are given the option to accept or reject the post on their profile. This particular social media application enhances the user experience by allowing the user to choose the content that becomes visible to other users. Furthermore, in the realm of automatic detection, detecting cyberbullying content just using expletives is often ineffective because youth will
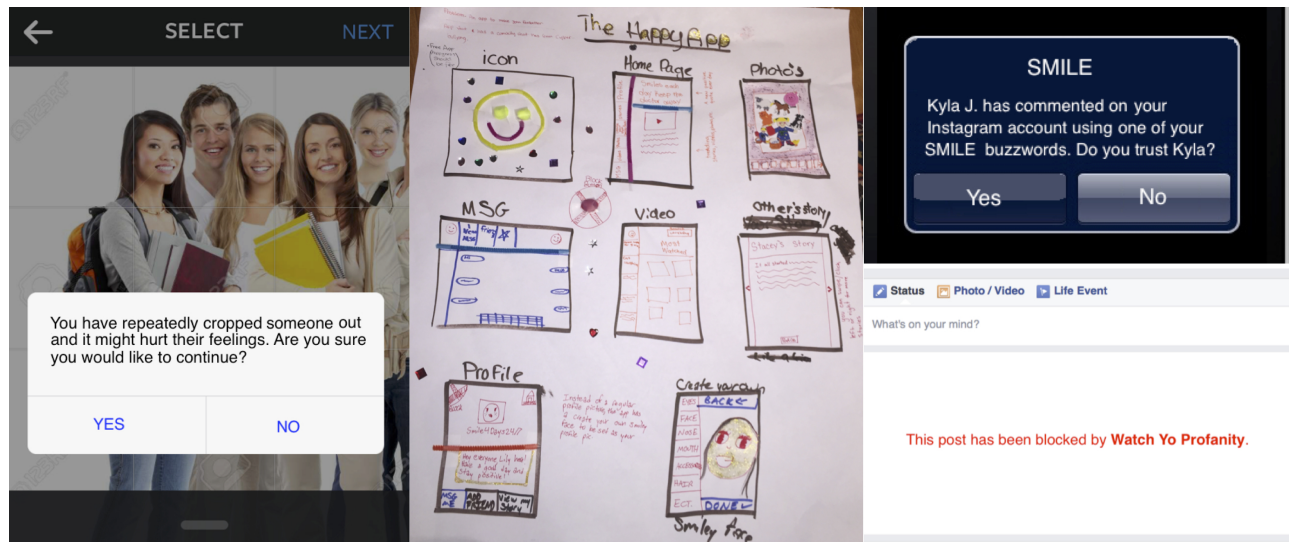
Figure 1. Application prototypes from sessions with participants, including Exclusion Prevention, Happy App, SMILE, and Watch Yo Profanity

use expletives emotionally [10]. One of the participants pointed out that if her best friend used an expletive in a post, she would have no concerns about accepting it because she trusts the source of the content. In this sense, SMILE offers an innovative solution to the automatic detection cyberbullying problem.

2. **"Happy App"**: Participants designed the Happy App for individuals who are upset or depressed due to cyberbullying. Users create their own profile and can share their experiences with cyberbullying. Interactive features allow users to connect with other cyberbullying victims and obtain peer support, which has been shown to promote psychological well-being and positively influence self-esteem [28]. The designers also suggested having a positive quote of the day on the app's homepage.

3. **"Fight Back"**: This application is designed to help people who have been cyberbullied. Features include a chat room to connect with trained therapists or a friend of your choice to speak with about your experiences and get advice on responding to the harassment. Users can block people they don't want to connect with through the app. The app also has a "happy room" that includes positive messages to mitigate cyberbullying harm (similar to the Happy App's homepage).

4. **"Exclusion Prevention"**: Repeatedly excluding someone is a form of bullying [43]. Exclusion can manifest in multiple forms on social media. Specifically, repeatedly excluding someone by cropping them out of a picture before posting it may make someone feel excluded and hurt their feelings. Using face-detection technology, the teens created an application that alerts a social media user when they using the cropping feature on sites like Instagram and crop out one or more people in the picture. A message pops up before the picture posts to the site, telling the user that cropping people out of the picture could hurt their feelings.

The user then decides whether they want to continue with posting the picture.

5. **"Watch Yo Profanity"**: This plugin features a filter that blocks out expletives and vulgar phrases from appearing on social media using an existing dictionary of words and phrases. Users can further customize the filter to block additional words and phrases that are not in the plugin's database. Users can also block specific people in social media, hiding all content from those users. For example, if someone keeps posting inappropriate content on social media, a plugin user can opt to block that person.

6. **"The Broiler"**: One student indicated that he had already chosen not to go to a specific university because of a campus-based website where participants could anonymously post mean comments about other participants at the school (in the same vein as the short-lived "Juicy Campus" website). As a means to fight it, he proposed an application or Twitter add-on called "The Broiler," that "would roast whoever is roasting others" on social media.

7. **"Reporting Bullies With Feedback"**: Participants discussed their dismay at the lack of feedback they receive when reporting abuse on various social media platforms. When they alert the site about negative content, they want to be notified not only of the abuse, but also receive feedback about how the situation was being handled and additional information about the victim post-abuse. They designed a feedback tool that reported back to a user that had reported malicious content about the current status of bullying with the user who may have been effected and if that user requires additional emotional support.

During the "Mixing Ideas" sessions, two additional prototypes were generated:

1. **Positivity Generator** This application was motivated by the "Watch Yo Profanity" application designed in the third

session. While "Watch Yo Profanity" merely censors potential cyberbullying content, the Positivity Generator *replaces* instances of negativity with positive and uplifting quotes from a selected celebrity. During the session, users chose Kanye West, a celebrity renowned for his self confidence [8]. In the sessions, the participants introduced the notion of expanding the "Kanye West Self-Confidence Generator" [1] to include self-confidence enhancing quotes by popular celebrities to boost the self-esteem of a cyberbullying victim. The participants suggested that a user's favorite artist could be inferred from their social media activity and that celebrity's most uplifting and encouraging sayings could be used replace negative content.

2. **"Hate Page Prevention":** The participants mentioned that the same face detection technology used for the "Exclusion Prevention" feature could also be leveraged on Instagram to automatically discover and report "Hate Pages". "Hate Pages" can be defined as pages where a cyberbully posts screenshots of another user's pictures and uses malicious captions under those photos to harass the person. The co-designers had a keen understanding of facial recognition software available through platforms like Facebook through their experience with tagging [5]. They proposed a monitoring system to prevent hate pages. If a user's photos looked too similar to another's then the page would be automatically flagged and investigated by platform administrators.

DISCUSSION
The wide range of cyberbullying scenarios we discussed with our design partners prompted them to consider forms of malicious online behavior that may not traditionally be deemed as cyberbullying. Below we discuss in more detail how our participant designers conceptualized cyberbullying and how their proposed solutions may be enacted in meaningful ways.

Defining Cyberbullying
In order to start designing for cyberbullying, the research team explored whether our co-designers were in agreement about the definition of cyberbullying. Participants unanimously agreed that all of the scenarios constituted cyberbullying, except for *Instagram (Exclusion)* in which a girl is continuously cropped from photos on Instagram. A heated discussion emerged among the participants as to whether cropping a user out of photos "is just rude and not targeted enough" or if it is more severe. The participants who argued that this scenario did not constitute cyberbullying claimed that in order for something to constitute as cyberbullying, 1) it must go "viral" so as to include a wide audience and 2) it must be directly targeted. For the male participants, the *Instagram (Exclusion)* scenario was missing these two components. One male student said, "Do you know why it's not bullying? Because [the girls] are still sitting with her when they take the photograph. It is still peaceful." One female student who adamantly believed the scenario constituted as bullying countered with, "Making someone

feel insecure about themselves is bullying too. Bullying isn't just physical."

While many researchers have surveyed youth to understand the climate of cyberbullying activities [9], exclusion cyberbullying has been largely ignored in cyberbullying research. Our study sheds light on the emotional trauma this kind of cyberbullying causes since one of our participants claimed that she had experienced exclusion repeatedly online.

Designing for Support
Many of the prototype designs fall under the emergent themes described by Bowler et al. [7]. For example. "Hate Page Prevention", "Reporting Bullies with Feedback", "SMILE", and "Watch Yo Profanity" fall under "Design for Control and Suppression," a theme that involves controlling content through a social media platform's administrators or a third party algorithm. "Exclusion Prevention" creates a pause in the cyberbullying process by asking users if they want to proceed by excluding someone from a picture, which is in line with Bowler et al.'s [7] "Design for Reflection" theme. Finally, "The Broiler" (though ethically questionable) and "Reporting Bullies With Feedback" are designed for consequence, because they ensure that their are consequences for bullying behavior.

While Bowler et al.'s [7] themes accurately describe most designs for cyberbullying mitigation, the design sessions highlighted a fourth critical theme: Designing for Support. Three of our designed prototypes–"Positivity Generator", "Happy App", and "Fight Back"–recognize that a bully's actions cannot always be controlled on social media. A bully may face negative consequences for their behaviors, but post-bullying solutions cannot *prevent* bullying from occurring. The seven emergent themes described by Bowler et al. [7] are all bully-centric. They focus on instilling fear or engendering empathy for the victim. Therefore, it is important to have tools to provide cyberbullying victims with emotional support and positivity *after the fact*. Based on the conversations and design sessions with participants, mitigation and support after the cyberbullying occurs is vital part of the mitigation process. This focus on mitigation after the cyberbullying is reflective of the lack of control over the bully and over the social media applications which afford bullying behavior.

We should note that while the "Designing for Empowerment" theme aims to redress the balance of power on social media by asking adults to play a more active role in intervention [7], student-designed automated and peer-focused solutions such as the "Positivity Generator" and "Happy App" may also potentially play an important role in empowering and supporting victims of cyberbullying; this should be outlined specifically in the list of design themes for cyberbullying mitigation.

Designing for Prevention
The prototypes generated from our design sessions varied in terms of who held control in either preventing a bullying scenario or mitigating it after it occurred. In addition to bullies and their victims, bystanders play an important

role in bullying scenarios, often offering implicit or explicit encouragement or discouragement of the bullying [41]. Furthermore, bullying roles are not always dichotomous, with individuals roles based on contextual factors.In our sessions, sample prototypes were generated in which all three actors (bullies, victims, and bystanders/systems) potentially had control over preventing or mitigating the bullying scenario. This particular taxonomy of cyberbullying solutions complements regarding power dynamics in bullying situations and identifying who–the victim, aggressor, and/or bystander–has the most power to mitigate the situation.

Cyberbullying Prevention by the Bully
While literature has discussed how to approach denigration and flaming [10], we have found no academic research discussion issues related to exclusion online. A *New York Times* parenting blog noted, "To be in a photo and to not be tagged is to be rendered socially invisible. Commenting on a party photo, my untagged daughter wrote, 'I was there too!'" [29]. The "Exclusion Prevention" application aims to remedy the potential emotional damage of exclusion-based cyberbullying by presenting the potential bully with a reflective notification. In "Exclusion Prevention", the bully decides whether she wants to continue with publishing content after the system warns the [potential] bully that they may be hurting someone by continuously cropping them out of photos. Ultimately, the decision of publishing the content lies with the potential bully. Dinakar et al. [10] have shared examples of preventive measures when discussing reflective interfaces, which ask users to reflect on their behavior before publishing malicious content online.

From an implementation standpoint, aiming to prevent cyberbullying by focusing on the [potential] bully requires some monitoring since it is attempting to prevent the cyberbullying before it occurs. While privacy advocates may find this monitoring particularly troubling, many parents believe they have the right to access and monitor their children's online activity [4]. There are three notions of a reflective practitioner: "reflection in action", "reflection on action", and "ladders of reflections" [37]. Reflective user interfaces aim to prevent cyberbullying by asking the aggressor to reconsider their actions and reflect on them through showing potential consequences of their actions, flagging their content and notifying them of the potential harm they can cause.

Cyberbullying Prevention by Victim
In the realm of cyberbullying prevention, cyberbullying applications that filter or report content can aid the victim in preventing further occurrences of bullying. A victim can choose to filter bullying content so they never see it (and subsequently experience negative emotional consequences). In "Watch Yo Profanity" and "SMILE", the victim decides if she would like some degree of filtering to be happening on his profile. Depending on the individual, these applications can serve a more proactive approach, whereby the individual chooses filters prior to negative events, or a more reactive approach, in which victims take apply filters to prevent future instances of cyberbullying content.

Cyberbullying Prevention by Automated Systems and Bystanders
From the suite of solutions produced in our design sessions, many attempted to mitigate negative emotional outcome of cyberbullying by sending positivity. This kind of prevention can be initiated by bystanders or automated systems. In the cyberbullying domain, the "Positivity Generator" allows victims to replace malicious content on their profiles with uplifting quotes from their favorite celebrities. This particular solution aims to do more than just filter negative content, but provide support and encouragement to counter the negative cyberbullying content they have experienced. Likewise, for "Hate Page Prevention", a bystander or a third-party automated system has the ultimate control over the cyberbullying content being published.

Limitations of Emergent Solutions
In our design sessions, all participants were encouraged to think outside of the box. In Session V, however, the participants explored the feasibility of their design solutions. For example, there was a discussion that "Exclusion Prevention" may be disruptive to a user's experience on Instagram if a user was prompted that they may hurt someone's feelings every time they tried to crop someone out of a picture. One participant said, "What if my friend doesn't look good in a picture and she would actually prefer for me to crop her out?" These questions prompted them to begin to think of allowing a user to opt-in/opt-out of using their application. The participants questioned the accuracy of filtering algorithms. To counter this limitation, they introduced the notion of letting user decide whether they wants to see the content based on the person who is sending the content. One student said, "I know that [close friend] would never send me anything malicious, so if I was notified that "Watch Yo Profanity" filtered somethings she sent me, I would know to undo it." When reflecting on "The Broiler", the participants decided that "bullying the bullies" was not a ethically sound solution; they ultimately decided that such an application was counter-productive to the cyberbullying mitigation movement.

Technologies and Tools for Implementation
In the design sessions, participants were encouraged to think out of the box regarding what would be technologically possible. That said, many of the solutions they designed are implementable. Below we discuss the technologies required to implement specific cyberbullying solutions.

1. **Application Programming Interfaces.** Many of our teen co-design partners' design solutions, whether standalone applications or browser plugins, required some degree of interaction with social media platforms (e.g., Snapchat, Facebook, Instagram). Just as social media platforms provide Application Programming Interface (API) services that developers and researchers use to expand our understanding of other technology-mediated social interactions (e.g., trending topics), they are important for developers and researchers who aim to design and implement cyberbullying solutions.

2. **Image Recognition Technology.**　Image recognition technology plays a vital role in two of the design solutions: "Hate Page Prevention" and "Exclusion Prevention." The participants described the ease with which social media platforms detect faces when making "tagging" suggestions. They noted that such face detection technologies could be leveraged to prevent "hate pages" which involved the use of screen-shot photos. Screen-shot photos were a reoccurring topic of discussion in our sessions. Many participants expressed that while screen-shotting a Snapchat photo was not necessarily cyberbullying, they would feel threatened if someone screen-shot their photos because it would be both be a violation of privacy and social media etiquette. Their solution to this violation was embedding image recognition technology into social media to detect when a photo was being re-used, and the using primary prevention tactics via reflective interfaces to prompt the aggressor to reconsider re-posting someone else's photo if it was indeed for malicious reasons. Furthermore, the participants spoke about the importance of face recognition technology to prevent exclusion via cropping on applications like Instagram.

3. **Automatic Detection of Malicious and Vulgar Content.** Within the applications that filtered malicious content, participants expressed that there should be filtering of some kind employed on the website.　While most of the filtering solutions included use of "negative buzzwords", participants references many of the same challenges researchers face in the automatic detection realm, especially false-positives in the cases where a negative word is used but the overall content of the post is not negative. To solve this problem, participants said they allow users to play a decision-making role in filtering items.　A victim would be notified if a buzzword was used, but would then get to evaluate whether the post was positive or negative based on the person sending the message. According to the participants in the sessions, the likelihood of being cyberbullied by someone you trust is not high so allowing users to evaluate content based on the person sending it could be a viable solution. While there have been many attempts to accurately identify expletives and negative words in the cyberbullying domain through sophisticated classifiers [11, 22], none has ever attempted to give the victims the power to choose what is acceptable or unacceptable to be posted on their profile. The challenge with automatic detection of cyberbullying is that often expletives can be used affectionately in this particular domain, so it would be counter-productive to filter those cases [22].

4. **Collaborative Filtering: Inferring Favorite Artists from Social Media Data.**　With the "Positivity Generator", participants discussed the possibility of an algorithm inferring a user's favorite artists. On many social media platforms, it is possible to follow/like celebrity pages, and collaborative filtering can be leveraged to infer a user's likes and interests [34].　Those likes could be used to provide support for a cyberbullying victim. Collaborative filtering and similar methodologies in recommendation system research could be leveraged in an application like the "Positivity Generator" to provide targeted support for a victim of cyberbullying.　Previous solutions in the non-academic sphere have sent out targeted song lyrics to a victim of cyberbullying based on the artists the cyberbullying victim enjoys.　The design participants expressed that this kind of targeted support, or the sophisticated understanding of a system of what things a victim enjoys would be particularly valuable in providing support to counter the negativity of cyberbullying.

5. **Networks.**　Prototype solutions like "Exclusion Prevention" and "Hate Page Prevention" can be implemented with higher granularity if they make use of the available information provided by a user's network on social media platforms.　For example, the formation of "cliques" based on tagging behavior on Instagram and the way that they change coupled with face detection technology can predict more accurately if someone is indeed being excluding maliciously.　By analyzing a user's network, it is possible to discover individuals who play an important role in someone's life based on common connections and the clusters within a user's network [17].　Leveraging this existing technology can be helpful in applications that require someone close to the victim, a peer, to provide support for a victim.

SELF-EVALUATION OF CO-DESIGN OF RESEARCHERS AND TEENAGERS

In our last participatory design session, we asked participants to reflect on their experience doing participatory design. From these discussions, three primary themes emerged: 1) appreciation of opportunities to discuss an important issue affecting participants and their peers, 2) surprise over discrepancy of opinion of definitions of cyberbullying and 3) excitement about collaborating with adults as equals. We have expanded on these three themes in the list below.

1. Participants appreciated inclusion in an initiative to prevent cyberbullying since many knew of the aftermath of such incidents, and participants were eager to have the designed solutions implemented.　While many of the participants did not personally experience cyberbullying, they knew peers who had experienced cyberbullying and the resulting negative repercussions. They expressed their excitement of being involved in an initiative that affected people around to them and that they considered a real-world problem.

2. Participants expressed surprise over the diversity of opinions within their peer group over the definition of cyberbullying (e.g., whether cropping an individual from a photo was cyberbullying).　These discussions fostered mutual appreciation between design partners and allowed them to consider nuances of these differences when designing mitigation tools.

3. Participants expressed excitement over the novelty of collaborating with adults for a shared cause and said that the process fostered communication skills with collaborators who were older.　Many of the participants had not had the opportunity work with adults as equal

design partners and this experience was novel for them. They enjoyed the dynamic of working with adults as equal design partners.

CONCLUSION AND FUTURE WORK: MEASURING THE EFFECTIVENESS OF CYBERBULLYING MITIGATION SOLUTIONS

While this study resulted in potential solutions for cyberbullying mitigation, much work lies ahead. We have proposed a number of potential mitigation solutions and the technologies required to implement these solutions. Future research should implement and evaluate these solutions with users through longitudinal studies to evaluate the behavioral impact they have on bullies, victims, and bystanders. In addition, future work should leverage the existing technologies to implement the proposed solutions which are a result of co-design between researchers and adolescents.

Our analysis and categorization of the different preventative types allows us to consider additional research questions, such as which preventative solution is most effective for cyberbullying prevention and how can we accurately measure this effectiveness. Until this point, technological cyberbullying prevention mechanisms have not been evaluated for effectiveness. The framework presented in this paper provides a straightforward way to begin to consider how one would compare different solutions. The ethical challenges of such a study are daunting, but would provide critical insights to preventing cyberbullying.

This paper presents solutions to cyberbullying that were designed by the users most vulnerable to it: adolescents. Specific ways in which this study contribute to HCI are: 1) extending existing cyberbullying intervention design themes (specifically, Bowler et al. [7]) through the analysis of solutions designed with teenagers; and 2) implementing new techniques within the participatory design process to generate cyberbullying solutions from teens' perspective (as compared to implementing teen feedback on designs first created by adults). Finally, the study demonstrates that participatory design using teenagers–who have a vital stake in cyberbullying prevention and mitigation–provides novel insights and solutions.

ACKNOWLEDGEMENTS

REFERENCES

1. The Kanye West Self Confidence Generator. `http://usatoday30.usatoday.com/exp/kanye/kanye.html`. (????). Accessed: 2015-05-16.

2. Louise Arseneault, Lucy Bowes, and Sania Shakoor. 2010. Bullying victimization in youths and mental health problems:Much ado about nothing? *Psychological medicine* 40, 05 (2010), 717–729.

3. Valerie Barker. 2009. Older adolescents' motivations for social network site use: the influence of gender, group identity, and collective self-esteem. *CyberPsychology & Behavior* 12, 2 (2009), 209–213.

4. Susan B Barnes. 2006. A privacy paradox: Social networking in the United States. *First Monday* 11, 9 (2006).

5. Brian Christopher Becker and Enrique G Ortiz. 2008. Evaluation of face recognition techniques for application to facebook. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on.* IEEE, 1–6.

6. Mathilde Bekker, Julie Beusmans, David Keyson, and Peter Lloyd. 2003. KidReporter: a user requirements gathering technique for designing with children. *Interacting with computers* 15, 2 (2003), 187–202.

7. Leanne Bowler, Eleanor Mattern, and Cory Knobel. 2014. Developing design interventions for cyberbullying: A narrative-based participatory approach. (2014).

8. Jake Brown. 2006. *Kanye West in the Studio: Beats Down! Money Up!(2000-2006).* Amber Books Publishing.

9. Francine DeHue, Catherine Bolman, and Trijntje Völlink. 2008. Cyberbullying: Youngsters' experiences and parental perception. *CyberPsychology & Behavior* 11, 2 (2008), 217–223.

10. Karthik Dinakar, Birago Jones, Catherine Havasi, Henry Lieberman, and Rosalind Picard. 2012. Common sense reasoning for detection, prevention, and mitigation of cyberbullying. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2, 3 (2012), 18.

11. Karthik Dinakar, Roi Reichart, and Henry Lieberman. 2011. Modeling the Detection of Textual Cyberbullying.. In *The Social Mobile Web*.

12. Allison Druin. 1996. A place called childhood. *Interactions* 3, 1 (1996), 17–22.

13. Allison Druin. 1999. Cooperative inquiry: developing new technologies for children with children. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 592–599.

14. Allison Druin, Ben Bederson, Angela Boltman, Adrian Miura, Debby Knotts-Callahan, and Mark Platt. 1998. Children as Our Technology Design Partners. (1998).

15. A Druin, B Bederson, JP Hourcade, L Sherman, G Revelle, M Platner, and S Weng. Designing a digital library for young children: An intergenerational partnership. CHI 2001. (????).

16. Dan Fitton, Beth Bell, Janet C Read, Ole Iversen, Linda Little, and Matthew Horton. 2014. Understanding teen UX: building a bridge to the future. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. ACM, 79–82.

17. Eric Gilbert and Karrie Karahalios. 2009. Predicting tie strength with social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 211–220.

18. Mona Leigh Guha, Allison Druin, Gene Chipman, Jerry Alan Fails, Sante Simms, and Allison Farber. 2004. Mixing ideas: a new technique for working with young children as design partners. In *Proceedings of the 2004 conference on Interaction design and children: building a community*. ACM, 35–42.

19. Sameer Hinduja and Justin W Patchin. 2010. Bullying, cyberbullying, and suicide. *Archives of Suicide Research* 14, 3 (2010), 206–221.

20. Sameer Hinduja and Justin W Patchin. 2013. Social influences on cyberbullying behaviors among middle and high school students. *Journal of youth and adolescence* 42, 5 (2013), 711–722.

21. Young Shin Kim and Bennett Leventhal. 2008. Bullying and suicide. A review. *International Journal of Adolescent Medicine and Health* 20, 2 (2008), 133–154.

22. April Kontostathis, Kelly Reynolds, Andy Garron, and Lynne Edwards. 2013. Detecting cyberbullying: query terms and techniques. In *Proceedings of the 5th Annual ACM Web Science Conference*. ACM, 195–204.

23. Nico Lang. 2014. Why teens are leaving Facebook: Its meaningless. (February 2014). `http://www.washingtonpost.com/news/the-intersect/wp/2015/02/21/why-teens-are-leaving-facebook-its-meaningless/`

24. Amanda Lenhart. 2015. Teens, social media, and technology overview 2015. *Pew Internet Project* (2015).

25. Amanda Lenhart, Mary Madden, Aaron Smith, Kristen Purcell, Kathryn Zickhur, and Lee Rainie. 2011. Teens, kindness and cruelty on social network sites. *Pew Internet Project* (2011).

26. Amanda Lenhart, Kristen Purcell, Aaron Smith, and Kathryn Zickuhr. 2010. Social Media & Mobile Internet Use among Teens and Young Adults. Millennials. *Pew Internet & American Life Project* (2010).

27. Sonia Livingstone and Ellen Helsper. 2007. Gradations in digital inclusion: children, young people and the digital divide. *New media & society* 9, 4 (2007), 671–696.

28. Micah L McCreary, Lesley A Slavin, and Eloise J Berry. 1996. Predicting problem behavior and self-esteem among African American adolescents. *Journal of Adolescent Research* 11, 2 (1996), 216–234.

29. Erika Milvy. 2015. Eavesdropping on the Seventh Grade Instagram Show. (2015). `http://parenting.blogs.nytimes.com/2015/02/20/the-unwritten-middle-school-ground-rules-of-the-instagram-show/?_r=0`

30. Michael J Muller, Daniel M Wildman, and Ellen A White. 1993. Equal opportunity PD using PICTIVE. *Commun. ACM* 36, 6 (1993), 64.

31. Daniel Nicolalde and Patricia Brennan. 2014. Involving Young Adults in the Design of Health Interventions. In *CHI 14 Workshop on Understanding Teen UX: Building a Bridge to the Future*.

32. Gwenn Schurgin O'Keeffe, Kathleen Clarke-Pearson, and others. 2011. The impact of social media on children, adolescents, and families. *Pediatrics* 127, 4 (2011), 800–804.

33. Privacy Protection. 2002. Children's Online Privacy Protection Act. (2002).

34. Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. 2001. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*. ACM, 285–295.

35. Mike Scaife and Yvonne Rogers. 1999. Kids as informants: Telling us what we didnt know or confirming what we knew already. *The design of childrens technology* (1999), 27–50.

36. Shari Kessel Schneider, Lydia O'Donnell, Ann Stueve, and Robert WS Coulter. 2012. Cyberbullying, school bullying, and psychological distress: A regional census of high school students. *American Journal of Public Health* 102, 1 (2012), 171–177.

37. Donald A Schön. 1983. *The reflective practitioner: How professionals think in action*. Vol. 5126. Basic books.

38. Brandy Shaul. 2015. Honestly Looks to Combat Cyberbullying on iOS, Android. (2015). `http://www.adweek.com/socialtimes/honestly-looks-to-combat-cyberbullying-on-ios-android/615873`

39. Patti M Valkenburg and Jochen Peter. 2009. Social consequences of the internet for adolescents a decade of research. *Current Directions in Psychological Science* 18, 1 (2009), 1–5.

40. Patti M Valkenburg, Jochen Peter, and Alexander P Schouten. 2006. Friend networking sites and their relationship to adolescents' well-being and social self-esteem. *CyberPsychology & Behavior* 9, 5 (2006), 584–590.

41. Heidi Vandebosch and Katrien Van Cleemput. 2009. Cyberbullying among youngsters: Profiles of bullies and victims. *New media & society* 11, 8 (2009), 1349–1371.

42. Molly White. 2014. The challenges of building a compassionate robot. (November 2014). `http://blog.mollywhite.net/the-challenges-of-building-a-compassionate-robot/`

43. Nancy Willard. 2007. Educators guide to cyberbullying and cyberthreats. *Center for safe and responsible use of the Internet* (2007).

44. Jason Yip, Tamara Clegg, Elizabeth Bonsignore, Helene Gelderblom, Emily Rhodes, and Allison Druin. 2013. Brownies or bags-of-stuff?: domain expertise in cooperative inquiry with children. In *Proceedings of the 12th International Conference on Interaction Design and Children*. ACM, 201–210.