

# Simulating Peer Support for Victims of Cyberbullying

Janneke van der Zwaan      Virginia Dignum      Catholijn Jonker

*Delft University of Technology, The Netherlands*

{j.m.vanderzwaan,m.v.dignum,c.m.jonker}@tudelft.nl

## Abstract

This paper proposes a design for an Embodied Conversational Agent (ECA) that empowers victims of cyberbullying by simulating peer support. The anti-cyberbullying buddy helps a child to cope with negative emotions due to a cyberbullying incident and it shows the child how to deal with future incidents of cyberbullying. The buddy interacts with the victim in three stages: first the child communicates her emotional state, next the buddy gathers information about the situation at hand, then the buddy gives practical advice on how to deal with the situation. During the interaction, the buddy gives emotional support by responding empathically. In addition to giving practical advice and providing emotional support, the buddy tries to establish a relationship with the user by referring to previous events when appropriate.

## 1 Introduction

According to a recent study, US children aged 8 to 18 on average spend 1.5 hours a day using the computer for recreational purposes [14]. Most of these activities take place on the Internet. The study also found 84% of children has access to the Internet at home. So, many children spend a lot of time online. They use the Internet not only as an educational tool, but also for fun, games and to develop and maintain social contacts. One of the risks they run online is to become a victim of cyberbullying. Cyberbullying can be defined as ‘any behavior performed through electronic or digital media by individuals or groups that repeatedly communicates hostile or aggressive messages intended to inflict harm or discomfort on others’ [27]. Cyberbullying takes place via e-mail, instant-messaging programs, Internet chat rooms, multi-player online games, or social websites or blogs. Our work is part of the multi-disciplinary Cyberbullying project that aims at designing social, legal and technological measures to protect and empower children and adolescents against bullying in virtual environments.

It is widely recognized that cyberbullying is a complex issue and a ‘quick fix’ does not exist [6, 19, 24, 26]. Education, both of children and adults, and increasing awareness are suggested to tackle the problem [6, 10, 28]. In particular, there is a growing recognition of the need for children to receive educational support for social and emotional learning, with awareness of its importance for both non-academic outcomes and improved academic performance. The eCircus project<sup>1</sup> developed innovative technology to support social and emotional learning through role-play and affective engagement for Personal and Social Education involving complex social situations. The project focused on enhanced learning through the use of an interactive 3D environment that explores virtual play and improvisational drama with synthetic characters that evoke empathy.

Even though awareness of the problem of cyberbullying is currently increasing at schools, teachers and parents often lack the knowledge and technical skills to truly help victims of cyberbullying [12]. Peer support has proven to be effective against traditional bullying [7]. However, peer supporters need excellent communication skills, such as active listening, adopting a problem-solving approach, being empathic and the ability to build up trust [7]. Not all children have access to (organized) peer support. In addition, peer support may not be available when a child needs it. In this paper, we describe work in progress on an Embodied Conversational Agent (ECA) that empowers victims of cyberbullying by acting as a supportive friend (peer). This anti-cyberbullying buddy makes peer support more accessible, because it is immediately available when a victim of cyberbullying needs help.

---

<sup>1</sup><http://www.macs.hw.ac.uk/EcircusWeb/>

The buddy is a virtual character that ‘lives’ on the computer screen of potential victims of cyberbullying. When a child feels uncomfortable because of a cyberbullying incident, it turns to the buddy for emotional support and practical advice on how to deal with the situation. Ideally, the buddy will be integrated with systems that the children use on a daily basis, so, children are familiar with the character. Implementing the buddy as a local application on the host computer, also serves as a protection of the child’s privacy, as its data will not be shared with the social web applications the child uses. The short term goal of the buddy is to lower the victim’s negative emotions (coping). On the long(er) term, the buddy aims at teaching the victim how to deal with cyberbullying. In order to be successful, the buddy must be able to persuade the user to follow its advice. Interaction between the buddy and the user happens in three stages: first the user indicates she is distressed because of a cyberbullying incident, next the buddy gathers information about the current situation by asking questions and finally the buddy gives practical advice on how to deal with the situation.

The paper is organized as follows. In section 2, we discuss related work on empathy and ECAs. Section 3 introduces the architecture of the anti-bullying buddy. In section 4, we discuss the empowerment strategies followed by the buddy. Section 5 gives an example of an interaction scenario between child and buddy. In section 6, we present our plan for evaluation of the buddy. Finally, in section 7, we present our conclusions and indicate directions for future research.

## 2 Related Work

Empathy is one of the most important communication skills of peer supporters [7]. Empathy is about understanding and responding to the affective state of another person. In recent years, researchers have started to explore the potential of empathic agents for improving interaction and achieving beneficial outcomes.

During the aforementioned eCircus project, Paiva et al. developed FearNot!, an Intelligent Virtual Environment (IVE), where synthetic characters act out bullying scenarios [23]. Paiva et al. demonstrated that the characters were able to elicit empathy from users of the application. While FearNot!’s characters elicit empathy from the user, the anti-cyberbullying agent should be empathic to the user.

Hone showed that empathic agents are able to help users relieving frustration [17]. The SEMAINE project is also about empathic agents that affect the emotional state of the user. The goal of the SEMAINE project is to create listening agents that represent different personalities and try to draw the user into their emotional state by choosing appropriate utterances and non-verbal behavior [16].

Bee et al. proposed an empathic listening agent that tries to keep the user engaged in a social relationship by either mirroring emotions recognized from speech sounds (parallel empathy) or responding to these emotions (reactive empathy) [3]. Adam and Ye designed an emotional dialogue module for an intelligent interactive toy aimed at engaging children by using different emotional strategies, including emotional support by expressing empathy and encouraging the child to take active steps to improve the situation (active coping) [1].

The FitTrack system attempted to change the behavior of its users by trying to establish a relationship [4]. Even though, a lasting change in exercising behavior did not occur, the relational agent was rated more positively than the non-relational agent from the control condition of the experiment.

Related work shows that the user’s emotional state can be affected in a positive way by expressing empathy. Different ways of expressing empathy exist, including linguistic utterances, mirroring the user’s emotion and responding to it. In addition, embodied agents can help users to deal with their negative emotions by applying coping strategies. Empathy also plays an important role in establishing a relationship between the user and an agent. Even though the results of [4] do not show a lasting behavioral change, researchers believe that ECAs have the potential to assist users in changing their behavior, e.g. in mental health promotion [8, 18].

## 3 Buddy Architecture

Interaction between the buddy and the child takes place in three stages:

- Understand the child’s emotional state
- Gathering of information about the bullying situation
- Give advice

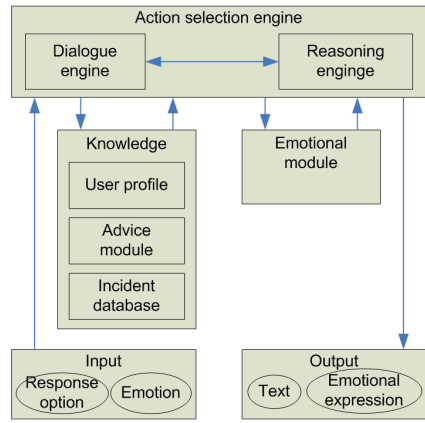


Figure 1: The architecture of the buddy agent (based on generic architecture by Steunebrink et al. [25]).

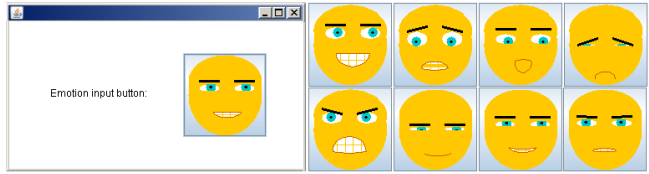


Figure 2: The AffectButton (left side) and several examples of emotional expressions (right side) (image from [5]).

This indicates that the buddy must be able to communicate with the user and understand user input. As a simulated peer, the buddy aims to provide the same skill set that is expected of human peer supporters. Essential skills are being empathic, adopting a problem-solving approach, active listening and the ability to build up trust [7]. In order to be empathic, the buddy will adapt its response to the emotional state of the child and display an appropriate facial expression. The problem-solving approach consists of gathering information about the current situation and giving practical advice. Active listening is implemented by showing appropriate listening behavior (e.g. nodding) and by asking follow up questions. We expect that implementing these three skills will foster trust between the child and the buddy. We intend to explore the extent to which trust is being built up during the evaluation (see section 6).

These requirements are reflected in the proposed architecture of the buddy depicted in Fig. 1. This architecture is a simplification of the generic design of social agents and robots by Steunebrink et al. [25], tailored to the specific characteristics of the buddy. The design of the buddy is modular, so the buddy can easily be extended. We intend to re-use existing modules to jumpstart the development of the buddy. In this section the different modules are discussed.

**3.1 Input and Output Modalities**

The interaction between the buddy and the user starts when the user tells the buddy how she feels. The user describes her emotional state by clicking the AffectButton [5]. The AffectButton is depicted in Fig. 2. By moving the mouse over the button, the user manipulates its facial expression. When the facial expression of the button matches the emotion the user wants to express, she clicks the button. Figure 2 also shows some of the possible affective states.

After the emotional state of the user is communicated, the conversation between the buddy and the user starts. The buddy communicates with the user by sending text messages to a chat-like interface. Ideally, the user would be allowed to respond by typing free text. Since online communication is mostly textual, exchanging text messages would be a natural way for children and adolescents to communicate. However, free text needs to be interpreted before the information content can be used for reasoning and this is not a trivial task. In the first version of the buddy, input from the user is limited to predefined options. After the buddy sends a message, a list of responses is shown on the screen and the user selects one of the options.

RE-phrase [15] offers a way to gather and rank response options in a dialogue. RE-phrase catches patterns in conversations by storing responses in a tree of choices. Possible response options are ordered and filtered based on usage statistics.

In addition to verbal communication, the buddy also communicates non-verbally. The buddy is represented by a virtual character on the screen that displays different facial expressions. The buddy is able to simulate all emotional states that can be expressed with the AffectButton.

The appearance of the buddy is very important. The way the agent looks influences the user's assumptions regarding the agent's similarity (to the user), attractiveness and credibility [2]. Users are more easily persuaded by embodied agents that are considered to be similar to them [13]. Since the buddy plays the role of a peer, its appearance should be similar to children and adolescents in certain age groups. Gender, visual style, and specific verbal and non-verbal behavior are other important aspects that should be taken into account. We are currently evaluating different possible appearances for the buddy. The possibility for users to adapt the appearance of the buddy is also being considered.

## 3.2 Knowledge

The knowledge of the buddy is stored in different databases. The buddy uses the following knowledge bases:

- **User profile:** In order to be able to adapt to the user, the buddy maintains a user profile, consisting of personal information (e.g. name, age, date of birth, etc.), the personality of the user, previous incidents and advice given
- **Incident database:** For each cyberbullying incident, the buddy gathers information about the type of cyberbullying, the method used, the identity of the bully (if known), the advice given, etc. This information is stored in the incident database
- **Advice module:** The advice module contains a mapping of characteristics of bullying incidents to specific pieces of advice

The action selection engine requests information from the knowledge bases and updates them when needed. All these modules are under development.

## 3.3 Emotional Model

The emotional model allows the buddy to reason about the user's emotional state. The emotional model will be based on OCC, a standard model for computerized emotion [22]. The emotional model of the buddy enables it to respond appropriately to emotions of the child, but also includes displaying emotions itself in response to the current situation and its internal state. The emotions experienced by the buddy are empathic to the state of the child. This means joy when the child is happy, sadness when the child is sad, disappointment when a proposed plan failed, etc.

The emotional model includes the components *appraiser* and *intensity* [11]. The appraiser is a process that triggers the creation of emotions based on the state of the environment. The intensity of triggered emotions indicates how strong the emotion is and is assigned with different values depending on the different situations that generated the particular emotion.

The emotional model receives the current affective state of the user through the action selection engine. As discussed in section 3.1, the emotional state of the user is expressed through the AffectButton. The emotional model then guides the synthesis of the buddy's emotions, by providing a mapping from experienced emotion to appropriate facial expressions and linguistic utterances.

## 3.4 Action Selection

In the action selection engine the reasoning process takes place. Based on the input from the user (either some emotional state or one of the predefined response options), the emotional model and the knowledge bases, the buddy selects some action consisting of one or more (textual) utterances and one or more movements of the body (including the face).

## 4 Strategies

In order to achieve its goal of empowerment, the buddy uses the following strategies:

- **Emotional support:** the buddy provides emotional support by responding empathically to the victim
- **Practical advice:** the buddy gives practical advice on how to deal with the current incident. The advice is tailored to the situation (e.g. the type and method of cyberbullying) and the user (based on the user profile)
- **Build relationship:** the buddy attempts to build a relationship with the victim by responding empathically and referring to previous incidents when appropriate

The first strategy of the buddy is to provide emotional support by showing empathy to the user. Different types of empathy exist, e.g. parallel empathy and reactive empathy [9]. Parallel empathy is emotional mimicry, i.e. mirroring the user's emotional state. Reactive empathy is being empathic by displaying other emotions than the ones expressed by the target. The buddy expresses parallel empathy immediately after the user has indicated her affective state with the AffectButton. In the next stage of the interaction, the buddy expresses reactive empathy, by responding emotionally to the responses of the user.

The second strategy of the buddy is to give practical advice. The advice module takes into account characteristics of the incident (e.g. anonymous bullies require a different approach than known bullies), the affective state of the user (including the intensity of the emotions) (e.g. if the user is frightened the buddy should respond in a different manner than if the user is angry), the user profile, and previous incidents (e.g. the first threatening e-mail should be handled differently than the tenth threatening e-mail).

The advice module contains mappings of incident characteristics to specific pieces of advice (cf. section 3.2). The rules in the advice module will be based on children's and adolescent's existing coping strategies that will be gathered from a survey and focus groups. Additional knowledge will be acquired from experts on cyberbullying. Rules are triggered based on the information the buddy gathers about the current situation. The buddy will ask the user questions either until it has sufficient information to give advice or until it concludes it will not be able to give advice. In this case the buddy falls back to the default advice, e.g. 'talk to someone about the incident'.

The buddy's third strategy is to build a relationship. Since bullying refers to a series of events rather than isolated incidents [21], users will interact repeatedly with the anti-cyberbullying buddy. This gives the buddy the opportunity to (try to) establish a relationship. The buddy uses its knowledge about earlier (similar) events and bullies to give the user the sense she is being understood by the buddy. In addition, the buddy requests explicit feedback about the way it handled situations, so it can adapt its approach to the preferences of the user.

## 5 Scenario

In this section, we describe a possible interaction scenario, showing how the buddy (B) will respond when a child (C) indicates she is being cyberbullied (cf. section 3.1). The dialogue of the buddy will be adapted to the child (gender, age group, etc.). Verbal responses are given in *italic* and emotional responses are put between brackets. The child responds to the buddy by selecting one of the response options that are displayed on the screen. These options have been left out of the scenario in order to keep the story brief. In some cases, the child can type in additional details (e.g. the name of the bully).

Situation: C has a profile on a social networking site. She just discovered someone set up a fake profile in her name and added nasty pictures and comments. C is upset and does not know how to respond.

C uses the AffectButton to tell B she is (sad).

B (mirror emotion, sadness): *I'm sorry you're sad. What happened?*

C selects: *fake profile of me on Facebook*

B (negative affect, increased sadness): *That's too bad! Who did that to you?*

C selects: *Someone in my class*

B (no change in expression): *What's this person's name?*

C selects: *Type name* and types: *Nicole*

B (positive affect, decreased sadness): *So, you know who did it! Has she cyberbullied you before?*

C selects: *No*

B (positive affect, moderate happiness): *That's good news. Does Nicole also bully you in real life?*

C selects: *Yes, sometimes*

B (negative affect, sadness): *I'm sorry you are a victim of real life bullying! Let's try to stop Nicole!*

The buddy matches the information it gathered to the advice base. The following advice is triggered: *request removal of the fake profile and talk to your teacher about the bullying*. The dialogue engine transforms this advice into appropriate text messages.

B (positive affect, neutral): *You can ask the administrators of Facebook to remove the fake profile. Does this sound reasonable?*

C selects: *Yes*

B (positive affect, neutral): *All right! Do you know how to request Facebook to remove the fake profile?*

C selects: *Yes*

B (positive affect, neutral): *I also recommend talking to your teacher about the bullying, because you say your classmate also bullies you in real life. Would you try that?*

C selects: *Yes*

B (positive affect, encouragement): *Good luck! I'm sure you can resolve this situation!*

B ends the conversation: *I'll be here, whenever you need me!*

Next time the buddy is started, it pro-actively requests feedback about the advice it gave, in order to adapt the user profile. This also gives the buddy the opportunity to determine whether the child followed the advice and how it worked. If needed, the buddy repeats the advice or provides additional emotional support.

## 6 Evaluation Plan

In order to understand the contribution of the buddy to the empowerment of children in cyberbullying situations, a thorough validation of the design and evaluation of the buddy is necessary. Since we are still in the development stage, we have not yet performed any evaluation. In this section we describe the plan for future work on evaluation.

The first step of evaluating the anti-cyberbullying agent consists of validating the approach for the Cyberbullying project. Since it is unethical to deliberately expose children to bullying, the buddy is evaluated indirectly. It is proposed to create a few different scenarios (see section 5 for an example) and a Wizard of Oz (WOZ) version of the buddy (i.e. a human experimenter selects the actions the buddy performs). The scenarios will be shown to a number of children and/or adolescents. Next, the participants will interact with the WOZ buddy. The children will be asked to pretend to be in situations similar to the scenarios shown before and to give feedback on the interaction. The feedback will be interpreted together with experts on cyberbullying.

When the approach is validated, a prototype system will be implemented. To further investigate how the buddy can convincingly mimic the communication skills needed for peer support, the target audience will be consulted on a regular basis during this phase. Feedback will be gathered on appearance, whether expressed emotions are recognizable and believable, whether the buddy seems trustworthy, whether the buddy is able to build a relationship, etc.

The final version of the anti-cyberbullying buddy will be compared to existing technological interventions against cyberbullying. Children and adolescents should be protected and/or empowered while still being able to enjoy virtual environments. Another important aspect will be the effectiveness of the buddy in different virtual environments compared to other technological measures (e.g. monitoring and filtering). Measures against cyberbullying will be evaluated based on a number of different scenarios. The buddy will also be compared to the social and legal interventions designed in the context of the Cyberbullying project.

## 7 Conclusion

This paper proposed a design for an embodied conversational agent that simulates peer support. The objective of this agent is to empower victims of cyberbullying. The buddy uses strategies to lower the victim's

negative emotions (short term) and teach her how to deal with cyberbullying (long term), including emotional support (empathy), practical advice and building a relationship. The next step in our research consists of creating a WOZ system in order to validate the approach for the Cyberbullying project.

After the approach is validated, we intend to further investigate behavior strategies for the buddy. We are particularly interested in issues concerning trust. The ability to build up trust is essential for human peer supporters [7]. Therefore we expect this to be a vital skill for the buddy as well. Trust and empathy are closely related, but it is still unclear how virtual characters can develop a trusting relationship with the user by being empathic [20].

At the moment, the buddy is being designed as a stand-alone application. Together with providers of online virtual environments that are popular among children and adolescents, we are exploring possibilities to integrate the buddy into these environments. Future research will explore whether an integrated system has added benefits compared to the stand-alone version.

## Acknowledgements

This work is funded by NWO under the Responsible Innovation (RI) program via the project ‘Empowering and Protecting Children and Adolescents Against Cyberbullying’.

## References

- [1] C. Adam and P. Ye. Reasoning about emotions in an engaging interactive toy (extended abstract). In *Proceedings of the 8th International Conference on Autonomous agents and Multiagent Systems (AAMAS 2009)*, pages 31–32, 2009.
- [2] A.L. Baylor. Promoting motivation with virtual agents and avatars: role of visual presence and appearance. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3559–3565, 2009.
- [3] N. Bee, E. André, and T. Vogd. First ideas on the use of affective cues in an empathic computer-based companion. In *Proceedings of the 8th International Conference on Autonomous agents and Multiagent Systems (AAMAS 2009)*, pages 13–17, 2009.
- [4] T.W. Bickmore and R.W. Picard. Establishing and maintaining long-term human-computer relationships. *ACM Trans. Comput.-Hum. Interact.*, 12(2):293–327, 2005.
- [5] J. Broekens and W.P. Brinkman. Affectbutton: Towards a standard for dynamic affective user feedback. In *Affective Computing and Intelligent Interaction (ACII) 2009*, 2009.
- [6] W. Cassidy, M. Jackson, and K.N. Brown. Sticks and stones can break my bones, but how can pixels hurt me?: Students’ experiences with cyber-bullying. *School Psychology International*, 30(4):383–402, 2009.
- [7] H. Cowie, P. Naylor, L. Talamelli, P. Chauhan, and P.K. Smith. Knowledge, use of and attitudes towards peer support: a 2-year follow-up to the prince’s trust survey. *Journal of Adolescence*, 25(5):453–467, 2002.
- [8] D. Coyle, G. Doherty, M. Matthews, and J. Sharry. Computers in talk-based mental health interventions. *Interacting with Computers*, 19(4):545–562, 2007.
- [9] M. Davis. *Empathy: A Social Psychological Approach*. Brown and Benchmark Publishers, 1994.
- [10] F. Dehue, C. Bolman, and T. Vollink. Cyberbullying: youngsters’ experiences and parental perception. *Cyberpsychology & Behavior: the Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, 11(2):217–223, 2008.
- [11] J. Dias and A. Paiva. Feeling and reasoning: A computational model for emotional characters. *Progress in Artificial Intelligence*, pages 127–140, 2005.
- [12] D. Finkelhor, K.J. Mitchell, and J. Wolak. Online victimization: A report on the nation’s youth. Technical report, 2000.

- [13] B.J. Fogg. *Persuasive Technology: Using Computers to Change What We Think and Do*, chapter Computers as Persuasive Social Actors. Acm, New York, NY, USA, 2002.
- [14] Kaiser Family Foundation. Generation m2: Media in the lives of 8- to 18-year-olds. Technical report, 2010.
- [15] J. Geuze, P. Desain, and J. Ringelberg. Re-phrase: chat-by-click: a fundamental new mode of human communication over the internet. In *CHI '08: CHI '08 extended abstracts on Human factors in computing systems*, pages 3345–3350, New York, NY, USA, 2008. Acm.
- [16] D. Heylen. Sensitive empathic agents. In *Proceedings of the 8th International Conference on Autonomous agents and Multiagent Systems (AAMAS 2009)*, pages 9–12, 2009.
- [17] K. Hone. Empathic agents to reduce user frustration: The effects of varying agent characteristics. *Interact. Comput.*, 18(2):227–245, 2006.
- [18] C.L. Lisetti and E. Wagner. Mental health promotion with animated characters: Exploring issues and potential. In *AAAI Spring Symposium – Technical Report*, volume Ss-08-04, pages 72–79, 2008.
- [19] F. Mishna, C. Cook, M. Saini, M.J. Wu, and R. MacFadden. Interventions to prevent and reduce cyber abuse of youth: A systematic review. *Research on Social Work Practice*, 2010.
- [20] R. Morrison. Empathy from avatars: Propositions for improving trust development in pseudo-social relationships with avatars. *European Journal of Social Sciences*, 12(2):298–309, 2009.
- [21] D. Olweus. *The Nature of School Bullying: a Cross-national Perspective*, chapter Sweden, pages 7–27. Routledge, New York, 1999.
- [22] A. Ortony, G.L. Clore, and A. Collins. *The cognitive structure of emotions*. Cambridge Univ. Press, 1988.
- [23] A. Paiva, J. Dias, D. Sobral, R. Aylett, S. Woods, L. Hall, and C. Zoll. Learning by feeling: Evoking empathy with synthetic characters. *Applied Artificial Intelligence: An International Journal*, 19(3):235–266, 2005.
- [24] P.K. Smith, J. Mahdavi, M. Carvalho, S. Fisher, S. Russell, and N. Tippett. Cyberbullying: its nature and impact in secondary school pupils. *Journal of Child Psychology and Psychiatry*, 49(4):376–385, 2008.
- [25] B.R. Steunebrink, N.L. Vergunst, C.P. Mol, F.P.M. Dignum, M.M. Dastani, and J.-J.C. Meyer. A generic architecture for a companion robot. In J. Filipe, J.A. Cetto, and J.-L. Ferrier, editors, *Proc. 5th Int. Conf. on Informatics in Control, Automation and Robotics (ICINCO'08)*, pages 315–321, 2008.
- [26] A.D. Thierer. Five online safety task forces agree: Education, empowerment & self-regulation are the answer. *Progress & Freedom Foundation Progress on Point Paper*, 16(13), 2009.
- [27] R.S. Tokunaga. Following you home from school: A critical review and synthesis of research on cyberbullying victimization. *Computers in Human Behavior*, 26(3):277–287, 2010.
- [28] M.L. Ybarra, K.J. Mitchell, J. Wolak, and D. Finkelhor. Examining characteristics and associated distress related to internet harassment: Findings from the second youth internet safety survey. *Pediatrics*, 118(4):1169–1177, 2006.