Computers in Human Behavior 66 (2017) 345–352

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

"Thinking before posting?" Reducing cyber harassment on social networking sites through a reflective message

Kathleen Van Royen [a, *], Karolien Poels [a], Heidi Vandebosch [a], Philippe Adam [b]

[a] Department of Communication Studies, University of Antwerp, Antwerp, Belgium
[b] Faculty of Arts and Social Sciences, University of New South Wales, Sydney, Australia

ARTICLE INFO

Article history:
Received 22 June 2016
Received in revised form
25 August 2016
Accepted 22 September 2016
Available online 8 October 2016

Keywords:
Cyber harassment
Social networking sites
Adolescents
Reflective interfaces
Experimental study

ABSTRACT

To deal with cyber harassment amongst youth on social networking sites (SNS), interfaces automatically showing a reflective message when harassment is recognized could be integrated. Such message encourages users to reconsider and to eventually self-censor their post. This study examines whether reflective messages reduce harassment among adolescents on SNS.

We conducted an experimental study to test the effect of three different types of reflective messages, as well as a mere time delay, on the change in intention to engage in harassment on SNS. Participants were 321 adolescents from 15 to 16 years old (59.1% was female). We measured their intention to harass prior and after exposure to a reflective message. Moreover, we tested whether certain groups (based on their gender, empathy trait and behavioral inhibition) are more susceptible to certain types of reflective messages.

The results show that, for all three conditions, the intention decreased after being exposed to the message, as well as following only a time delay. Furthermore, in one condition a stronger effect was found for individuals with a higher sensitivity of the behavioral inhibition system.

1. Introduction

Social networking sites (SNS) provide adolescents with multiple options for interaction (Ellison, Steinfeld, & Lampe, 2007; Livingstone, 2008). However, for many, harassment has become part of these interactions (Lenhart et al., 2013). Cyber harassment has no standard definition but is usually described as rude, threatening or offensive content directed at others by friends or strangers and performed via electronic means such as Internet or mobile phones (Lwin, Li, & Ang, 2012; Wolak, Mitchell, & Finkelhor, 2007). Cyber harassment is a broad concept and includes multiple forms of harassment, such as cyberbullying; single insults; hate speech; spamming; cyber stalking; identity theft or online sexual harassment (Lwin et al., 2012).

Studies have found prevalence rates of 21% of adolescents in Canada that had been victim of cyber harassment in the past year (Beran & Li, 2005) and 14% of Swedish boys and 20% of girls (Fridh, Lindström, & Rosvall, 2015). In Singapore, 51% of the adolescents indicated to have been harassed at least once online (Lwin et al., 2012) and among Portuguese adolescents, even 69.9% of the adolescents reported such an event in their lifetime (Pereira, Spitzberg, & Matos, 2016). Reducing cyber harassment is important because various negative health outcomes have been observed among victims, such as depressive symptoms, anxiety, loneliness, somatic complaints or suicidal behavior (Beran, Rinaldi, Bickham, & Rich, 2012; Fridh et al., 2015; Nixon, 2014). In addition to digital literacy and public policy initiatives, one way to protect adolescents from cyber harassments is through the monitoring of these situations by the industry itself (e.g., Internet Service Providers, SNS) (Coyne & Gountsidou, 2013; EC Social Networking Task Force, 2009; Mc Guckin et al., 2013). However, preventing cyber harassment on SNS is a complex matter since adolescents want SNS platforms to provide protective mechanisms to prevent harm, but at the same time prefer to maintain their freedom of expression (Van Royen, Poels, & Vandebosch, 2016). So, while ensuring their

* Corresponding author.
E-mail addresses: kathleen.vanroyen@uantwerpen.be (K. Van Royen), karolien.poels@uantwerpen.be (K. Poels), heidi.vandebosch@uantwerpen.be (H. Vandebosch), philippe.adam@unsw.edu.au (P. Adam).

http://dx.doi.org/10.1016/j.chb.2016.09.040
0747-5632/

right for protection, their right for online participation and freedom of expression should not be infringed (United Nations, 1989).

Reflective interfaces or "automated messages encouraging the user to reflect on online behavior" (Dinakar, Jones, Havasi, Lieberman, & Picard, 2012; Jones, 2012) may be an answer to this challenge since they embrace both the aspects of protection and freedom. Concretely, when an automated detection system (through text and image mining techniques) recognizes that content, about to be posted, is "harassing", it prompts the user to reflect on his behavior, but the user still maintains the autonomy to decide whether to post (Bretschneider, Wöhner, & Peters, 2014; Chen, Zhou, Zhu, & Xu, 2012; Delort, Arunasalam, & Paris, 2011; Sood, Churchill, & Antin, 2012; Yin et al., 2009). In that way, such interfaces can protect adolescents by creating reflection upon harassment but at the same time do not censor content and respect adolescents' freedom. Moreover, creating reflection might be effective since adolescence is a developmental period characterized by low self-control and impulsive actions (Casey, Jones, & Hare, 2008) which have been linked with perpetration of cyber harassment (Holt, Bossler, & May 2011; Vazsonyi, Machackova, Sevcikova, Smahel, & Cerna, 2012).

In regard with *desirability*, some studies indicated that adolescents support the idea of warnings as preventive strategies for cyber harassment (Bowler, Mattern, & Knobel, 2014; Van Royen et al., 2016). Apart from one explorative study on the *effectiveness* of creating a pause before posting harassment (Google Science Fair, 2014), no academic evidence on the effectiveness of different types of reflective messages to reduce cyber harassment amongst adolescents, is available yet. Therefore, the objective of this paper is to further examine whether and how reflective interfaces can reduce harassment on SNS. In the current study, we tested three different types of reflective messages: a message indicating that parents could read the post; a message referring to the disapproval by others and a third message reminding the user of the potential harm for the receiver. We measured the intention to harass prior and after exposure to the reflective message and compared these with a condition with a mere time delay (without a message).

2. Facilitating factors of cyber harassment

2.1. Self-control and impulsivity

Large-scale campaigns that aim to prevent children and adolescents from engaging in online risk behaviors often advertise with slogans such as 'think before you post'(Child Focus, 2010;Common Sense Media, 2015; Kidscape, 2016). Similar messages could also be shown on SNS, thus in the right space, in the heat of the moment of potential harassment (Jones, 2012). An interface providing a reflective message at the exact moment between the typing and posting of potentially harassing content creates a delay and facilitates reflection. This strategy may be effective in countering several facilitating factors of cyber harassment, including low levels of self-control and impulsivity (Holt et al., 2011; Vazsonyi et al., 2012). Self-control refers to the ability to suppress inappropriate emotions, desires and actions (Casey & Caudle, 2013). Gottfredson and Hirschi (1990) state that individuals low in self-control are more likely to engage in offending and deviant behaviors as well as impulsive conduct and are less capable to see the consequences of their action. The association between low self-control and in particular cyber harassment was found for adolescents (Holt et al., 2011; Vazsonyi et al., 2012). Furthermore, impulsivity, the converse of self-control, is an important characteristic of human social behavior. According to dual-process theories, human behavior is a product of reasoned and impulsive behavior (Chaiken & Trope, 1999; Strack & Deutsch, 2004), and eventually all information is processed via the impulsive system (Strack & Deutsch, 2004). Depending on dispositional (e.g. trait of impulsivity) and situational factors (e.g. impulsive states), one of the two systems may be activated more. Some studies suggest SNS users often regret the content they posted (Ferwerda, Schedl, & Tkalčič, 2014; Sleeper et al., 2013; Wang et al., 2011) when they were impulsive and not thinking about consequences while being in a "hot" state of emotions (Wang et al., 2011).

2.2. Online disinhibition

In addition, specific characteristics of computer-mediated communication (CMC) may lower behavioral inhibitions (Kiesler, Siegel, & McGuire, 1984; Suler, 2004) and foster cyber harassment. In particular, online disinhibition, has been linked with several characteristics that can be addressed in the reflective message such as minimization of authority, invisibility, dissociative imagination, and anonymity (Suler, 2004).

First, minimization of authority is a characteristic of the online environment that may foster harassment (Suler, 2004). Especially for adolescents, the internet might appear as a "free zone", compared to the home and school environment. Ritter (2014) demonstrated that perceiving the online environment as more tolerant towards harassment was related to multiple forms of cyber sexual harassment. In the online environment, a harasser is rarely informed that his/her behavior is crossing the line, which can be addressed by a reflective message.

A second important characteristic of CMC is the invisibility of nonverbal cues such as hand movements and gestures, facial expressions and other bodily signs of harm or disapproval (Barak, 2007; Suler, 2004). The lack of nonverbal cues can increase misunderstandings, offenses, and miscommunications (Kiesler et al., 1984) and has been linked with harassment and cyberbullying (a form of peer harassment) (Mehari, Farrell, & Le, 2014). In particular, the lack of nonverbal cues of disapproval online (e.g. a shaking head, a sigh) can encourage people to engage in behaviors they otherwise would not (Suler, 2004). Moreover, since the audience is invisible, users may not be aware of the large amount of people that is (Bernstein, Bakshy, Burke, & Karrer, 2013).

Thirdly, one may feel disassociated, which refers to a feeling that the online world appears like a game. Suler (2004) suggests that someone feeling disinhibited online may be disconnected from the real world and therefore engages more in deviant behavior online. A reflective message might remind the user about the online presence of his/her offline network.

Finally, anonymity is a major factor in determining antinormative behavior (Christopherson, 2007; Postmes & Spears, 1998; Suler, 2004). However, since SNS offer a "nonymous" (i.e. the opposite of anonymous) setting where relationships are often anchored through institutions, residence and mutual friends (Zhao, Grasmuck, & Martin, 2008), anonymity is a less relevant factor here than for example in chat rooms and forums.

3. Previous work related to cues on SNS

Detection techniques are being developed to automatically recognize harassment, through text and image mining techniques (Bretschneider et al., 2014; Chen, , Zhou, , Zhu, , & Xu, 2012; Delort et al., 2011; Sood et al., 2012; Yin et al., 2009). Technology provides therefore opportunities to anticipate and stimulate reflection, which can be integrated in interfaces. Jones (2012) elaborated on the idea of reflective interfaces to assist teenagers on social media, by suggesting options such as: providing notifications, action

delays, interactive help and education, and information about hidden consequences. Effectiveness of messages to stimulate rethinking were tested for the first time in the 2014 Google Science fair, by the teenager Trisha Prabhu (Google Science Fair, 2014; Rethink, 2014). Adolescents were presented with a hurtful message and asked if they would post this on a social media site. If the adolescent clicked "Yes", they were provided an alert message "This message may be hurtful to others. Would you like to pause, review and rethink before posting?" She found this to be effective in convincing the student not to post the hurtful message.

Furthermore, in similar domains several authors have been examining "cues" on SNS to help users consider the content they post. Social cues to decrease the anonymity underlying trolling and negative behavior online were examined, by increasing the degree of identifiability (Cho & Acquisti, 2013; Cho & Kwon, 2015). Ferwerda et al. (2014) demonstrated the effectiveness of persuasive cues, such as a cue that predicts how the user's audience would possibly respond, in order to provide guidance on whether it is safe to post. Nevertheless, there is a lack of research empirically testing the theory of reflection before posting potentially harassing content on SNS. Further, no research evidence is available on the effectiveness of different reflective messages to decrease cyber harassment among adolescents.

4. Reflective messages

4.1. Parents as potential audience

In every communicative act, we have an imagined audience in mind (Goffman, 1959) and in this way users assess the appropriateness of the content they post on SNS (De Wolf, Gao, Berendt, & Pierson, 2015; Marwick & boyd, 2011). However, the imagined and the actual audience are not always aligned, and sometimes one is not aware of who is reading the post (De Wolf et al., 2015).

A message visualizing five random Facebook friends when the user was about to post content was found to make adult users more cautious when disclosing (Wang et al., 2013). For adolescents, a message informing them about parents as a potential audience can be a trigger for withholding negative content. It has been argued that young people will self-censor their posts in face of an imagined audience including parents and significant others, referred to as "nightmare readers" (Marwick & boyd, 2011; Oolo & Siibak, 2013). Moreover, adolescents report on the use of similar strategies, such as minimizing screens, deleting browsing history, and using slang, to evade parental surveillance (Livingstone, 2006; Media Awareness Network, 2004).

H1a. The intention to engage in harassment on SNS will significantly decrease after being exposed to a message indicating that parents might read the post.

In addition, we expect that the behavioral inhibition system (BIS) of an individual may influence the effect of this type of message. BIS is a personality trait in the form of a motivational system that underlies behavior (Gray, 1981; 1970). The primary purpose of the BIS is preventing or stopping behavior that is expected to lead to punishment or the cessation/loss of reward (Gray, 1981). Since parents' disapproval may be perceived as potentially leading to a sanction, those individuals whose personality is characterized by a high level of behavioral inhibition will be more likely to withdraw from engaging in harassment.

H1b. Adolescents with a high behavioral inhibition will show a higher decrease in their intention after being exposed to the message referring to parents than those who have a low behavioral inhibition.

4.2. Disapproval by bystanders

Individuals, part of an online group, tend to comply with the either positive or negative norms (Douglas & McGarty, 2001; Postmes, Spears, & Lea, 2000). When harassment is perceived as acceptable by others, one's intention to engage in it is higher (Ritter, 2014). For cyberbullying, perceiving a negative social pressure from significant others, decreases the intention to perpetrate (Heirman & Walrave, 2012). Furthermore, SNS users censor their content depending on the perceptions they have of their audience's response to the posted content (Ferwerda et al., 2014).Wang et al. (2013) investigated privacy cues on Facebook and found that warning users that others might perceive the post as negative changed the users' posting behavior. Therefore, we will test whether a message indicating others' disapproval of the harassment can decrease the intention to harass.

H2a. The behavioral intention to engage in harassment on SNS will decrease after being exposed to a message indicating others' disapproval.

Moreover, we expect that the disapproval message will have a stronger effect on girls, as they are more likely to engage in behaviors approved by significant others in their social group (Roberts, 1991; Venkatesh & Morris, 2000). In the context of cyberbullying, girls that perceive a higher subjective norm of cyberbullying disapproval are more likely than boys to feel inhibited to bully online (Wong, Cheung, Xiao, & Chan, 2015).

H2b. Girls will show a higher decrease in the intention to engage in harassment when exposed to the message indicating others' disapproval compared to boys.

4.3. Harm for the receiver

Since in CMC the receiver is invisible (Suler, 2004), empathy is reduced, because it is not possible to see the victim's immediate emotional reaction (Mehari et al., 2014), leading to a lack of emotional understanding (Derks, Fischer, & Bos, 2008). A lack of perspective-taking (i.e. the ability of a person to empathize with the situation of another person) and lower levels of both affective (i.e. the sharing of one's emotional state) and cognitive (i.e. the ability to understand another's emotional state) empathy have been linked with cyberbullying perpetration and cyber harassment (Ang & Goh, 2010; Schultze-Krumbholz & Scheithauer, 2009). Furthermore, perspective-taking or imagining how the other feels can produce empathy (Batson, Early, & Salvarani, 1997) and can inhibit aggressive and antisocial behavior (Miller & Eisenberg, 1988).Since text messages can be used to increase empathic motivations (Konrath et al., 2015), this argues for reminding the harasser of the potential emotional impact of his/her message on the receiver.

H3a. The intention to engage in harassment on SNS will decrease after being exposed to a message referring to the impact for the receiver.

Higher levels of the trait empathy has been shown to relate to more prosocial and altruistic behaviors (Roberts & Strayer, 1996). Consequently, we expect that high-empathic individuals will be more prone to change their intention after exposure to a message that requires them to imagine one others' feelings.

H3b. Highly empathic people will show a higher decrease in the intention to engage in harassment on SNS after exposure to the message referring to the impact for the receiver, in comparison with low empathic people.

5. Methodology

5.1. Study participants

Study participants were 321 girls (59.1%) and boys (40.9%) between 15 and 16 years old (M age = 15.60, SD = 0.68). All participants were recruited from the fourth year of 6 secondary schools in Flanders (Dutch speaking part of Belgium) and all of them provided informed consent to participate in the study.

5.2. Study design

This study used a 4 level pre-test/post-test between-subjects design. The experimental design consisted of 4 conditions according to our hypotheses: 1) Message with parents as audience (1st experimental condition), 2) Message with disapproval by others (2nd experimental condition), 3) Message with potential harm for the receiver (3rd experimental condition), 4) only time delay, without message (4th experimental condition). In all conditions, a time delay was introduced by showing the message. This 4thcondition was included to test whether the effects are caused by the messages and not only because adolescents can rethink their harassing message during a time delay. The participants were asked again about their intention, which introduced a mere time delay, without showing any message.

As a scenario, we used a prevalent form of cyber harassment, slut-shaming (labeling someone as a whore for perceived sexual activity) (Tanenbaum, 2015). A fictitious but realistic scenario on an SNS, provoking slut-shaming, was shown to the participants. The scenario represents a situation on Facebook in which a girl, named Merel, posts a public status update revealing that her friend, named Hanna, has stolen her boyfriend. The respondents were also told that Merel was a "friend" of theirs. Based on theoretical assumptions (Lippman & Campbell, 2014; Tanenbaum, 2015), we expected that Hanna could be penalized for violating gendered norms and for using a "provocative" profile picture, by being labeled as a slut.

5.2.1. Pilot test

This scenario was extensively tested in a pilot test among 54 adolescents in the second grade of 2 secondary schools. First, a scenario was shown to the participants in which they had to indicate their intention to comment 'slut' or 'whore' as a reaction on a rating scale from 1 to 7. In this pilot test, for 26% of the participants the intention to harass was higher than 1. The results of the effect of the interfaces shown to the participants were indicating a reduction in intention, however this decrease was not significant because of the small sample size. Second, we inquired their opinion on the scenario with an open question. Some participants commented they would not interfere, as they didn't know the person. Therefore, for the actual experiment we decided to reframe the scenario as if the girl was their friend. Next, the scenario was rated on a 7- point rating scale based on criteria such as perceived credibility (M = 4.31; SD = 1.73) and realism (M = 4.98; SD = 1.85).

Several profile pictures of girls were shown and participants were asked to rate the intention to post 'slut' or 'whore' as a reaction. In addition, several formulations of the reflective messages, based on theoretical assumptions, were rated on a scale to which degree they were perceived as convincing, effective, useless and annoying. Based on these results, we chose the three messages that were tested in this study.

5.3. Research instrument and measures

This study received ethical approval of the University Ethical Committee. Secondary schools were contacted, we obtained consent of the school principals and informed parents about the study. They were given the opportunity to object the participation of their child. The experiment was organized in a computer classroom within the school. Participants logged into a web-based survey. First, they were introduced to the study, without mentioning the exact purpose of the study. Participants were told that the study sought their opinion on common online interactions among adolescents. After giving informed consent, they were exposed to the scenario of Merel posting about Hanna stealing her boyfriend (see supra). Immediately following the scenario, participants were presented with several potential comments on this post, among which harassing comments aimed at Hanna ('whore' and 'slut'), but also supportive comments ('don't mind Hanna'). Participants were asked to indicate how likely it would be that they react with the each of the comments *(Pre intention)*. The intention was measured with a rating scale ranging from 1 to 7. If participants indicated to be "somewhat likely" up to "very likely" to post at least one of the harassing comments (score from 2 to 7), they were randomly exposed to one of the reflective messages or to the condition with a mere time delay. Each participant was exposed to only one message. In the first experimental condition with parents as potential audience, the message was the following: "The comment could be read by your parents and friend's parents. Are you sure to post it?". In the second experimental condition with disapproval by bystanders, this message was shown: "Many others disapprove this comment. Are you sure to post it?". Thirdly, the message in the condition where the user is reminded about the harm for the receiver read: "This comment may be hurtful for the receiver. Are you sure to post it?".

Immediately following the message exposure or the time delay, participants were asked a second time to indicate their intention to engage in this behavior on a similar rating scale from 1 to 7 *(Post intention)*.

Empathy was measured with the validated AMES scale (Vossen, Piotrowski, & Valkenburg, 2015), using 8 items for both cognitive and affective empathy. The AMES scale has been shown to be a reliable and valid measure of empathy and sympathy in Dutch-speaking teens (Vossen et al., 2015). The cronbach's alpha in the present study was reliable (α = 0.795, M = 25.64, SD = 5.16).

Behavioral inhibition system was assessed with the BIS scale of Carver and White (1994). The Dutch version of the BIS/BAS scales has adequate reliability and construct validity (Franken, Muris, & Rassin, 2005) and previous research has found this to be a reliable scale to use with Dutch-speaking adolescents (De Cock et al., 2016; Matton, Goossens, Braet, & Vervaet, 2013; Roose, Bijttebier, Claes, & Lilienfeld, 2011). The scale proved to be reliable in this study (α = 0.729, M = 19.73, SD = 4.16).

In addition, participants were asked for their age, gender and education level.

5.4. Data analysis

Preliminary analyses confirmed that there were no significant differences between the participants in the experimental conditions in terms of gender, age, education type, pre-intention to harass, and previous experience with slut-shaming (either as perpetrator or as a victim) ($0.85 \leq \chi^2 \leq 20.53$, all $p > 0.05$).

In order to test for main effects, the mean score for the pre-intention versus the post-intention in each condition were compared. Repeated measures ANOVA models were used to test the effects with time as a within-subjects factor (pre versus post). For all statistical tests, significance was tested using a 95% confidence interval.

Differences in effect were assessed by calculating effect sizes for each message effect, to examine whether the effect differs for the

conditions. We calculated the effect sizes by using the omega squared (ω^2) as this is less biased than for instance the r^2 (Field & Hole, 2003).

In order to analyze for the moderating effects, we first created subgroups such as high-versus low-empathic people and low BIS and high BIS. We used a median split to transform the continuous variable into a categorical one with two subgroups. There were no significant differences observed between the subgroups in regard with their pre-intention, both for the variable BIS ($t = 1.91$; $p > 0.05$), as for empathy ($t = 0.693$; $p > 0.05$). We entered the new categorical moderating variables as a between-subjects variable in our repeated measures ANOVA model to look for an interaction effect.

6. Results

In total, 48.6% of the participants ($N = 156$) demonstrated an intention (with a score of 2 or higher) to post the comment 'slut' and consequently were randomly exposed either to one of the three reflective messages or to the condition with a mere time delay.

Results indicate that in all conditions the intention to engage in cyber harassment decreased (See Fig. 1 and Table 1). The first message, indicating parents as an audience, significantly reduced the mean intention to harass ($F(1,43) = 33.32$; $p < 0.001$; $\omega^2 = 0.65$) (See Table 2). The second message, referring to the disapproval by others, significantly decreased one's intention to harass ($F(1,41) = 6.23$; $p = 0.017$; $\omega^2 = 0.33$). The third message, reminding the user about the potential harmful impact, also led to a significant reduction in intention ($F(1,38) = 17.28$; $p < 0.001$; $\omega^2 = 0.54$). Finally, the adolescents who were exposed to a mere time delay, demonstrated a significant decrease in intention to harass as well ($F(1,35) = 9.55$; $p = 0.004$; $\omega^2 = 0.44$). All effect sizes are large effects since an omega squared value above 0.14 is considered a large effect (Field & Hole, 2003). The highest decrease in the mean intention was found for the message using parents as a potential audience (from $M = 3.68$ to $M = 2.57$). The lowest decrease in mean intention was found for the condition in which the message indicates the disapproval by bystanders (from $M = 3.33$ to $M = 2.71$). See further in Table 1 for the difference scores of the four experimental conditions.

When testing these effects for only the group of participants that indicated an intentional score higher than 3, as well as from 4, to post such comment, all effects remained significant.

In regard with moderating effects, we found that adolescents with a low behavioral inhibition, demonstrated a significantly ($p = 0.019$) lower decrease in their intention after exposure ($M = 0.79$) to the message (parents as audience), than those with high behavioral inhibition levels ($M = 1.73$). Thereby, our hypothesis (H1b) was confirmed. There was no significant difference in the decreased intention between girls and boys, after exposure to the message with other's approval ($p = 1.0$). For the variable of empathy, too, no significant difference was found in the changed intention after exposure to the message indicating the receivers' harm ($p = 0.97$).

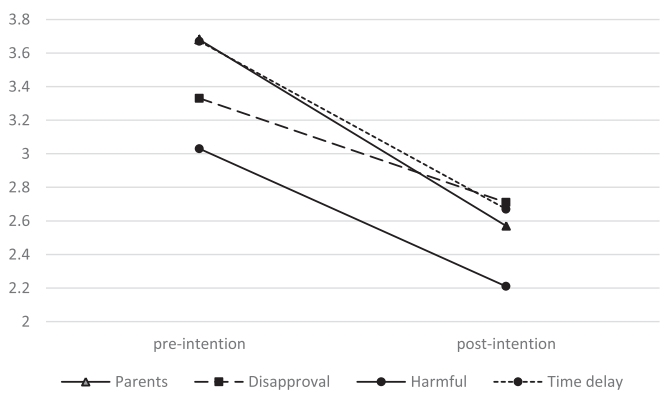

Fig. 1. Means for pre- and post-intention scores.

7. Discussion

Since adolescents are increasingly involved with SNS and cyber harassment is prevalent on these sites among adolescents (Lenhart et al., 2013), there is an urgent need to find new ways to create more self-control for harassers in order to curb their harassing comments. We proposed a technology and in particular interfaces showing reflective messages as a strategy to address this problem. This study tested the effect of three different types of reflective messages using an experimental design. Findings indicate that after exposure to a message, for all conditions, a significant reduction in the intention to harass was observed. This study shows that reflective technology is potentially a substantial way forward to achieve a reduction in harassment on SNS. Important to note is that the same effect of decreasing harassment was found for a mere time delay, suggesting the relevance of curbing the impulsivity and increasing self-control. In a sense, online risky behavior and cyber harassment can be seen as a consequence of low self-control and a need for immediate gratification for the desire of harassing someone (Bossler & Holt, 2010). Delaying this gratification with a time out might be therefore effective in decreasing impulsive actions. Similarly, in a previous study, a timer cue, designed to encourage users to stop and think, so as to avoid online disclosure that would threaten their privacy, was found helpful in reconsidering their posts (Wang et al., 2013). However, a mere time delay may annoy the user when they have to wait until time expires and comes at the cost of delaying every post (Wang et al., 2013). Therefore these messages may be a good alternative to vary and prompt users to reflect.

The content of the reflective message can be tailored based on people's characteristics. In this regard, we can look at which effects are strongest for which groups. The highest effect was found for the message that notified the adolescent about parents as a potential audience of the harassing publicly visible comment. This may be related to their fear of parents as readers (Marwick & boyd, 2011; Oolo & Siibak, 2013). Adolescents are worried about their online privacy for mostly their parents (boyd, 2014). However, in terms of desirability, it may be the least favorite message, as it may be perceived as too patronizing. This might remind the adolescents of being monitored which interferes with the child's right on autonomy and freedom (United Nations, 1989).

For adolescents with high levels of behavioral inhibition, the effect of this message was stronger than for those low in behavioral inhibition. Since these individuals are more sensitive for cues that signal punishment, they may fear sanctions from their parents when they would read the post. Studies confirm that for cyberbullying, often adolescents prefer to hide this event for their parents because of fear for sanctions or more discipline (Hoff & Mitchell, 2009; Mishna, Saini, & Solomon, 2009). Age may be another important factor in determining the effectiveness of this message, as the youngest adolescents might be more susceptible for this. More different age groups should be involved in future research.

The message indicating disapproval by others had the least strong effect. This weaker effect may be due to the lack of specification of the "others" that would disapprove the behavior in the tested message. Injunctive norms are dependent on significant

Table 1
Descriptive statistics: Means and standard deviations by experimental condition.

Experimental condition	*N*	Pre-intention	Post-intention	Difference score
		M (SD)	*M (SD)*	
Parents as potential audience	44	3.68 (1.67)	2.57 (1.87)	1.11
Disapproval by bystanders	42	3.33 (1.62)	2.71 (2.11)	0.62
Receiver's harm	39	3.03 (1.71)	2.21 (1.58)	0.82
Mere time delay	36	3.67 (1.97)	2.67 (2.01)	1.00

Table 2
Repeated measures ANOVA.

Experimental condition	*df*	*MSE* (mean standard error)	*F*	ω^2
Parents as potential audience	1	0.82	33.32^{***}	0.65
Disapproval by others	1	1.29	6.23^{*}	0.33
Receiver's harm	1	0.76	17.28^{***}	0.54
Mere time delay	1	1.89	9.55^{**}	0.44

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

others and it was found that the effect differs depending on behavior-specific important others (i.e. those having greater influence on specific behaviors) (Trafimow & Fishbein, 1994). Furthermore, reflective messages could communicate simultaneously descriptive norms (i.e. what others do). Dolan et al. (2012) argue that if the norm is desirable, informing that other people display conformity with that norm can be effective. Moreover, Fishbein argues in his integrative model of behavioral prediction (Fishbein & Yzer, 2003) that both injunctive and descriptive norms execute normative influence in attitude - behavior relations and should be modelled together.

In future research, reflective messages can be applied and tested for other situations such as in regard with online risk behavior (e.g., posting provocative pictures). In this context, other messages may be relevant, such as reminding the user about the number of people that would be able to see the post, as often social media users underestimate how many friends they reach by a factor of four (Bernstein et al., 2013). This was also proposed by Jones (2012), in regard with cyberbullying. Nevertheless, unintended (and undesirable) effects of reflective messages may occur. In the case of cyberbullying, for instance, perpetrators may be even more motivated to post their harmful content when they know that many people will see their post.

Limitations of this study include the use of an experimental design. Effects could have been better tested in a real setting, however practical constraints did not allow us to test it as such. The established effects may have been socially desirable answers. Nevertheless, measures were taken to avoid this such as not explicitly informing the participants about the exact purpose of the study and using a between-subjects design.

The messages used in this study may have required too much cognitive effort from the participants. More subtle messages that can be easily processed should be tested, using intuitive judgment processes (Greene & Haidt, 2002).

Finally, the scenarios were non-interactive, since respondents were not able to actually post something but were given possible response options.

To conclude, the applicability of these interfaces needs further elaboration. When a user is repeatedly exposed to such messages, this may cause annoyance, and might be perceived intrusive and privacy invading. Other options can be explored, such as for new users, providing them with interfaces for a short period of time, sufficient in duration though to train the users in controlling impulsive posts. Studies can be conducted to understand users' experiences by measuring emotions and reactions when adolescents are targeted by these messages, in order to optimize the user experience when exposed to these messages.

Acknowledgements

This work was supported by the Agency for Innovation by Science and Technology (IWT) [grant number 120007].

References

Ang, R. P., & Goh, D. H. (2010). Cyberbullying among adolescents: The role of affective and cognitive empathy, and gender. *Child Psychiatry & Human Development, 41*(4), 387–397.
Barak, A. (2007). Phantom emotions: Psychological determinants of emotional experiences on the internet. In *Oxford handbook of internet psychology* (pp. 303–329). Oxford: Oxford University Press.
Batson, C. D., Early, S., & Salvarani, G. (1997). Perspective taking: Imagining how another feels versus imaging how you would feel. *Personality and Social Psychology Bulletin, 23*(7), 751–758. http://dx.doi.org/10.1177/0146167297237008.
Beran, T., & Li, Q. (2005). Cyber-harassment: A study of a new method for an old behavior. *Journal of Educational Computing Research, 32*(3), 265–277.
Beran, T. N., Rinaldi, C., Bickham, D. S., & Rich, M. (2012). Evidence for the need to support adolescents dealing with harassment and cyber-harassment: Prevalence, progression, and impact. *School Psychology International, 33*(5), 562–576. http://dx.doi.org/10.1177/0143034312446976.
Bernstein, M. S., Bakshy, E., Burke, M., & Karrer, B. (2013). Quantifying the invisible audience in social networks. In *Proceedings of the SIGCHI conference on human factors in computing systems (pp. 21–30). New York, NY, USA*. ACM. http://dx.doi.org/10.1145/2470654.2470658.
Bossler, A. M., & Holt, T. J. (2010). The effect of self-control on victimization in the cyberworld. *Journal of Criminal Justice, 38*(3), 227–236. http://dx.doi.org/10.1016/j.jcrimjus.2010.03.001.
Bowler, L., Mattern, E., & Knobel, C. (2014). Developing design interventions for cyberbullying: A narrative-based participatory approach. In *iConference 2014 Proceedings* (pp. 153–162). http://dx.doi.org/10.9776/14059.
boyd, D. (2014). *It's complicated: The social lives of networked teens*. New Haven, CT: Yale University Press.
Bretschneider, U., Wöhner, T., & Peters, R. (2014). Detecting online harassment in social networks. In *ICIS 2014 proceedings*. Retrieved from http://aisel.aisnet.org/icis2014/proceedings/ConferenceTheme/2.
Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology, 67*(2), 319–333. http://dx.doi.org/10.1037/0022-3514.67.2.319.
Casey, B., & Caudle, K. (2013). The teenage brain: Self control. *Current Directions in Psychological Science, 22*(2), 82–87. http://dx.doi.org/10.1177/0963721413480170.
Casey, B. J., Jones, R. M., & Hare, T. A. (2008). The adolescent brain. *Annals of the New York Academy of Sciences, 1124*(1), 111–126. http://dx.doi.org/10.1196/annals.1440.010.
Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. Guilford Press.
Chen, Y., Zhou, Y., Zhu, S., & Xu, H. (2012). Detecting offensive language in social media to protect adolescent online safety. In: *Privacy, Security, Risk and Trust (PASSAT), 2012 international conference on and 2012 international confernece on social computing* (*SocialCom*) (pp. 71–80).http://doi.org/10.1109/SocialCom-PASSAT.2012.55.
Child Focus, (2010). Retrieved from http://www.childfocus.be/sites/default/files/manual_uploads/think_before_you_post_nl.pdf.
Cho, D., & Acquisti, A. (2013). The more social cues, the less trolling? An empirical study of online commenting behavior. In *Proceedings of the twelfth workshop on the economics of information security*.
Cho, D., & Kwon, K. H. (2015). The impacts of identity verification and disclosure of social cues on flaming in online user comments. Part A *Computers in Human Behavior, 51*, 363–372. http://dx.doi.org/10.1016/j.chb.2015.04.046.
Christopherson, K. M. (2007). The positive and negative implications of anonymity in internet social interactions: "On the internet, nobody knows You're a dog."

Computers in Human Behavior, 23(6), 3038–3056. http://doi.org/10.1016/j.chb.2006.09.001.
Common Sense Media. (2015). Retrieved, from https://www.commonsensemedia.org/sites/default/files/uploads/classroom_curriculum/edu-vd_6-8_oversharingthinkbeforeyoupost.pdf.
Coyne, I., & Gountsidou, V. (2013). The role of industry in reducing cyberbullying. In P. K. Smith, & G. Steffgen (Eds.), *Cyberbullying: Research on coping with negative and enhancing positive uses of new technologies* (pp. 83–98). London: Psychology Press.
De Cock, N., Van Lippevelde, W., Goossens, L., De Clercq, B., Vangeel, J., Lachat, C., … Van Camp, J. (2016). Sensitivity to reward and adolescents' unhealthy snacking and drinking behavior: The role of hedonic eating styles and availability. *International Journal of Behavioral Nutrition and Physical Activity, 13*(17). http://dx.doi.org/10.1186/s12966-016-0341-6.
De Wolf, R., Gao, B., Berendt, B., & Pierson, J. (2015). The promise of audience transparency. Exploring users' perceptions and behaviors towards visualizations of networked audiences on facebook. *Telematics and Informatics, 32*(4), 890–908.
Delort, J.-Y., Arunasalam, B., & Paris, C. (2011). Automatic moderation of online discussion sites. *International Journal of Electronic Commerce, 15*(3), 9–30. http://dx.doi.org/10.2753/JEC1086-4415150302.
Derks, D., Fischer, A. H., & Bos, A. E. R. (2008). The role of emotion in computer-mediated communication: A review. *Instructional Support for Enhancing Students' Information Problem Solving Ability, 24*(3), 766–785. http://dx.doi.org/10.1016/j.chb.2007.04.004.
Dinakar, K., Jones, B., Havasi, C., Lieberman, H., & Picard, R. (2012). Commonsense reasoning for detection, prevention and mitigation of cyberbullying. *ACM Transactions on Interactive Intelligent Systems, 2*(3). http://dx.doi.org/10.1145/2362394.2362400.
Dolan, P., Hallsworth, M., Halpern, D., King, D., Metcalfe, R., & Vlaev, I. (2012). Influencing behaviour: The mindspace way. *Journal of Economic Psychology, 33*(1), 264–277. http://dx.doi.org/10.1016/j.joep.2011.10.009.
Douglas, K. M., & McGarty, C. (2001). Identifiability and self-presentation: Computer-mediated communication and intergroup interaction. *British Journal of Social Psychology, 40*(3), 399–416. http://dx.doi.org/10.1348/014466601164894.
EC Social Networking Task Force. (2009). *Safer social networking principles for the EU*. Luxembourg: European Commission. Retrieved from https://ec.europa.eu/digital-agenda/sites/digital-agenda/files/sn_principles.pdf.
Ellison, N. B., Steinfeld, C., & Lampe, C. (2007). The benefits of facebook "friends": Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication, 12*(4). http://dx.doi.org/10.1111/j.1083-6101.2007.00367.x.
Ferwerda, B., Schedl, M., & Tkalčič, M. (2014). To post or not to post: The effects of persuasive cues and group targeting mechanisms on posting behavior. In *Proceedings of the 6th ASE international conference on social computing, Stanford, USA* (pp. 1–12). May 27-31.
Field, A., & Hole, G. (2003). *How to design and report experiments*. London: SAGE Publications Ltd. Retrieved from https://uk.sagepub.com/en-gb/eur/how-to-design-and-report-experiments/book219351.
Fishbein, M., & Yzer, M. C. (2003). Using theory to design effective health behavior interventions. *Communication Theory, 13*(2), 164–183. http://dx.doi.org/10.1111/j.1468-2885.2003.tb00287.x.
Franken, I. H. A., Muris, P., & Rassin, E. (2005). Psychometric properties of the Dutch BIS/BAS scales. *Journal of Psychopathology and Behavioral Assessment, 27*(1), 25–30. http://dx.doi.org/10.1007/s10862-005-3262-2.
Fridh, M., Lindström, M., & Rosvall, M. (2015). Subjective health complaints in adolescent victims of cyber harassment: Moderation through support from parents/friends - A swedish population-based study. *BMC Public Health, 15*(1), 949. http://dx.doi.org/10.1186/s12889-015-2239-7.
Goffman, E. (1959). *The presentation of self in everyday life* (1 edition). Garden City, N.Y.: Anchor.
Google Science Fair. (2014). *Rethink: An effective way to prevent cyberbullying*. Retrieved from https://docs.google.com/presentation/d/1bpP5s6GTrCnd9tmTDcD68fqZO-rOYDYxJZCgByBZ6p4/preview?usp=embed_facebook.
Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime* (Vol. xvi). Stanford University Press: Stanford University Press.
Gray, J. A. (1970). The psychophysiological basis of introversion-extraversion. *Behaviour Research and Therapy, 8*(3), 249–266. http://dx.doi.org/10.1016/0005-7967(70)90069-0.
Gray, J. A. (1981). A critique of Eysenck's theory of personality. In P. H. J. Eysenck (Ed.), *A model for personality* (pp. 246–276). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-67783-0_8.
Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences, 6*(12), 517–523. http://dx.doi.org/10.1016/S1364-6613(02)02011-9.
Heirman, W., & Walrave, M. (2012). Predicting adolescent perpetration in cyberbullying: An application of the theory of planned behavior. *Psicothema, 24*(4), 614–620.
Hoff, D. L., & Mitchell, S. N. (2009). Cyberbullying: Causes, effects, and remedies. *Journal of Educational Administration, 47*(5), 652–665. http://dx.doi.org/10.1108/09578230910981107.
Holt, T. J., Bossler, A. M., & May, D. C. (2011). Low self-control, deviant peer associations, and juvenile cyberdeviance. *American Journal of Criminal Justice, 37*(3), 378–395. http://dx.doi.org/10.1007/s12103-011-9117-3.
Jones, B. K. (2012). *Reflective Interfaces: Assisting teens with stressful situations online*. Unpublished PhD dissertation. Massachusetts Institute of Technology. Retrieved from http://dspace.mit.edu/handle/1721.1/76519.
Kidscape. (2016). Retrieved May 12, 2016, from https://www.kidscape.org.uk/advice/advice-for-young-people/dealing-with-cyber-bullying/think-before-you-post.
Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist, 39*(10), 1123–1134. http://dx.doi.org/10.1037/0003-066X.39.10.1123.
Konrath, S., Falk, E., Fuhrel-Forbis, A., Liu, M., Swain, J., Tolman, R., … Walton, M. (2015). Can text messages increase empathy and prosocial Behavior? The development and initial validation of text to connect. *PLoS One, 10*(9), e0137585. http://dx.doi.org/10.1371/journal.pone.0137585.
Lenhart, A., Madden, M., Smith, A., Purcell, A., Zickuhr, K., & Rainee, L. (2013). *Teens, kindness and cruelty on social network sites*. Pew Research Center's Internet & American Life Project.
Lippman, J. R., & Campbell, S. W. (2014). Damned if you do, damned if you don't…if you're a girl: Relational and normative contexts of adolescent sexting in the United States. *Journal of Children and Media, 8*(4), 371–386. http://dx.doi.org/10.1080/17482798.2014.923009.
Livingstone, S. (2006). Children's privacy online: Experimenting with boundaries within and beyond the family. In R. Kraut, M. Brynin, & S. Kiesler (Eds.), *Computers, phones, and the Internet: Domesticating information technology* (pp. 145–167). New York, USA: Oxford University Press. Retrieved from http://www.oup.com/us.
Livingstone, S. (2008). Taking risky opportunities in youthful content creation: Teenagers' use of social networking sites for intimacy, privacy and self-expression. *New Media & Society, 10*(3), 459–477.
Lwin, M. O., Li, B., & Ang, R. P. (2012). Stop bugging me: An examination of adolescents' protection behavior against online harassment. *Journal of Adolescence, 35*(1), 31–41.
Marwick, A. E., & boyd, D. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society, 13*(1), 114–133. http://dx.doi.org/10.1177/1461444810365313.
Matton, A., Goossens, L., Braet, C., & Vervaet, M. (2013). Punishment and Reward Sensitivity: Are Naturally Occurring Clusters in These Traits Related to Eating and Weight Problems in Adolescents? *European Eating Disorders Review, 21*(3), 184–194. http://dx.doi.org/10.1002/erv.2226.
Mc Guckin, C., Perren, S., Corcoran, L., Cowie, H., Dehue, F., Ševčíková, A., … Völlink, T. (2013). *Coping with cyberbullying: How can we prevent cyberbullying and how victims can cope with it*. Psychology Press. Retrieved from http://www.muni.cz/research/publications/1136495.
Media Awareness Network. (2004). *Young Canadians in a wired World: Phase II. Focus Groups*. Ottawa: Media Awareness Network.
Mehari, K. R., Farrell, A. D., & Le, A.-T. H. (2014). Cyberbullying among adolescents: Measures in search of a construct. *Psychology of Violence*. http://dx.doi.org/10.1037/a0037521.
Miller, P. A., & Eisenberg, N. (1988). The relation of empathy to aggressive and externalizing/antisocial behavior. *Psychological Bulletin, 103*(3), 324–344. http://dx.doi.org/10.1037/0033-2909.103.3.324.
Mishna, F., Saini, M., & Solomon, S. (2009). Ongoing and online: Children and youth's perceptions of cyber bullying. *Children and Youth Services Review, 31*(12), 1222–1228.
Nixon, C. L. (2014). Current perspectives: The impact of cyberbullying on adolescent health. *Adolescent Health, Medicine and Therapeutics, 5*, 143–158. http://dx.doi.org/10.2147/AHMT.S36456.
Oolo, E., & Siibak, A. (2013). Performing for one's imagined audience: Social steganography and other privacy strategies of Estonian teens on networked publics. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 7*(1). http://dx.doi.org/10.5817/CP2013-1-7.
Pereira, F., Spitzberg, B. H., & Matos, M. (2016). Cyber-harassment victimization in Portugal: Prevalence, fear and help-seeking among adolescents. *Computers in Human Behavior, 62*, 136–146. http://dx.doi.org/10.1016/j.chb.2016.03.039.
Postmes, T., & Spears, R. (1998). Deindividuation and antinormative behavior: A meta-analysis. *Psychological Bulletin, 123*(3), 238–259. http://dx.doi.org/10.1037/0033-2909.123.3.238.
Postmes, T., Spears, R., & Lea, M. (2000). The formation of group norms in computer-mediated communication. *Human Communication Research, 26*(3), 341–371. http://dx.doi.org/10.1111/j.1468-2958.2000.tb00761.x.
Rethink. (2014). Retrieved from https://www.googlesciencefair.com/projects/en/2014/f4b320cc1cedf92035dab51903bdd95a846ae7de6869ac40c909525efe7c79db.
Ritter, B. A. (2014). Deviant behavior in computer-mediated Communication: Development and validation of a measure of cybersexual harassment. *Journal of Computer-Mediated Communication, 19*(2), 197–214. http://dx.doi.org/10.1111/jcc4.12039.
Roberts, T. A. (1991). Gender and the influence of evaluations on self-assessments in achievement settings. *Psychological Bulletin, 109*(2), 297–308.
Roberts, W., & Strayer, J. (1996). Empathy, emotional expressiveness, and prosocial behavior. *Child Development, 67*(2), 449–470. http://dx.doi.org/10.1111/j.1467-8624.1996.tb01745.x.
Roose, A., Bijttebier, P., Claes, L., & Lilienfeld, S. O. (2011). Psychopathic traits in adolescence: Associations with the revised reinforcement sensitivity theory systems. *Personality and Individual Differences, 50*(2), 201–205. http://dx.doi.org/10.1016/j.paid.2010.09.028.
Schultze-Krumbholz, A., & Scheithauer, H. (2009). Social-behavioral correlates of

cyberbullying in a German student sample. *Zeitschrift Für Psychologie/Journal of Psychology, 217*(4), 224–226.

Sleeper, M., Balebako, R., Das, R., McConahy, A. L., Wiese, J., & Cranor, L. F. (2013). The post that wasn't: Exploring self-censorship on facebook. In *CSCW '13 proceedings of the 2013 conference on computer supported cooperative work* (pp. 793–802). http://dx.doi.org/10.1145/2441776.2441865.

Sood, S. O., Churchill, E. F., & Antin, J. (2012). Automatic identification of personal insults on social news sites. *Journal of the American Society for Information Science and Technology, 63*(2), 270–285. http://dx.doi.org/10.1002/asi.21690.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc, 8*(3), 220–247. http://dx.doi.org/10.1207/s15327957pspr0803_1.

Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior, 7*(3), 321–326.

Tanenbaum, L. (2015). *I Am not a slut*. New York: Harper Collins Publishers.

Trafimow, D., & Fishbein, M. (1994). The moderating effect of behavior type on the subjective norm-behavior relationship. *The Journal of Social Psychology, 134*(6), 755–763. http://dx.doi.org/10.1080/00224545.1994.9923010.

United Nations. (1989). *Convention on the rights of the child*. Retrieved from http://www.ohchr.org/Documents/ProfessionalInterest/crc.pdf.

Van Royen, K., Poels, K., & Vandebosch, H. (2016). Harmonizing freedom and Protection: Adolescents' voices on automatic monitoring of social networking sites. *Children and Youth Services Review, 64*, 35–41. http://dx.doi.org/10.1016/j.childyouth.2016.02.024.

Vazsonyi, A. T., Machackova, H., Sevcikova, A., Smahel, D., & Cerna, A. (2012). Cyberbullying in context: Direct and indirect effects by low self-control across 25 European countries. *European Journal of Developmental Psychology, 9*(2), 210–227.

Venkatesh, V., & Morris, M. G. (2000). Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. *MIS Quarterly, 24*(1), 115–139. http://dx.doi.org/10.2307/3250981.

Vossen, H. G. M., Piotrowski, J. T., & Valkenburg, P. M. (2015). Development of the adolescent measure of empathy and sympathy (AMES). *Personality and Individual Differences, 74*, 66–71. http://dx.doi.org/10.1016/j.paid.2014.09.040.

Wang, Y., Leon, P. G., Scott, K., Chen, X., Acquisti, A., & Cranor, L. F. (2013). Privacy nudges for social media: An exploratory facebook study. In *WWW 2013 companion* (pp. 763–770). Rio de Janeiro (Brasil).

Wang, Y., Norcie, G., Komanduri, S., Acquisti, A., Leon, P., & Cranor, L. (2011). I regretted the minute I pressed share": A qualitative study of regrets on Facebook. In *Proceedings of the seventh symposium on usable privacy and security (SOUPS 2011)*. New York, NY, USA: ACM.

Wolak, J., Mitchell, K. J., & Finkelhor, D. (2007). Does online harassment constitute bullying? An exploration of online harassment by known peers and online-only contacts. *Journal of Adolescent Health, 41*(6, Supplement 1), S51–S58.

Wong, R., Cheung, C., Xiao, S. B., & Chan, T. (2015). The instigating, impelling, and inhibiting forces in cyberbullying perpetration across gender. Retrieved from. In *PACIS 2015 proceedings* http://aisel.aisnet.org/pacis2015/109.

Yin, D., Xue, Z., Hong, L., Davison, B., Kontostathis, A., & Edwards, L. (2009). Detection of harassment on web 2.0. In *Presented at the proceedings of the 1st content analysis in web 2.0 workshop, Madrid, Spain*.

Zhao, S., Grasmuck, S., & Martin, J. (2008). Identity construction on Facebook: Digital empowerment in anchored relationships. *Computers in Human Behavior, 24*(5), 1816–1836. http://dx.doi.org/10.1016/j.chb.2008.02.012.