Information Systems Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Standing Up or Standing By: Understanding Bystanders' Proactive Reporting Responses to Social Media Harassment

Randy Yee Man Wong, Christy M. K. Cheung, Bo Xiao, Jason Bennett Thatcher

To cite this article:
Randy Yee Man Wong, Christy M. K. Cheung, Bo Xiao, Jason Bennett Thatcher (2021) Standing Up or Standing By: Understanding Bystanders' Proactive Reporting Responses to Social Media Harassment. Information Systems Research 32(2):561-581. https://doi.org/10.1287/isre.2020.0983

http://pubsonline.informs.org/journal/isre

INFORMATION SYSTEMS RESEARCH
Vol. 32, No. 2, June 2021, pp. 561–581
ISSN 1047-7047 (print), ISSN 1526-5536 (online)

Standing Up or Standing By: Understanding Bystanders' Proactive Reporting Responses to Social Media Harassment

Randy Yee Man Wong,[a] **Christy M. K. Cheung,**[b] **Bo Xiao,**[c] **Jason Bennett Thatcher**[d]

[a] Department of Information Systems and Operations Management, The University of Auckland, Auckland 1010, New Zealand; [b] Department of Finance and Decision Sciences, Hong Kong Baptist University, Hong Kong; [c] Shidler College of Business, University of Hawaii at Manoa, Honolulu, Hawaii 96822; [d] Fox School of Business, Temple University, Philadelphia, Pennsylvania 19122

Contact: rrymwong@gmail.com, https://orcid.org/0000-0001-6585-9973 (RYMW); ccheung@hkbu.edu.hk, https://orcid.org/0000-0003-4411-0570 (CMKC); boxiao@hawaii.edu (BX); jason.thatcher@temple.edu, https://orcid.org/0000-0002-7136-8836 (JBT)

Received: May 10, 2017
Revised: September 12, 2018; August 24, 2019; July 8, 2020; September 7, 2020
Accepted: September 14, 2020
Published Online in Articles in Advance: February 17, 2021

https://doi.org/10.1287/isre.2020.0983

Abstract. Social media harassment, a cyberbullying behavior, poses a serious threat to users and platform owners of social media. A growing body of research suggests involving bystanders in interventions to combat deviant behaviors. In this paper, we contextualize the bystander intervention framework and reporting literature to social media in order to understand why bystanders report social media harassment. Our contextualized intervention framework focuses on three sociotechnical aspects—the online social environment, characteristics of the technology platform, and their interplay—that explain bystander reporting on social media platforms. We tested the model using data gathered from 291 active Facebook users. We found that four contextualized factors, (1) perceived emergency of the social media harassment incident, (2) perceived responsibility to report, (3) perceived self-efficacy in using built-in reporting functions, and (4) perceived outcome effectiveness of built-in reporting functions for tackling social media harassment, shaped bystanders' willingness to intervene against social media harassment. In addition, we showed that perceived anonymity of the reporting system counterbalances the negative influence of the presence of others on bystanders' willingness to intervene. For research, we contribute to the cyberbullying literature by offering a novel sociotechnical explanation of mechanisms that shape bystanders' willingness to report social media harassment. For practice, we offer insight into how to build safer and secure social media platforms for all users.

History: Alessandro Acquisti, Senior Editor; Likoebe Maruping, Associate Editor.

Funding: This work was supported by a Hong Kong Baptist University Faculty Research Grant [Grant FRG2/16-17/014] and by a fellowship award from the Research Grants Council of the Hong Kong Special Administrative Region, China [Award HKBU SRFS2021-2H03].
Supplemental Material: The online appendix is available at https://doi.org/10.1287/isre.2020.0983.

Keywords: social media harassment • bystanders • built-in reporting functions • anonymity • presence of others • Facebook • cyberbullying • reporting intervention • sociotechnical perspective

1. Introduction

Online harassment, the most prevalent form of cyberbullying, involves individuals deliberately disseminating rude, threatening, or offensive content directed at individuals or groups through information communication technologies (Wolak et al. 2007). Studies have found that online harassment on social media platforms (hereafter referred to as social media harassment) is a widespread phenomenon that affects all age groups. Plan International UK found that 43.9% of adolescents between 11 and 18 experienced social media harassment (Plan International UK 2017). The Pew Research Center reported that 58% of adults experienced social media harassment (Duggan 2017). Evidence suggests that social media harassment negatively impacts social media users, resulting in adverse outcomes such as suicidal ideations, social anxiety, substance abuse, diminished life satisfaction, and delinquency (Hinduja and Patchin 2010, Slonje et al. 2013, Veletsianos et al. 2018).

Social media harassment is pervasive across social media platforms (Van Royen et al. 2017), exerting increasing public pressure on platforms such as Facebook and Twitter to actively fight harassment. Furthermore, 84% of users believe that platform owners have a responsibility to protect them from social media harassment (Anti-Defamation League 2019). Users' attributions of responsibility to social media platforms are echoed in educators' and government agencies' requests for social media platforms to alleviate social media harassment by introducing digital safety plans (Duggan 2017), which essentially makes social media owners not only responsible for providing platforms for user activities but also for moderating public discourse (Boyd 2010, Crawford and Gillespie 2016, Turel et al. 2019).

Social media harassment typically involves three actors: bystanders, perpetrators, and victims (Wong-Lo and Bullock 2014). Bystanders who witness harassment can intervene in three major ways: reporting the harassment, defending the victim, and supporting the victim (Dillon and Bushman 2015). Because administrators control access to platform features, reporting social media harassment to platforms may be the most efficient and effective way to stop cyberbullying (Wong-Lo and Bullock 2014). Though many social media platform owners (e.g., Facebook, Twitter, LinkedIn, and YouTube) have implemented built-in social media harassment reporting functions (see Table A-1 in Online Appendix A) (Facebook Safety 2011, Kiss 2010), we lack a rich understanding of the factors that shape bystanders' willingness to use them.

A clear understanding of why bystanders report social media harassment is important because platforms need help from bystanders to combat social media harassment. Given the volume of social media content, it is difficult for platform owners to monitor, identify, and review every post. Although Facebook and Instagram have experimented with machine learning and artificial intelligence to detect problematic posts (Griffiths 2019), humans are more apt than machines at recognizing sarcasm and other subtle forms of social media harassment employed by perpetrators (Harris 2017). For instance, Facebook reported that it removed 2.6 million content items related to bullying or harassment in the first quarter of 2019; however, automated technology identified only 14.1% of that content (Facebook 2019). Thus, bystanders can play a critical role in mitigating the impact of social media harassment by identifying offensive materials or behavior on social media platforms.

Furthermore, given that victims of social media harassment tend not to take action (Price and Dalgleish 2010, Paul et al. 2012), motivating bystanders on social media platforms to use reporting functions may be an effective strategy to deter social media harassment. A recent survey found that although 60% of social media users believe that bystanders have the responsibility to mitigate social media harassment, just 30% of bystanders reported intervening after witnessing social media harassment (Duggan 2017). Of these bystanders, 17% used the platforms' reporting tools to flag content, whereas 12% used these tools to report the perpetrator (Duggan 2017). Thus, there is a need to understand the factors driving bystanders to use built-in reporting functions to report social media harassment incidents.

To achieve this objective, we draw on the literature on bystander intervention (Latané and Darley 1970) and reporting behavior (e.g., Lowry et al. 2013) and adopt a sociotechnical perspective (Lee 2004, Bostrom et al. 2009, Sarker et al. 2019) to develop a contextualized research model. Using a sociotechnical perspective allows us to capture both social and technical aspects in framing and investigating technology-related phenomena (Lee 2004, Bostrom et al. 2009, Sarker et al. 2019). This perspective, with a focus on aligning the system's design with the social environment in order to produce desired outcomes, fits well into our investigation of bystanders' willingness to use built-in reporting functions to further firm goals of responding to social media harassment.

Developing a contextualized understanding of bystanders' willingness to report social media harassment is important for several reasons. First, whereas information systems (IS) studies have investigated the positive (e.g., well-being) (Wenninger et al. 2019) and negative consequences (e.g., envy and loneliness) (Krasnova et al. 2015, Matook et al. 2015) of social media use, limited research has investigated the impact of these characteristics in mitigating bad behavior (e.g., social media harassment) and encouraging good behavior (e.g., bystander intervention) on social media platforms. In addition, the context of social media harassment differs from traditional applications of the bystander intervention framework in that social media harassment is reported in the same place as the harassing act—namely, on the social media platform itself. In contrast, traditional bystander interventions separate the physical location of the harassing act from the reporting act. For example, it would not be unusual for a report of social media harassment to be filed on the same platform, whereas a charge of face-to-face harassment would typically require a bystander to change locations to report the incident to a supervisor or human resource professional. Thus, bystander reporting interventions on social media platforms are unique in that the harassment and the reporting take place within the same sociotechnical system offered by the social media platform.

This paper unfolds as follows. First, we provide a theory and model for bystander reporting interventions on social media platforms. Then, we describe our research design and results. Finally, we conclude with a discussion of the results, their limitations, and their implications for future work. Our work contributes to efforts to mitigate social media harassment by shedding light on the social and technical aspects that shape bystanders' willingness to use reporting functions on social media platforms.

2. Research Background

In this section, we review research on online harassment, bystanders' responses, and reporting behavior and provide a summary of built-in reporting functions on popular social media platforms. We use this review to contextualize the bystander intervention framework (Latané and Darley 1970) to social media harassment reporting.

2.1. Online Harassment and Bystanders' Responses

Online harassment refers to any threatening and/or offensive message sent directly to a victim or posted publicly about a victim by means of an online communication medium. It often involves a disruptive event that follows the pattern of cyberbullying[1] and results in distress to the victim (Jones et al. 2013). Studies of online harassment focus on (1) comparisons between offline and online harassment (e.g., Wolak et al. 2007, Sumter et al. 2012), (2) the way in which online features accelerate the prevalence of online harassment (e.g., Moore et al. 2012, Lowry et al. 2016b), and (3) the characteristics (or profiles) of perpetrators and victims (e.g., Finn 2004, Wong et al. 2018, Calvete et al. 2010, Huang and Chou 2010). Although research has directed attention toward perpetrators and victims (Ybarra et al. 2007, Lindsay et al. 2016, Chan et al. 2019), few studies examine bystanders in the context of harassment occurring on social media platforms (Jones et al. 2015, Chan et al. 2021).

Bystanders are witnesses of harassment and other acts of violence that they neither perpetrate nor by which they are directly victimized (Twemlow et al. 2004). Bystanders can play positive or negative roles in online harassment (Salmivalli 2010, Pozzoli and Gini 2013a). *Reinforcers* are bystanders who engage in negative bystander behaviors—such as actively supporting harassment (via forwarding, leaving comments, or clicking the "like" button on harassing posts) or passively ignoring incidents (Shultz et al. 2014)—in an attempt to reinforce the undesirable behavior of the perpetrators and magnify the negative impact on victims (Macháčková et al. 2013, Runions et al. 2013). In contrast, *upstanders* are bystanders who are willing to engage in positive behaviors—such as reporting harassment to platform administrators, or defending, consoling, or supporting the victims—in an attempt to mitigate the harm to victims and/or stop the harassment (Cassidy et al. 2013). Bystander intervention is considered a highly effective means of curbing online harassment (Wong-Lo and Bullock 2014). Our research seeks to identify factors that encourage bystanders to become upstanders who are willing to use a platform's built-in reporting functions to intervene in social media harassment.

2.2. Prior Literature on Reporting Behavior

Our literature review reveals that although reporting behavior has been studied in different contexts (see Table B-1 in Online Appendix B), prior research has not examined bystander reporting intervention in the context of social media harassment. Among the various forms of reporting behavior, whistleblowing, which refers to "the disclosure by organization members . . . of illegal, immoral, or illegitimate practices under the control of their employers, to persons or organizations that may be able to effect action" (Near and Miceli 1995, p. 680), has been found to be a particularly effective tool to fight unethical practices/wrongdoing in organizations (Park et al. 2008, Park and Keil 2009, Smith and Keil 2003). For instance, the model of whistleblowing has been used to explain the reporting of computer abuse (Lowry et al. 2013) and bad news about information technology projects in organizational contexts (Park et al. 2008, Park and Keil 2009). Although previous IS studies offer valuable insights on reporting behaviors and reporting systems, such work requires nuance and extension to explain bystanders' reporting of social media harassment outside of organizational contexts.

Absent specific guidance from academic research, it is not surprising that platform owners implement different functions for facilitating bystander reports of social media harassment. Table A-1 in Online Appendix A presents a list of popular social media platforms and their harassment reporting functions (GlobalWebIndex 2018). Reporting mechanisms generally share technical features; however, social media platforms require different information from bystanders who report an incident. For example, Facebook, Instagram, and LinkedIn ask users to specify the incident type, whereas Twitter requires users to provide their name and email when reporting. Furthermore, some social media platforms (e.g., Instagram, LinkedIn, and YouTube) provide limited feedback to bystanders who report posts, giving them no indication as to whether review teams actually responded to reports of social media harassment.

2.3. The Bystander Intervention Framework

The bystander intervention framework (Latané and Darley 1970) was developed to explain why bystanders take action to intervene in order to curtail personal harassment. The framework suggests that bystanders' assessment of a harassment event they are witnessing, their sense of personal responsibility to intervene, and their actual capacity to intervene are all factors that shape their helping responses. The framework also directs attention to bystanders' evaluation of the presence of others (i.e., pluralistic ignorance, diffusion of responsibility, and evaluation apprehension) as shaping responses (Darley and Latané 1968; Latané and Darley 1968, 1970). The bystander intervention framework has been used to examine helping behavior in various contexts, such as sexual violence (Banyard 2011), domestic violence (Hoefnagels and Zwikker 2001), and traditional bullying (Pozzoli and Gini 2013b).

Prior research has drawn on the bystander intervention framework to investigate bystanders in the context of cyberbullying (e.g., Allison and Bussey 2016, Brody and Vangelisti 2016, Obermaier et al. 2016). However, such research has mostly considered *social psychological* factors, such as demographic variables (e.g., age, gender, and education level) (e.g., Quirk and Campbell 2015, DeSmet et al. 2016), relationship with the victim/perpetrator (e.g., Macháčková et al. 2013, Song and Oh 2018), and empathy (e.g., Van Cleemput et al. 2014) to further the understanding of bystander intervention in the online environment (see Table C-1 in Online Appendix C). Notably, scant research has examined how *sociotechnical* factors shape bystanders' willingness to report social media harassment.

2.4. Understanding the Social Media Bystander Reporting Intervention: A Sociotechnical Perspective

Social media refer to "online platforms where people form communities in which they create, exchange, comment, recreate, and cocreate content" (Karahanna et al. 2018, p. 738). Platform owners have implemented built-in reporting functions for bystanders to use if they witness social media harassment. Bystander reporting interventions on social media platforms are distinct from interventions performed in offline environments in that the willingness to intervene requires considering social and technical components because social media can be viewed both as online social environments in which social media harassment incidents take place and as technology platforms on which reporting functions are embedded. In this study, we adopt the sociotechnical perspective to explain why bystanders report harassment on social media platforms.

The sociotechnical perspective, which views systems as comprised of elements of the technology and the social environment, fits well into our investigation of drivers of bystanders' willingness to use built-in reporting functions to report social media harassment. First, social media are inherently social technology platforms that enable connectivity and communication among users and offer visibility of those social interactions, facilitated by built-in design mechanisms (such as message transparency and network translucence) (Leonardi 2015). Specifically, bystanders can easily understand a harassing incident on social media, assess their personal responsibility to intervene, and observe how other bystanders react to the incident (i.e., the presence of others). Second, bystanders' use of built-in reporting functions likely results from social considerations (e.g., privacy and presence of others) within the context of a technically enabled environment. Specifically, bystanders' willingness to report through built-in reporting functions reflects a trade-off between confidence in the anonymity of the reporting function and awareness of the presence of others in a social environment. Third, the influence of social and technical components may be intertwined in how they shape bystanders' view of how design features (e.g., built-in reporting functions) enable a social process (e.g., reporting behaviors). As technology platforms, social media make them easier for bystanders to articulate the implications of reporting social media harassment because they directly interact with the platforms and the embedded reporting functions. Because bystanders understand the platforms and control whether they use built-in reporting functions, they are likely better positioned to evaluate their own ability to use the technology and predict the likely outcomes of such use (i.e., self-efficacy and outcome expectancy) (DeSmet et al. 2016). Furthermore, bystanders can assess the managerial practices and regulatory structures of platform owners (i.e., perceived reporting climate and perceived reporting justice) to judge the likely effectiveness of using built-in reporting functions to mitigate social media harassment incidents.

In sum, our study uses the three aspects noted above to develop a sociotechnical perspective on social media reporting. Placing salience on both social and technical components (as well as their interplay) in framing and investigating social media bystander reporting interventions, our study aims to capture the distinctive nature of this technology-related phenomenon (Sarker et al. 2019).

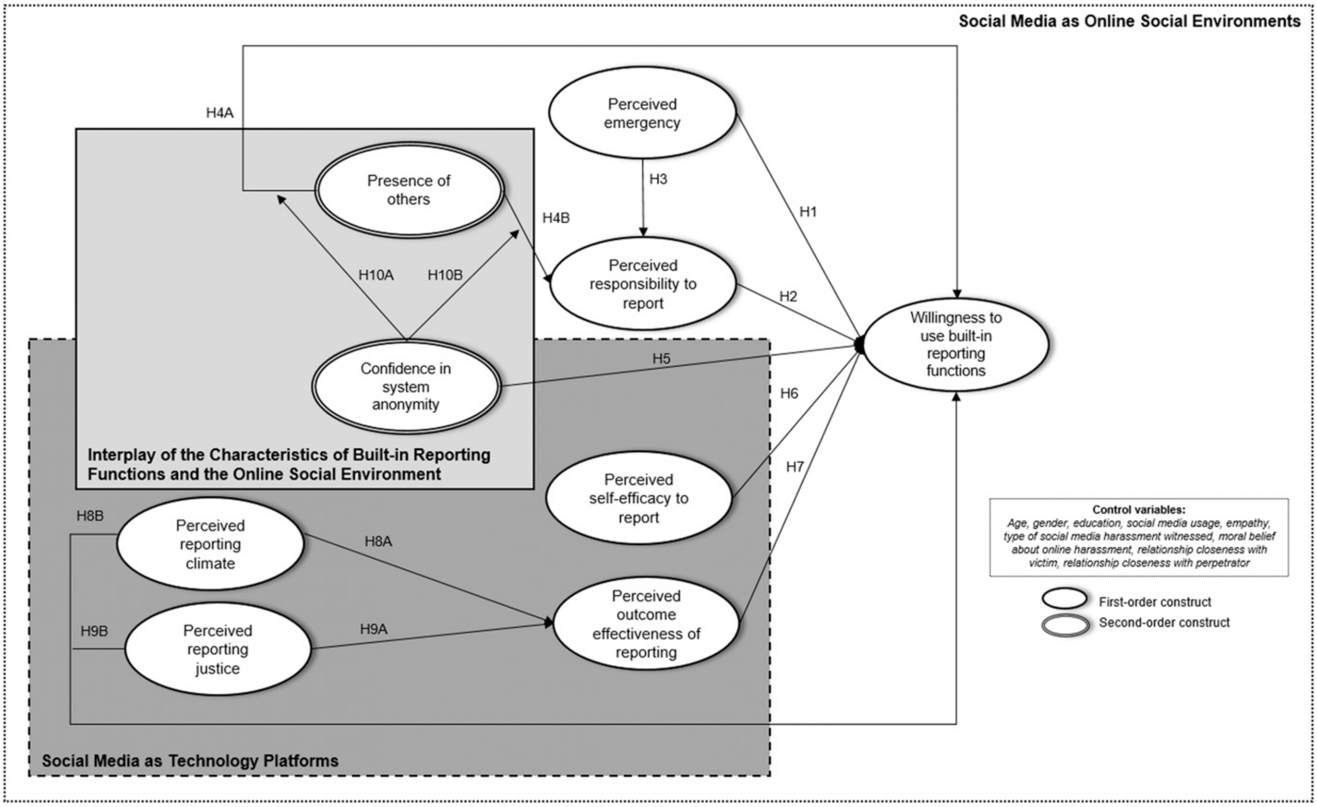
3. Research Model and Hypothesis Development

In this section, we articulate a contextualized social media bystander reporting intervention framework (Latané and Darley 1970) in order to develop a sociotechnical model (Sarker et al. 2019) of bystanders' willingness to use the platform's built-in reporting functions to report social media harassment (see Figure 1).

3.1. Contextualizing the Factors of the Bystander Intervention Framework

Our contextualized social media bystander reporting intervention framework suggests that four main factors

Figure 1. The Research Model


Social Media as Online Social Environments
H4A
Perceived emergency
Presence of others
H3
H4B
H1
H10A
H10B
Perceived responsibility to report
H2
Willingness to use built-in reporting functions
H5
Confidence in system anonymity
H6
Interplay of the Characteristics of Built-in Reporting Functions and the Online Social Environment
H7
Perceived self-efficacy to report
Control variables:
Age, gender, education, social media usage, empathy, type of social media harassment witnessed, moral belief about online harassment, relationship closeness with victim, relationship closeness with perpetrator
H8B
Perceived reporting climate
H8A
Perceived outcome effectiveness of reporting
First-order construct
Second-order construct
H9B
Perceived reporting justice
H9A
Social Media as Technology Platforms

shape bystanders' willingness to use built-in reporting functions: (1) their assessment of the emergency of the social media harassment situation, (2) their sense that it is their personal responsibility to report the incident, (3) the capacity to intervene, and (4) the presence of others—that is, whether other bystanders are also present for helping behaviors. We map the core concepts of the bystander intervention framework to constructs in our model in order to predict bystanders' willingness to use built-in reporting functions to report social media harassment (see Table 1).

3.2. Social Media and Bystander Reporting Interventions

As discussed in Section 2.4, bystander reporting interventions on social media platforms are distinctive in that bystanders' willingness to intervene involves a consideration of both social and technical components. In this study, our contextualized model of the social media bystander reporting intervention includes three distinct sociotechnical components: online social environments where social media harassment incidents take place, technology platforms on which reporting functions are embedded, and the interplay of the characteristics of built-in reporting functions and the online social environment.

3.2.1. Social Media as Online Social Environments

Social media platforms have changed how users communicate. Social media make the content of users' message exchanges transparent (i.e., message transparency) and their network connections translucent (i.e., network translucence) (Leonardi 2014, 2015). In the context of social media harassment, the metaconcepts of message transparency (e.g., awareness of the authorship on and the content of the harassing posts) and network translucence (e.g., awareness of the network ties of those involved in the harassing events) influence bystanders' assessments of their reporting interventions on social media platforms (summarized in Table 2).

Perceived emergency refers to the extent to which bystanders perceive that a social media harassment incident needs to be urgently addressed. The assessment of emergency is important because it exerts a significant influence on individuals' willingness to engage in prosocial interventions (Manstead and Fischer 2001, Burn 2009, Nickerson et al. 2014). Emergencies are thought to draw bystanders' attention to observable details of an incident and motivate them to take actions (Dovidio et al. 2006, Loewenstein and Small 2007). Emergency is relevant to social media harassment because social media create online environments that increase communication visibility—that is,

Table 1. Social Media Bystander Reporting Intervention Model

The bystander intervention framework	Core constructs in our research model
Decision of intervention	*Willingness to use built-in reporting functions* is defined as bystanders' willingness to report social media harassment incidents to platform owners by using built-in reporting functions of social media platforms
Assessment of the event	Assessment of the social media harassment incident —*Perceived emergency* is defined as the extent to which bystanders believe that the social media harassment incident needs to be addressed urgently
Assessment of personal responsibility	Assessment of personal responsibility to report the incident —*Perceived responsibility to report* is defined as bystanders' subjective assessment of their sense of personal obligation to deal with social media harassment incidents
Assessment of capability to intervene (personal and situational factors)	Assessment of capability to intervene —*Perceived self-efficacy to report* is defined as bystanders' subjective assessment of their ability to successfully report the harassment using built-in reporting functions on social media platforms —*Perceived outcome effectiveness of reporting* is defined as the extent to which bystanders believe that using built-in reporting functions on social media platforms will effectively tackle social media harassment
Presence of others	Presence of others —*Pluralistic ignorance* is defined as the extent to which bystanders believe that other bystanders who have also witnessed the incident will remain unconcerned with the social media harassment incident on the social media platform —*Diffusion of responsibility* is defined as the extent to which bystanders believe that reporting responsibility should be transferred to other bystanders who have also witnessed the incident —*Evaluation Apprehension* is defined as bystanders' fear of being judged or negatively evaluated when using built-in reporting functions to report social media harassment incidents

it makes it possible to directly observe not only the harassing content itself in the original messages (i.e., message transparency) but also how other bystanders react to such content (e.g., via reposts, comments, and likes) (i.e., network translucence) (McFarland and Ployhart 2015), which can therefore invoke feelings of emergency among bystanders. If bystanders perceive a social media harassment incident as an emergency that requires prompt action, they are more likely to report the incident using

Table 2. Social Media Characteristics and Bystander Intervention Decision

	Social media characteristics	
Bystander intervention	Message transparency	Network translucence
Assessment of the event (perceived emergency)	Social media allow the content of messages (i.e., harassing posts) to be easily and effortlessly seen by users. Bystanders can assess the emergency of a harassment incident based on the harassing content.	Social media allow users to observe other users involved in social media harassment (e.g., perpetrators, victims, bystanders). Bystanders can assess the emergency of the incident based on the characteristics of other users (e.g., how many bystanders have viewed/shared/commented on the harassing posts).
Assessment of personal responsibility (perceived responsibility to report)	Social media allow the content of messages (i.e., harassing posts) to be easily and effortlessly seen by users. Bystanders can assess their personal responsibility for intervention based on the harassing content.	Social media allow users to observe other users involved in social media harassment (e.g., perpetrators, victims, bystanders). Bystanders can assess their personal responsibility to intervene based on the characteristics of other users (e.g., how many bystanders have viewed/shared/reacted/commented on the harassing posts).
Presence of others	Social media allow the content of others' interactions (i.e., what they like, comment, and share) to be easily and effortlessly seen by users. Bystanders can assess the influence of the presence of others on helping behaviors based on others' interactions (e.g., how other bystanders comment on an intervention).	Social media allow users to observe other users who are involved in a social media harassment incident (e.g., other bystanders). Bystanders can assess the influence of the presence of others on helping behaviors based on the characteristics of others' interactions (e.g., how many bystanders there are to help; who the other bystanders are).

built-in reporting functions. We thus hypothesize the following.

Hypothesis 1. *Perceived emergency will have a positive effect on bystanders' willingness to use built-in reporting functions to report social media harassment incidents to platform owners.*

Perceived responsibility to report refers to bystanders' subjective assessments of their sense of obligation to respond to social media harassment incidents personally (Gracia et al. 2008). Across many different types of crime, violence, and wrongdoing (Finkelhor and Wolak 2003, Tarling and Morris 2010, Edwards et al. 2013), bystanders who feel moral responsibility are likely to intervene to address an incident (Laible et al. 2008). In the context of social media, bystanders' feelings of moral responsibility may be evoked by viewing harassing content (i.e., message transparency) or the parties involved in the incident (i.e., network translucence). For example, a bystander may feel more responsibility to report if the incident is particularly extreme (e.g., a post threatening physical harm) or if the victim of social media harassment is part of their social network (e.g., a close friend, acquaintance, or online follower). Therefore, we expect that bystanders who perceive a personal responsibility to report will report greater willingness to use built-in reporting functions to report social media harassment incidents to platform owners. We thus hypothesize the following.

Hypothesis 2. *Perceived responsibility to report will have a positive effect on bystanders' willingness to use built-in reporting functions to report social media harassment incidents to platform owners.*

Because social media increase the visibility of harassing content and other users' responses to such content (Zhao et al. 2011), we suspect that greater feelings of emergency triggered by social media harassment will lead to bystanders' feelings of personal responsibility to intervene because feelings of emergency will make threat assessments more salient (Latané and Darley 1970). We thus hypothesize the following.

Hypothesis 3. *Perceived emergency will have a positive effect on perceived responsibility to report.*

The presence of others refers to bystanders being less likely to intervene when they know others are present and observing their behavior. The bystander intervention literature suggests that such "nonaction" results from bystanders demonstrating pluralistic ignorance, diffusion of responsibility, and evaluation apprehension (Latané and Nida 1981, Fischer et al. 2011). Specifically, because social media bystanders can easily observe who has reacted to a disruptive event as well as how they reacted (Thornberg 2007, Schacter et al. 2016), they may form intentions to act based on others' responses. If bystanders see that no one else is intervening (i.e., no feedback in the form of likes, comments, or shares on harassing posts), they may conclude that no action is needed (i.e., pluralistic ignorance). Also, if social media bystanders perceive the presence of others on the social media platform (e.g., via the list of connections or the "Who is available to chat" function provided by Facebook in real time), they may diffuse responsibility and expect other witnesses to the incident to take action (i.e., diffusion of responsibility) (Obermaier et al. 2016), potentially resulting in no help being offered to the victim. Finally, humans care about social rewards and punishments and are generally concerned about how others evaluate them and their actions (i.e., through easy access to online profiles displaying personal information and a record of activities) (Manstead and Fischer 2001, Burn 2009, Nickerson et al. 2014). Because social media content is traceable and users' can directly observe the interactions of others on social media (Leonardi and Vaast 2017), bystanders tend to be more conservative in their actions in order to avoid retaliation by a perpetrator or negative evaluation by other community members (Brody and Vangelisti 2016, Song and Oh 2018). As such, bystanders witnessing social media harassment may be reluctant to intervene if they believe that their actions will be judged negatively or retaliated against by others on the social media platform (i.e., evaluation apprehension). Thus, we hypothesize the following.

Hypothesis 4A. *The presence of others will have a negative effect on bystanders' willingness to use built-in reporting functions on social media platforms.*

Empirical evidence suggests that the presence of others negatively impacts a bystander's sense of personal responsibility. For instance, Koedinger and Aleven (2007) observed that when bystanders perceive the presence of others in their social circles, they tend to minimize their responsibility to help and underuse available intervention resources. Obermaier et al. (2016) also found that when the number of witnesses to cyberbullying incidents on Facebook increased, bystanders were less likely to feel a personal responsibility to respond and less inclined to attempt to intervene in cyberbullying incidents. We expect that the presence of others reduces bystanders' sense of personal responsibility to report social media harassment. Therefore, we hypothesize the following.

Hypothesis 4B. *The presence of others will have a negative effect on bystanders' perceived responsibility to report social media harassment.*

3.2.2. Social Media as Technology Platforms.

Platform owners offer not only space for users to engage in online activities but also tools for users to help monitor discourse on this space (Crawford and Gillespie 2016). Such help from bystanders is invaluable for making social media platforms safer because automated solutions have been found to miss harassment incidents (Facebook 2019). Platform owners solicit help through affording bystanders access to built-in reporting functions, which allow users to flag or report offensive content or harassing incidents. When a bystander flags a problematic post or behavior, in most cases, the platform's review team examines the report and either removes materials deemed offensive or blocks the perpetrator from accessing his or her account. To better understand how platform owners can encourage the use of built-in reporting functions, we direct attention to the platform's technical features, bystanders' personal attributes, and perceived outcomes.

Confidence in system anonymity refers to the extent to which bystanders believe in the anonymity of the system when using built-in reporting functions. Confidence in system anonymity is important, particularly because well-publicized security scandals (e.g., Facebook's Cambridge Analytica data scandal in 2018) have intensified users' concerns about privacy on social media platforms (*eMarketer* 2019). Studies suggest that when individuals perceive reporting to be anonymous, they feel less concerned about personal costs and may thus be more likely to report disruptive events (Keil et al. 2010, Lowry et al. 2013). Similarly, if users have confidence in the anonymity of an online reporting function, they will be more likely to report social media harassment incidents via the built-in reporting function because they will be less concerned about potential retaliation from a perpetrator on the social media platform. We predict that confidence in system anonymity increases bystanders' willingness to use built-in reporting functions to report social media harassment incidents. Thus, we hypothesize the following.

Hypothesis 5. *Confidence in system anonymity will have a positive effect on bystanders' willingness to use built-in reporting functions on social media platforms.*

Appraisal theories suggest that individuals evaluate resources available to effect change when deciding whether to take action in a situation (Folkman et al. 1986). Resources can be found within the self (e.g., in the form of efficacy beliefs) or externally (e.g. in the form of support from an organization). Ample evidence suggests that if users believe they are capable of effecting change, they will perform protective actions (Lee and Larsen 2009, Liang and Xue 2010, Bala and Venkatesh 2015, Tu et al. 2015). We examine two appraisals of capabilities that shape how bystanders respond to harassment incidents (Latané and Darley 1970): perceived self-efficacy to report and perceived outcomes of reporting. *Perceived self-efficacy to report* refers to bystanders' personal judgment of their ability to perform reporting acts using built-in reporting functions on social media platforms. We also include in our research model the *perceived outcome effectiveness of reporting*, which refers to the extent to which bystanders believe that using built-in reporting functions on social media platforms is an effective means of tackling social media harassment.

If bystanders have confidence in their ability to use built-in reporting functions and in the outcomes of such use, they will likely be more willing to report social media harassment to platform owners. Although it is well-established that efficacy shapes behavior (Compeau and Higgins 1995), it is less certain how efficacious social media users believe built-in reporting functions to be. Moreover, it is less certain how beliefs about the platform's capability (e.g., outcomes of use) relate to users willing to use built-in reporting functions. Our uncertainty is echoed in industry reports on cyberbullying and social media. Consider a recent Safety Net report (The Children's Society and YoungMinds 2018) showing that users believe they should receive training on reporting and that they want access to tools and technological solutions for addressing online harassment. The report's findings align with our intuition that individuals' beliefs about personal capability in conjunction with beliefs about tools shape bystanders' willingness to report social media harassment incidents (by using reporting tools or other technological solutions). Thus, we hypothesize the following:

Hypothesis 6. *Perceived self-efficacy to report will have a positive effect on bystanders' willingness to use built-in reporting functions on social media platforms.*

Hypothesis 7. *Perceived outcome effectiveness of reporting will have a positive effect on bystanders' willingness to use built-in reporting functions on social media platforms.*

Research on reporting (Dozier and Miceli 1985, Miceli and Near 1985) suggests that organizations (i.e., platform owners) play a key role in influencing reporting interventions. Individuals generally evaluate the effectiveness of reporting interventions in terms of structures and policies (i.e., perceived reporting justice) and the managerial practices (i.e., perceived reporting climate) of organizations (i.e., platform owners). In the context of social media reporting interventions, bystanders' appraisal of the managerial and regulatory practices of a platform (i.e., perceived reporting climate and perceived reporting justice) influences their willingness to intervene.

Perceived reporting climate refers to bystanders' perception of the extent to which a social media platform encourages and supports reporting. If individuals perceive that the platform climate encourages expressing personal views and opinions, they are more likely to speak up (Wei et al. 2015). Bullying research suggests that the perceived climate reinforces how bystanders perceive the likely efficacy of an intervention, which, in turn, may encourage bystanders to defend victims (Gini et al. 2008, Barchia and Bussey 2011, Pöyhönen et al. 2012). In the context of our study, supporting the perception that bystanders' reporting behavior is safe and welcome in the social media community (e.g., through a "thank you" message sent by the platform to bystanders reporting social media harassment or an initiative taken by community members to raise awareness of social media harassment) may enhance bystanders' beliefs about the effectiveness of using these tools to combat social media harassment and encourage bystanders to report social media harassment incidents using reporting tools. Therefore, we hypothesize the following.

Hypothesis 8A. *Perceived reporting climate will have a positive effect on bystanders' perceived outcome effectiveness of reporting social media harassment incidents.*

Hypothesis 8B. *Perceived reporting climate will have a positive effect on bystanders' willingness to use built-in reporting functions on social media platforms.*

Perceived reporting justice refers to the extent to which bystanders believe that their reporting behaviors will be treated fairly by social media platform owners. It involves bystanders' perceptions that they will be treated fairly by the report-receiving authorities (i.e., the social media platform owners) and that the outcome of reporting procedures will also be fair. Researchers have found that if bystanders lack confidence in the fairness of reporting processes, particularly regarding how the relevant authorities will handle submitted reports, they tend not to report incidents (Miceli and Near 1992). Studies on computer crime reporting, peer reporting, and whistleblowing (Skinner and Fream 1997, Lewis 2011, Sulkowski 2011) demonstrate that perceived justice affects response outcomes (e.g., outcome effectiveness and intervention adoption). Echoing this intuition, social media users have urged platform owners to implement transparent review processes with concrete review policies for reporting social media harassment (The Children's Society and YoungMinds 2018). As the perception of reporting justice grows, evidence suggests that bystanders will expect positive outcomes of using built-in reporting functions to report social media harassment and feel more motivated to report social media harassment. Thus, we hypothesize the following.

Hypothesis 9A. *Perceived reporting justice will have a positive effect on bystanders' perceived outcome effectiveness of reporting social media harassment incidents.*

Hypothesis 9B. *Perceived reporting justice will have a positive effect on bystanders' willingness to use built-in reporting functions on social media platforms.*

3.2.3. The Interplay of the Characteristics of Built-In Reporting Functions and the Online Social Environment.

In contrast to traditional reporting interventions (in which the harassing act and the reporting act typically occur in different locations), social media reporting interventions often occur in the same place as the harassing act—namely, on the social media platform itself. Parties involved in the social media harassment incidents (i.e., perpetrators, victims, and bystanders) use the same social media platform to harass and report harassment. As such, their offline identities and social networks (i.e., lists of friends and followers) may be easily identifiable through information posted on their public or private profiles. Furthermore, platform owners or users with application programming interface (API)[2] access can retrieve and trace posts back to their sources (McFarland and Ployhart 2015), making authors identifiable and potentially responsible for the content of their posts. Thus, bystanders have legitimate concerns about protecting their identity when they use built-in reporting functions embedded in the social media platform.

We expect that confidence in system anonymity not only drives bystanders to use built-in reporting functions but also inhibits the negative impact of the presence of others on bystander reporting interventions. Reporting behavior and perceived responsibility to report social media harassment are more likely when the bystander effect (i.e., the presence of others) is low and confidence in system anonymity is high. In other words, perceived anonymity of the reporting system likely counterbalances the negative influence of the presence of others on bystanders' willingness to intervene. For instance, if bystanders perceive the reporting system to be anonymous (i.e., perceive that no sensitive personal information is captured during the reporting process), they will be less concerned about the possibility of being negatively evaluated or retaliated against by others. Confidence in reporting system anonymity can serve as a safeguard against the bystander effect on social media platforms. We thus hypothesize that confidence in system anonymity mitigates the negative impact of the presence of others on bystanders' willingness to use built-in reporting functions.

Hypothesis 10A. *Confidence in system anonymity will positively moderate the relationship between the presence of* *others and bystanders' willingness to use built-in reporting functions on social media platforms. Specifically, the negative effect will be weakened when bystanders' confidence in system anonymity is high.*

Hypothesis 10B. *Confidence in system anonymity will positively moderate the relationship between the presence of others and bystanders' perceived responsibility to report social media harassment. Specifically, the negative effect will be weakened when bystanders' confidence in system anonymity is high.*

4. Research Method

In this section, we detail our research setting, data collection method, and measures.

4.1. Setting and Data Collection

We recruited social media users who had witnessed social media harassment incidents on Facebook in the six months prior to data collection. We focused on Facebook because (1) it is recognized as a prominent platform for social media harassment (Kim and Hancock 2015), and (2) its built-in reporting function directly reports content that may violate Facebook's terms of use to a review team.

We drew our sample of Facebook users from Amazon's Mechanical Turk (MTurk). Registered MTurk users participate in tasks (such as completing surveys) in exchange for remuneration (Ward and Broniarczyk 2011). Consistent with practices suggested by Lowry et al. (2016a), we included several "attention check" questions to detect careless, random, or haphazard responses (Mason and Suri 2012). For example, we asked, "Is Facebook a social networking site?" and "Does heat make ice melt?" In addition, to minimize potential response bias, we followed general principles (i.e., autonomy, beneficence, justices, privacy, and confidentiality) for ethical research practices on human subjects in data collection (Mishna et al. 2012). The participants were informed that the survey was voluntary and anonymous. After completing the questionnaire, participants were debriefed.

We collected 291 useable responses from active Facebook users—161 (55.3%) were female and 130 (44.7%) were male. The age of the participants ranged from 17 to 70. The participants were generally well educated, with approximately 69.4% holding a bachelor's degree or higher. The average participant spent 28 minutes completing our survey (49 minutes maximum, 18 minutes minimum). Online Appendix D, Table D-1 summarizes the sample demographics.

4.2. Procedure and Measurement

The questionnaire included three parts. First, participants recalled a recent harassment incident on Facebook that they had witnessed during the past six months. Second, participants described the incident and assessed it in a questionnaire. Third, they were introduced to the built-in reporting function on Facebook (see Online Appendix E) and answered questions that measured our constructs.

We included nine control variables (see Table F-1 in Online Appendix F) to reduce the possibility of spurious relationships: *moral belief about online harassment*, *empathy*, *type of social media harassment witnessed*, *relationship closeness with victim*, *relationship closeness with perpetrator*, *social media usage* (self-reported), and demographic characteristics, namely, *age*, *gender*, and *education*. Research suggests that these factors may influence individuals' decisions to intervene (Miceli and Near 1992, Tavakoli et al. 2003, Kirkman et al. 2009). We used existing construct measures with the exception of *relationship closeness with victim* and *relationship closeness with perpetrator*. Where necessary, minor modifications tailored items to fit the social media context or to direct attention to the built-in reporting function on Facebook. All measures used a seven-point Likert scale, except for *type of social media harassment witnessed*, which used a nominal scale. Apart from two constructs considered formative at the second-order levels (i.e., presence of others and confidence in system anonymity), all constructs were modeled as reflective indicators. Online Appendix F details the measurement items and their sources.

We validated the second-order formative constructs in consistency with prescriptions found in the research methods literature (MacKenzie et al. 2011, Polites et al. 2012). Our second-order constructs were operationalized as superordinate at the second level and reflective at the first level (Diamantopoulos and Winklhofer 2001, Cenfetelli and Bassellier 2009). *Presence of others* was conceptualized as a formative second-order construct determined by three first-order constructs—pluralistic ignorance, diffusion of responsibility, and evaluation apprehension—because prior research (Darley and Latané 1968, Latané and Darley 1970) suggests that these dimensions constitute bystanders' perceptions of the presence of others in the environment. We measured the three first-order dimensions using scales adapted from Burn (2009), La Greca and Lopez (1998), and Prentice and Miller (1993). *Confidence in system anonymity* was operationalized as a formative second-order construct determined by dissociated anonymity and visual anonymity, the essential anonymity components identified by Suler (2004). To measure these two first-order constructs, we used scales adapted from Lowry et al. (2013).

To ensure the content validity of the measures, we conducted a pretest and pilot test. The pretest involved 30 undergraduate and graduate students evaluating the online questionnaire, with a particular focus on the clarity of instructions, wording of the questions, relevance of the measures, presence of biased words and phrases, use of standard English, and format (Fowler and Cosenza 2009). Based on initial feedback, we removed six items see Online Appendix F for more detail). We then conducted an online pilot test with 100 active users of Facebook. Except for minor modifications to formatting, no major issues were identified in the pilot test.

5. Data Analysis and Results

We used covariance-based structural equation modeling (CB-SEM) through AMOS 22 to run the data. CB-SEM is considered appropriate for validating models with multidimensional constructs (Roberts and Thatcher 2009, Wright et al. 2012). We employed a two-step analytical approach to evaluate the research model, first estimating the psychometric assessment of the measures, and then the structural model.

5.1. Measurement Model Evaluation

5.1.1. Reflective Constructs. We conducted a confirmatory factor analysis that included the reflective latent constructs *(i.e., willingness to use built-in reporting functions, perceived emergency, perceived responsibility to report, perceived self-efficacy to report, perceived outcome effectiveness of reporting, perceived reporting justice, and perceived reporting climate)*, the first-order dimensions of the presence of others *(i.e., pluralistic ignorance, diffusion of responsibility, and evaluation apprehension)*, and confidence in system anonymity *(i.e., dissociative anonymity and visual anonymity)*, as well as the two reflective control variables (i.e., *empathy* and *moral belief about online harassment*).

To evaluate the reflective constructs, we assessed construct reliability, convergent validity, and discriminant validity. First, we estimated composite reliability indices (Fornell and Larcker 1981). All constructs exceeded the 0.7 benchmark, indicating satisfactory construct reliability (see Tables G-1a and G-1b in Online Appendix G). Second, we examined item loadings and the square root of the average variance extracted (AVE) for each construct to assess convergent and discriminant validity. All item loadings were greater than the recommended 0.5 cutoff (Carmines and Zeller 1979), suggesting that the items loaded well on their respective constructs (see Table G-4 in Online Appendix G). In addition, the AVEs of all the constructs were greater than the recommended level of 0.5 (Fornell and Larcker 1981), demonstrating good convergent validity (see Tables G-1a and G-1b in Online Appendix G). The square root of the AVE of each construct was found to be greater than the correlations of the construct with other constructs, demonstrating satisfactory discriminant validity (see Table G-3a in Online Appendix G).

We also assessed the psychometric properties of the overall measurement model. Following the two-index strategy suggested by Hu and Bentler (1999), we evaluated model fit using the comparative fit index (CFI; where values approaching or surpassing 0.9 indicate satisfactory fit) and standardized root mean square residual (SRMR; where values approaching or below 0.08 indicate good fit). Our measurement model shows satisfactory fit with the data (χ^2 = 2,226.10, degrees of freedom (df) = 1081, χ^2/df = 2.06, CFI = 0.91, root mean square error of approximation (RMSEA) = 0.06, SRMR = 0.06).

5.1.2. Second-Order and First-Order Formative Constructs. Following the guidelines and recommendations of operationalizing multidimensional and formative constructs (Petter et al. 2007, Cenfetelli and Bassellier 2009, Wright et al. 2012), we operationalized the presence of others and confidence in system anonymity as second-order aggregate constructs formed from first-order reflective dimensions, whereas the type of social media harassment witnessed was operationalized as a first-order formative construct. First, we evaluated the variance inflation factor (VIF) of the measures of the second-order and first-order formative constructs. All VIFs were below 3.33, indicating the absence of multicollinearity (see Table G-1b in Online Appendix G) (Diamantopoulos and Siguaw 2006, Petter et al. 2007, Cenfetelli and Bassellier 2009). Second, we assessed the zero-order correlation (i.e., absolute contribution) for each second-order construct against the overall average for each construct (see Tables G-2a and G-2b in Online Appendix G). All the items showed significant associations with the overall measure at the 0.05 level of significance. Third, we assessed the weight (i.e., relative contribution) and loading (i.e., absolute contribution) of the first-order formative indicators (see Table G-2c in Online Appendix G) and found that the weights of all indicators were significant. Our results also show that there was no unexpectedly high correlation among the formative indicators (below the 0.9 thresholds). In sum, the results provide evidence of the validity of our second-order and first-order formative constructs.

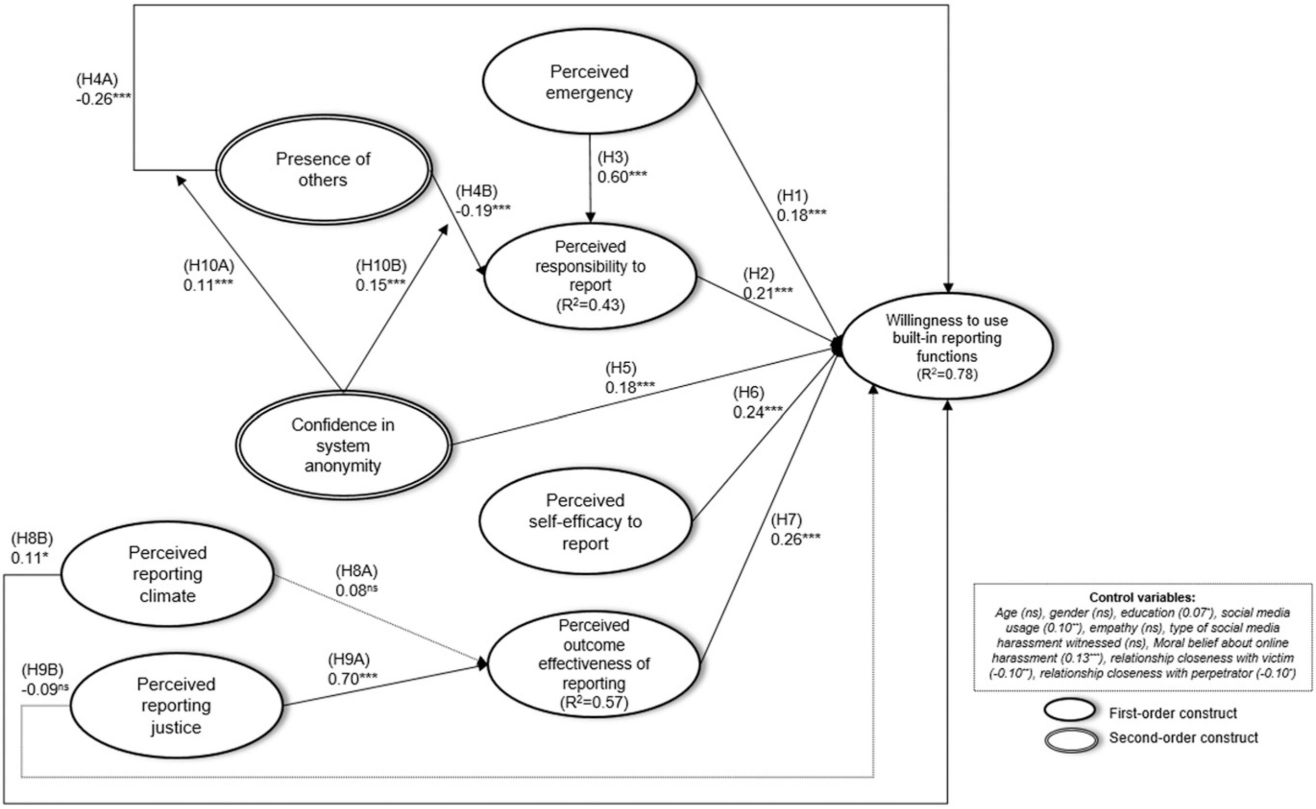
5.1.3. Common Method Bias and Social Desirability Bias. We also tested for common method variance influence (Schwarz et al. 2017) and social desirability bias (Podsakoff et al. 2003). Our results suggest that common method bias and social desirability bias had minimal impact on this study. Online Appendix H details the assessment of common method bias and social desirability bias.

5.2. Structural Model Evaluation

We analyzed our model using a maximum likelihood parameter estimation in CB-SEM. We first estimated a baseline model with only the main effect (Model 1 in Table G-5 in Online Appendix G). The model fit was deemed satisfactory (χ^2 = 69.29, df = 27.00, χ^2/df = 2.57, CFI = 0.98, RMSEA = 0.07, SRMR = 0.04). All hypothesized effects, except those in Hypotheses 8A and 9B, were statistically significant. We then added the interaction terms of confidence in system anonymity and the presence of others on willingness to use built-in reporting functions and perceived responsibility to report, respectively, to test the moderating effects of confidence in system anonymity (Model 2 in Table G-5 in Online Appendix G). The research model accounts for 77.7% of the variance in bystanders' willingness to use built-in reporting functions, 43.4% of the variance in perceived responsibility to report, and 56.5% of the variance in perceived outcome effectiveness of reporting (see Figure 2).

Table 3 presents a summary of our results. Bystanders' perceived emergency (Hypothesis 1, β = 0.18, p < 0.001), perceived responsibility to report (Hypothesis 2, β = 0.21, p < 0.001), presence of others (Hypothesis 4A, β = −0.26, p < 0.001), confidence in system anonymity (Hypothesis 5, β = 0.18, p < 0.001), perceived self-efficacy to report (Hypothesis 6, β = 0.24, p < 0.001), perceived outcome effectiveness of reporting (Hypothesis 7, β = 0.26, p < 0.001), and perceived reporting climate (Hypothesis 8B, β = 0.11, p < 0.05) were significant predictors of bystanders' willingness to use built-in reporting functions. Also, consistent with our predictions, perceived emergency (Hypothesis 3, β = 0.60, p < 0.001) and the presence of others (Hypothesis 4B, β = −0.19, p < 0.001) exerted significant positive and negative influence on perceived responsibility to report, respectively. Perceived reporting justice had a significant impact on bystanders' perceived outcome effectiveness of reporting (Hypothesis 9A, β = 0.70, p < 0.001). When confidence in system anonymity was added to the model as a moderator, it exerted significant positive moderating effects on both the relationship between presence of others and willingness to use built-in reporting functions (Hypothesis 10A, β = 0.11, p < 0.001) as well as the relationship between the presence of others and perceived responsibility to report (Hypothesis 10B, β = 0.15, p < 0.001). Thus, the negative effects of the presence of others on both willingness to use built-in reporting functions and perceived responsibility to report were alleviated when bystanders had a higher

Figure 2. Results of the Research Model


Perceived emergency
(H4A) -0.26***
Presence of others
(H3) 0.60***
(H4B) -0.19***
(H1) 0.18***
Perceived responsibility to report (R²=0.43)
(H10A) 0.11***
(H10B) 0.15***
(H2) 0.21***
Willingness to use built-in reporting functions (R²=0.78)
(H5) 0.18***
(H6) 0.24***
Confidence in system anonymity
Perceived self-efficacy to report
(H8B) 0.11*
Perceived reporting climate
(H7) 0.26***
(H8A) 0.08ns
Control variables:
Age (ns), gender (ns), education (0.07*), social media usage (0.10**), empathy (ns), type of social media harassment witnessed (ns), Moral belief about online harassment (0.13***), relationship closeness with victim (-0.10**), relationship closeness with perpetrator (-0.10*)
(H9B) -0.09ns
Perceived reporting justice
(H9A) 0.70***
Perceived outcome effectiveness of reporting (R²=0.57)
First-order construct
Second-order construct

Note. H, Hypothesis; n.s., not significant (also depicted with a dotted line).
*p < 0.05; **p < 0.01; ***p < 0.001.

Table 3. Results of Hypothesis Testing

Hypotheses	Path (sig.)	Supported?
H1: Perceived emergency → willingness to use built-in reporting functions	0.18***	Yes
H2: Perceived responsibility to report → willingness to use built-in reporting functions	0.21***	Yes
H3: Perceived emergency → perceived responsibility to report	0.60***	Yes
H4A: Presence of others → willingness to use built-in reporting functions	−0.26***	Yes
H4B: Presence of others → perceived responsibility to report	−0.19***	Yes
H5: Confidence in system anonymity → willingness to use built-in reporting functions	0.18***	Yes
H6: Perceived self-efficacy to report → willingness to use built-in reporting functions	0.24***	Yes
H7: Perceived outcome effectiveness of reporting → willingness to use built-in reporting functions	0.26***	Yes
H8A: Perceived reporting climate → perceived outcome effectiveness of reporting	0.08 (n.s.)	No
H8B: Perceived reporting climate → willingness to use built-in reporting functions	0.11*	Yes
H9A: Perceived reporting justice → perceived outcome effectiveness of reporting	0.70***	Yes
H9B: Perceived reporting justice → willingness to use built-in reporting functions	−0.09 (n.s.)	No
H10A: Presence of others × confidence in system anonymity → willingness to use built-in reporting functions	0.11***	Yes
H10B: Presence of others × confidence in system anonymity → perceived responsibility to report	0.15***	Yes

Note. H, Hypothesis; sig., significance; n.s., not significant.
*p < 0.05; ***p < 0.001.

level of confidence in system anonymity. Contrary to our expectations, perceived reporting climate (Hypothesis 8A) and perceived reporting justice (Hypothesis 9B) did not have a statistically significant influence on bystanders' perceived outcome effectiveness of reporting social media harassment and willingness to use built-in reporting functions, respectively.

5.3. Post Hoc Analyses

5.3.1. Interaction Effects. To assess the nature of confidence in system anonymity, we conducted simple slope analyses following the guidelines suggested by Aiken et al. (1991). We plotted the significant interactions one standard deviation above and below the mean for the confidence in system anonymity. Online Appendix I shows the conditional effects of the moderator. Figures 3 and 4 show the interaction plots. For the interaction between the presence of others and confidence in system anonymity, we observed a weaker negative relationship between presence of others and willingness to use built-in reporting functions when confidence in system anonymity was perceived to be high ($\beta = -0.27$, $p < 0.001$), and a stronger negative relationship when confidence in system anonymity was perceived to be low ($\beta = -0.76$, $p < 0.001$). Furthermore, we found a moderate negative relationship between the presence of others and perceived responsibility to report when confidence in system anonymity was perceived to be low ($\beta = -0.49$, $p < 0.001$). These results imply that, compared with bystanders with low confidence in system anonymity, those with high confidence in system anonymity were more likely to accept the responsibility to report and more willing to use built-in reporting functions when they witnessed social media harassment, despite their perception of the presence of others. The results, therefore, confirm that confidence in reporting system anonymity reduces the negative effect of the presence of others on reporting social media harassment.

5.3.2. Mediation Effects. We used bootstrapping to conduct post hoc tests of mediation (Hayes 2009, Lowry et al. 2016b, Vance et al. 2015). In the bootstrapping process, we resampled with replacement from the obtained sample 5,000 times (Hayes 2009) and specified a 95% confidence interval (CI). Figure J-1 in Online Appendix J depicts the mediation relationships for our study. We examined (1) the effects of perceived emergency and the presence of others on perceived responsibility to report (a_1 and a_2), as well as the effect of perceived reporting justice on the perceived outcome effectiveness of reporting (a_3); (2) the effects of the perceived responsibility to report and perceived outcome effectiveness of reporting on willingness to use built-in reporting functions (b_1 and b_2); and (3) the effects of the assessments (i.e., perceived emergency, presence of others, and perceived reporting justice) on willingness to use built-in reporting functions (c_1'–c_3'). When the CIs of

Figure 3. Interaction of Presence of Others and Confidence in System Anonymity on Willingness to Use Built-In Reporting Functions

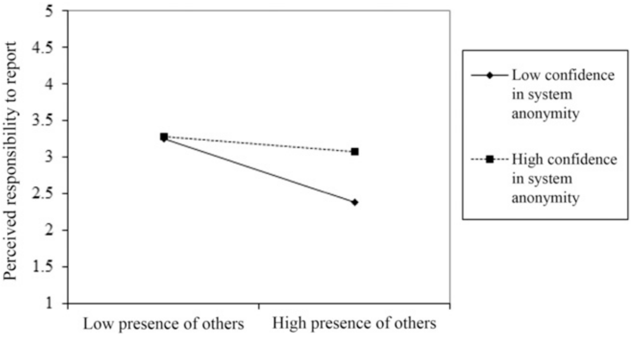
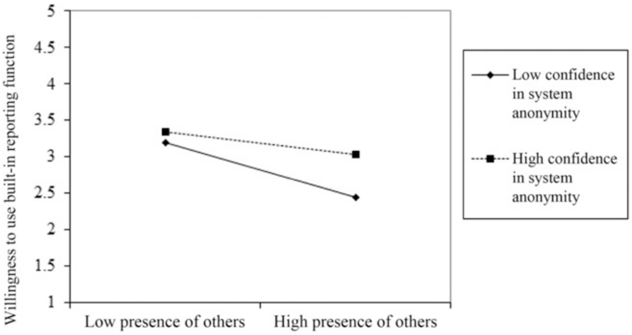


Figure 4. Interaction of Presence of Others and Confidence in System Anonymity on Perceived Responsibility to Report



the indirect effects (i.e., *ab*) did not include zero between the upper and lower bounds and the CIs of the direct effects (i.e., *c′*) did, full mediation was indicated; on the other hand, if the CIs of neither the indirect effects (i.e., *ab*) nor the direct effects (i.e., *c′*) included zero between the upper and lower bounds, partial mediation was indicated.

We found partial mediation. Table J-1 in Online Appendix J summarizes the results of the mediation test. The effects of perceived emergency and presence of others on bystanders' willingness to use built-in reporting functions were partially mediated by perceived responsibility to report. The effects of perceived reporting justice were partially mediated by the perceived outcome effectiveness of reporting. In other words, in addition to exerting direct impact on bystanders' willingness to use built-in reporting functions, perceived emergency, presence of others, and perceived reporting justice also indirectly influenced willingness to report through influencing bystanders' perceived responsibility to report and perceived outcome effectiveness of reporting, respectively.

6. Discussion and Implications

6.1. Discussion of Results

This study was motivated by a desire to understand bystanders' willingness to use built-in reporting functions on social media platforms. To explore this, we drew on the bystander intervention framework and reporting literature to develop a contextualized research model for the social media bystander reporting intervention. Adopting a sociotechnical perspective, we considered the distinctive elements of technology and social factors relevant to social media bystanders' willingness to report incidents. The contextualized social media constructs (i.e., perceived emergency, perceived responsibility to report, perceived self-efficacy to report, perceived outcome effectiveness of reporting, and presence of others) had significant effects on bystanders' willingness to use built-in reporting functions. In addition, our study's findings provide empirical evidence of the importance of considering the social and technical components in bystander reporting interventions on social media platforms and also demonstrate the interplay of the characteristics of built-in reporting functions and the online social environment—namely, how confidence in system anonymity of the reporting system counterbalances the negative influence of the presence of others, a frequently cited reason for bystanders not reporting harassment. Contrary to our expectations, perceived reporting justice did not have a significant direct effect on bystanders' willingness to use built-in reporting functions. It was only in terms of perceived outcome effectiveness of reporting that justice in the regulatory structures mattered for reporting interventions on social media platforms. We thus infer that perceived reporting justice is an essential criterion for evaluating the outcome effectiveness of reporting, which influences bystanders' reporting interventions.

One interesting caveat worth noting is that perceived reporting climate did not influence perceived outcome effectiveness of reporting. One of the possible reasons for this is that the reporting climate empowers bystanders and builds their social confidence in the reporting intervention, which may thus encourage them to intervene through using built-in reporting functions on social media platforms directly. We therefore found a positive effect of perceived reporting climate on bystanders' willingness to use built-in reporting functions. Prior whistleblowing studies also demonstrate that managerial practices and regulatory structures do not always have the same strength of relationship to the effectiveness of the outcome. For example, Perry (1993) found that the effectiveness of whistleblowing was not related to the organization's climate in terms of discouraging dissent. In the following sections, we discuss the implications for research and practice, limitations, and avenues for future research.

6.2. Implications for Research

Our work represents one of the first academic studies, if not the first, examining bystanders' reporting behavior in response to social media harassment. Understanding why social media bystanders are willing to intervene is critical, as it is considered to be one of the most effective means of curbing social media harassment (Wong-Lo and Bullock 2014). Furthermore, bystanders' responses to social media harassment can influence how the harassment incident unfolds (Leung et al. 2018). Given that many social media platform owners have implemented built-in reporting functions for reporting social media harassment, this study complements practice by offering a theoretical understanding of why bystanders use reporting tools offered by social media platforms.

To explain the willingness to report, we drew on the bystander intervention and reporting literature. We mapped the core concepts of the bystander intervention framework and reporting literature to constructs in our research model. In contrast to established reporting frameworks (which typically apply to situations in which the harassing act and the reporting act occur at different locations), our social media bystander reporting intervention framework accounts for harassing acts and reporting acts that take place on the same social media platform. Specifically, we used a sociotechnical perspective to identify distinctive elements of the social media context that make bystanders' reporting of social media harassment different from face-to-face reporting. Our work directs attention to how the interplay of the online social environment and characteristics of technology (Crawford and Gillespie 2016) shape bystanders' willingness to use built-in reporting functions. By delineating the distinctiveness of bystander reporting interventions on social media from face-to-face environments, our study not only advances the literature on reporting and harassment but underscores the value of using a sociotechnical approach to study how to identify and mitigate social media harassment (Sarker et al. 2019).

Our social media bystander reporting intervention framework identifies three distinctive elements: (1) online social media environments, (2) social media technology platforms with built-in reporting functions, and (3) the interplay of the characteristics of built-in reporting functions and the online social environment. In doing so, we draw on social media, cyberbullying, and reporting literature to examine how social media's characteristics shape bystander reporting behavior. We also explain how communication visibility of social media (via message transparency and network translucence) (Leonardi 2014, 2015) increases bystanders' enhanced awareness of the harassing incident, personal responsibility to report, and presence of others on social media platforms. To explain why bystanders use built-in reporting functions, we built upon appraisal theory (Folkman et al. 1986) and highlighted the importance of considering the perception of the reporting function (i.e., confidence in system anonymity), the managerial practices and regulatory structures of the platform owners (i.e., perceived reporting climate and perceived reporting justice), and bystanders' assessment of their usage of the reporting function (i.e., perceived self-efficacy to report and perceived outcome effectiveness of reporting). Our research underscores the tension between technological and social solutions to social media harassment. Although our work emphasizes the importance of technological remedies or the perception of them, it does not discount that paying attention to social elements is important to further the fight against social media harassment. Empirical evidence of the interplay of the characteristics of built-in reporting functions and the online social environment provides further support to the sociotechnical perspectives offered by Sarker et al. (2019). Through the social media bystander reporting intervention framework's application and further investigation of these distinctive elements, we believe researchers could shed light on how to more effectively encourage social media bystanders to intervene on behalf of victims of social media harassment, cyberstalking, cyberimpersonation, and cybertrolling.

6.3. Implications for Practice

The findings of this study have important implications for practitioners, including platform owners, schools, organizations, and government agencies. Platform owners often invest in online social platforms with the goal of building social relationships among individuals who share similar interests, activities, backgrounds, or real-life connections. Specific guidelines are discussed below.

First, our findings show that the perceived emergency associated with a social media harassment incident influences bystanders' willingness to use a built-in reporting function to report the incident to platform owners. To increase bystanders' awareness of the emergency of social media harassment incidents, platform owners could use machine learning techniques to detect harassment language on social media platforms, classify posts into benign or hurtful categories, and add automated alerts for negative harassment posts. Platform owners could also experiment with showing other bystanders (anonymously) that a user has flagged a post or message as offensive or harassing as a means of evoking powerful social forces to be well behaved. Such design features may also draw bystanders' attention to the emergency of the harassing post and motivate them to report social media harassment.

Second, perceived responsibility predicts bystanders' willingness to report. The concept of responsibility implies that bystanders take a more active role in supporting and protecting their communities' interests and that they feel more broadly accountable to their communities for their actions. To promote the development of social and moral responsibility, campaigns and training programs (e.g., on Internet etiquette, advanced moral development, and acceptable online behavior) should be developed or implemented in school curricula and in other public forums, which could help mitigate the negative consequences of social media harassment. For example, to advance bystanders' sense of responsibility as well as their sense of personal efficacy in using built-in reporting functions, platform owners could design lively and interactive "take action" modules that educate users on how to recognize and differentiate social media harassment incidents from acceptable online posting and sharing behaviors, introduce the anonymous reporting system mechanism (e.g., by instructing users on how to submit or complete reports by using built-in reporting functions), and provide samples demonstrating how platform owners handle bystander reports.

Third, bystanders' efficacy beliefs could potentially be engendered through the effective design of online reporting systems and training on how to use them. Whereas we know that efficacy beliefs in technology translate to its use (McKnight et al. 2011), we know much less about how to design reporting systems that encourage such personal efficacy. For practice, this suggests a need for sandboxing and experimenting with ways to design reporting functions. For example, it would be useful to examine whether the placement and prominence of reporting functions shape bystanders' perceptions of their personal efficacy. Platform owners could potentially increase bystanders' efficacy beliefs through improving the reliability, dependability (Tams et al. 2018), and quality of the user interface of the reporting function. Moreover, platform owners could emphasize the efficacy of reporting tools and focus on directing attention toward both enactive mastery and mindfulness. An extensive body of work underscores that enactive mastery and vicarious learning—that is, watching others perform tasks—encourages users to perform new tasks on computers (Compeau and Higgins 1995). Also, a growing body of work underscores the importance of mindfulness—that is, attentiveness to the context—as a driver of value-added technology use (Sun and Fang 2016, Thatcher et al. 2018). Bystander intervention training programs that provide users with illustrations of the effective use of response tools and that underscore the context for when to use them could increase bystanders' willingness to use reporting functions.

Fourth, platform owners should establish clear and fair standards for bystander reporting and handling harassment on social media platforms and make such standards easily accessible, as bystanders may need guidance on what reporting procedures should be followed and what actions are expected of platform owners after reporting. Moreover, platform owners should actively foster a supportive reporting climate and instill greater transparency in the review process. For instance, the platform's review team could update bystanders who report social media harassment on the progress and results of the investigation. Such information sharing could increase bystanders' beliefs about the efficacy of built-in reporting functions and, consequently, enhance their willingness to report future social media harassment incidents.

Fifth, the results of our study show that the presence of others inhibits bystanders' reporting of social media harassment. To reduce the influence of bystanders' perception of the presence of others, platform owners should consider mechanisms that empower bystanders to use reporting tools. The social media bystander reporting intervention model suggests that platform designers should focus attention on reinforcing positive perceptions and outcomes of bystander reporting interventions as a means of reducing evaluation apprehension, thus mitigating users' tendency to diffuse responsibility by underscoring bystanders' responsibility to report social media harassment, and reducing pluralistic ignorance by publicizing stories recounting how bystanders' helping behavior can reduce or mitigate the impact of social media harassment.

Finally, our study directs attention to the need for bystanders' confidence in reporting system anonymity. Platform owners should raise awareness of the availability of the reporting tool and focus on designing tools that positively influence the subdimensions of confidence in system anonymity—dissociative anonymity and visual anonymity—in order to avoid potential social impact or retaliation threats within bystanders' social circles. Deidentification in online reporting could be designed by removing cues, prompts, and any contextual information as a means of maintaining user anonymity in the reports sent to victims and perpetrators. As the findings of our study suggest, enhanced confidence in system anonymity may counterbalance the negative effect of the presence of others on bystander intervention.

6.4. Limitations and Future Research

Our research has a few limitations. First, our study offers a general test of a contextualized model of bystander intervention on a social media platform. Future research could consider the effects of types of incidents (e.g., purposeful embarrassment, threatening events, or sexual harassment) and levels of emergency (e.g., high, medium, and low) on bystanders' reporting interventions. In addition, other situational factors, such as the number and characteristics of individuals in a harassment incident (e.g., who started harassing whom, how many people joined in on the harassment), need to be examined to assess whether they encourage positive or negative responses from bystanders (Darley and Latané 1968, Latané and Darley 1970).

Second, our study assesses the effects of a limited number of social and technical components on bystanders' willingness to use built-in reporting functions on social media platforms. Future research could explore additional components as well as their interplays. The need for such research is supported by our study, which indicates that sociotechnical factors contribute to social media bystanders' willingness to report harassment to the platform. Future research in this area should delve into additional social factors (e.g., collective empowerment) and technical factors (e.g., synchronicity of the reporting function) that enhance the understanding of how bystanders assess the acuity of incidents, trust in the platform owners, and willingness to engage in additional behaviors to ameliorate the impact of social media harassment.

Third, our work on social media bystander reporting opens the door to at least five streams of work on online reporting functions: (1) Additional research is needed to validate our findings in the field across various social media platforms. It would be interesting to see whether our model is robust across platforms with varying levels of message transparency and network translucence. (2) Future research could investigate the critical factors influencing bystanders' decisions to report/not report social media harassment. (3) Future research could explore the relationship between bystander reporting interventions (e.g., no intention to report social media harassment) and negative bystander behaviors (e.g., joining in the social media harassment incidents) on social media platforms. (4) Future research could explore how other design features, such as instant responses, feedback from the review teams to users who reported the harassing post, and incentives to report, influence users' willingness to use built-in reporting functions. (5) Future work should also examine the interplay of online and offline reporting of social media harassment; for example, it would be interesting to examine how to create effective coordination mechanisms between platform owners and local authorities (e.g., government agencies, police, and counseling centers) in order to develop privacy-enabled referral policies and procedures for mitigating social media harassment. Thus, future research should consider looking into how the specific artifact design of the reporting system and the culture of the social media platform influence bystanders' reporting of social media harassment to social media authorities as well as offline authorities.

7. Conclusion

This study fills a gap in the understanding of social media harassment and, more specifically, in the social media harassment literature. It introduces the social media bystander reporting intervention framework and identifies factors relevant to understanding bystanders' willingness to report social media harassment through using built-in reporting functions on social media platforms. The results of our empirical analysis confirm that three distinctive elements—the online social environment, the technology platform, and their interplay—shape bystanders' willingness to report social media harassment. By enriching the understanding of the distinctive elements that shape bystanders' willingness to use built-in online reporting functions, our work sheds light on how to more effectively mitigate social media harassment and provides a foundation for future work on how to build safer communities on social media platforms.

Acknowledgments

The authors express their gratitude to the senior editor and associate editor for their motivating guidance, encouragement, and insightful direction in developing this manuscript. The authors thank three anonymous reviewers for their most helpful comments and suggestions during the review process. The authors greatly appreciate their feedback and expertise.

Endnotes

[1] Kowalski et al. (2008) categorized cyberbullying into six types: harassment, denigration, outing and trickery, exclusion, impersonation, and cyberstalking.

[2] An application programming interface is a computing interface which allows interactions between multiple software intermediaries. The hypertext transfer protocol–based API offers a means of getting data into and out of a social media platform and can be used by applications to programmatically query data on social media. Users with an access token can read any post (including status updates) on a social media platform. The availability of such applications means that social media reporting may not be entirely anonymous because users' profiles, social networks, and posts may be visible and accessible through the use of such applications.

References

Aiken LS, West SG, Reno RR (1991) *Multiple Regression: Testing and Interpreting Interactions* (Sage Publications, Thousand Oaks, CA).

Allison KR, Bussey K (2016) Cyber-bystanding in context: A review of the literature on witnesses' responses to cyberbullying. *Children Youth Services Rev.* 65(June):183–194.

Anti-Defamation League (2019) More than one-third of Americans experience severe online hate and harassment, new ADL study finds. Accessed May 19, 2019, https://www.adl.org/news/press-releases/more-than-one-third-of-americans-experience-severe-online-hate-and-harassment.

Bala H, Venkatesh V (2015) Adaptation to information technology: A holistic nomological network from implementation to job outcomes. *Management Sci.* 62(1):156–179.

Banyard VL (2011) Who will help prevent sexual violence: Creating an ecological model of bystander intervention. *Psych. Violence* 1(3):216–229.

Barchia K, Bussey K (2011) Individual and collective social cognitive influences on peer aggression: Exploring the contribution of aggression efficacy, moral disengagement, and collective efficacy. *Aggressive Behav.* 37(2):107–120.

Bostrom RP, Gupta S, Thomas D (2009) A meta-theory for understanding information systems within sociotechnical systems. *J. Management Inform. Systems* 26(1):17–48.

Boyd D (2010) Social network sites as networked publics: Affordances, dynamics, and implications. Papacharissi Z, ed. *A Networked Self: Identity, Community, and Culture on Social Network Sites* (Routledge, New York), 47–66.

Brody N, Vangelisti AL (2016) Bystander intervention in cyberbullying. *Comm. Monographs* 83(1):94–119.

Burn SM (2009) A situational model of sexual assault prevention through bystander intervention. *Sex Roles* 60(11–12):779–792.

Calvete E, Orue I, Estévez A, Villardón L, Padilla P (2010) Cyberbullying in adolescents: Modalities and aggressors' profile. *Comput. Human Behav.* 26(5):1128–1135.

Carmines EG, Zeller RA (1979) *Reliability and Validity Assessment*, vol. 17 (Sage, Thousand Oaks, CA).

Cassidy W, Faucher C, Jackson M (2013) Cyberbullying among youth: A comprehensive review of current international research and its implications and application to policy and practice. *School Psych. Internat.* 34(6):575–612.

Cenfetelli RT, Bassellier G (2009) Interpretation of formative measurement in information systems research. *MIS Quart.* 33(4):689–708.

Chan TCH, Cheung CMK, Wong RYM (2019) Cyberbullying on social networking sites: The crime opportunity and affordance perspectives. *J. Management Inform. Systems* 36(2):574–609.

Chan TCH, Cheung CMK, Lee ZWY (2021) Cyberbullying on social networking sites: A literature review and future research directions. *Inform. Management* 58(2):103411.

Compeau DR, Higgins CA (1995) Application of social cognitive theory to training for computer skills. *Inform. Systems Res.* 6(2):118–143.

Crawford K, Gillespie T (2016) What is a flag for? Social media reporting tools and the vocabulary of complaint. *New Media Soc.* 18(3):410–428.

Darley JM, Latané B (1968) Bystander intervention in emergencies: Diffusion of responsibility. *J. Personality Soc. Psych.* 8(4):377–383.

DeSmet A, Bastiaensens S, Van Cleemput K, Poels K, Vandebosch H, Cardon G, De Bourdeaudhuij I (2016) Deciding whether to look after them, to like it, or leave it: A multidimensional analysis of predictors of positive and negative bystander behavior in cyberbullying among adolescents. *Comput. Human Behav.* 57(April): 398–415.

Diamantopoulos A, Siguaw JA (2006) Formative vs. reflective indicators in organizational measure development: A comparison and empirical illustration. *British J. Management* 17(4):263–282.

Diamantopoulos A, Winklhofer HM (2001) Index construction with formative indicators: An alternative to scale development. *J. Marketing Res.* 38(2):269–277.

Dillon KP, Bushman BJ (2015) Unresponsive or un-noticed?: Cyberbystander intervention in an experimental cyberbullying context. *Comput. Human Behav.* 45(April):144–150.

Dovidio JF, Piliavin JA, Schroeder DA, Penner L (2006) *The Social Psychology of Prosocial Behavior* (Psychology Press, New York).

Dozier JB, Miceli MP (1985) Potential predictors of whistle-blowing: A prosocial behavior perspective. *Acad. Management Rev.* 10(4): 823–836.

Duggan M (2017) Online harassment 2017. *Pew Research Center* (July 11), http://www.pewinternet.org/2017/07/11/online-harassment-2017/.

Edwards MS, Lawrence SA, Ashkanasy NM (2013) The role of perceptions, appraisals and anticipated emotions in shaping reporting behavior in response to wrongdoing. Burke RJ, Cooper CL, eds. *Voice and Whistleblowing in Organizations: Overcoming Fear, Fostering Courage and Unleashing Candour* (Edward Elgar Publishing, Northampton, MA), 254–278.

eMarketer (2019) Level of privacy/security concern toward select social networks among US internet users, March 2019. *eMarketer* (April 29), https://www.emarketer.com/Chart/Level-of-PrivacySecurity-Concern-Toward-Select-Social-Networks-Among-US-Internet-Users-March-2019-of-respondents/228185.

Facebook (2011) Creating a culture of respect. (March 10), https://www.facebook.com/notes/facebook-safety/creating-a-culture-of-respect/196123070408483/.

Facebook (2019) Facebook community standards enforcement report. Accessed May 19, 2019, https://transparency.facebook.com/community-standards-enforcement#bullying-and-harassment.

Finkelhor D, Wolak J (2003) Reporting assaults against juveniles to the police barriers and catalysts. *J. Interpersonal Violence* 18(2): 103–128.

Finn J (2004) A survey of online harassment at a university campus. *J. Interpersonal Violence.* 19(4):468–483.

Fischer P, Krueger JI, Greitemeyer T, Vogrincic C, Kastenmüller A, Frey D, Heene M, Wicher M, Kainbacher M (2011) The bystander-effect: A meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psych. Bull.* 137(4):517–537.

Folkman S, Lazarus RS, Dunkel-Schetter C, DeLongis A, Gruen RJ (1986) Dynamics of a stressful encounter: Cognitive appraisal, coping, and encounter outcomes. *J. Personality Soc. Psych.* 50(5):992–1003.

Fornell C, Larcker DF (1981) Evaluating structural equation models with unobservable variables and measurement error. *J. Marketing Res.* 18(3):39–50.

Fowler F Jr, Cosenza C (2009) Design and evaluation of survey questions. Bickman L, Rog DJ, eds. *The SAGE Handbook of Applied Social Research Methods*, 2nd ed. (Sage Publications, Thousand Oaks, CA), 375–412.

Gini G, Albiero P, Benelli B, Altoe G (2008) Determinants of adolescents' active defending and passive bystanding behavior in bullying. *J. Adolescence* 31(1):93–105.

GlobalWebIndex (2018) Social: Flagship report 2018. Accessed May 19, https://www.globalwebindex.com/hubfs/Downloads/Social-H2-2018-report.pdf.

Gracia E, Garcia F, Lila M (2008) Police involvement in cases of intimate partner violence against women: The influence of perceived severity and personal responsibility. *Violence Against Women* 14(6):697–714.

Griffiths S (2019) Can this technology put an end to bullying? *BBC* (February 7), http://www.bbc.com/future/story/20190207-how-artificial-intelligence-can-help-stop-bullying.

Harris R (2017) Detecting cyber bullying: But can it be stopped? *WVTF/Radio IQ* (March 2), https://www.wvtf.org/post/detecting-cyber-bullying-can-it-be-stopped#stream/0.

Hayes AF (2009) Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Comm. Monographs* 76(4): 408–420.

Hinduja S, Patchin JW (2010) Bullying, cyberbullying, and suicide. *Arch. Suicide Res.* 14(3):206–221.

Hoefnagels C, Zwikker M (2001) The bystander dilemma and child abuse: Extending the Latane and Darley model to domestic violence. *J. Appl. Soc. Psych.* 31(6):1158–1183.

Hu L, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria vs. new alternatives. *Structural Equation Model.* 6(1):1–55.

Huang Y, Chou C (2010) An analysis of multiple factors of cyberbullying among junior high school students in Taiwan. *Comput. Human Behav.* 26(6):1581–1590.

Jones LM, Mitchell KJ, Finkelhor D (2013) Online harassment in context: Trends from three Youth Internet Safety Surveys (2000, 2005, 2010). *Psych. Violence* 3(1):53–69.

Jones LM, Mitchell KJ, Turner HA (2015) Victim reports of bystander reactions to in-person and online peer harassment: A national survey of adolescents. *J. Youth Adolescence* 44(12):2308–2320.

Karahanna E, Xu SX, Xu Y, Zhang NA (2018) The needs–affordances–features perspective for the use of social media. *MIS Quart.* 42(3):737–756.

Keil M, Tiwana A, Sainsbury R, Sneha S (2010) Toward a theory of whistleblowing intentions: A benefit-to-cost differential perspective. *Decision Sci.* 41(4):787–812.

Kim SJ, Hancock JT (2015) Optimistic bias and Facebook use: Self–other discrepancies about potential risks and benefits of Facebook use. *Cyberpsych. Behav. Soc. Networks* 18(4):214–220.

Kirkman BL, Chen G, Farh JL, Chen ZX, Lowe KB (2009) Individual power distance orientation and follower reactions to transformational leaders: A cross-level, cross-cultural examination. *Acad. Management J.* 52(4):744–764.

Kiss J (2010) Facebook announces new safety measures but no panic button. *Guardian* (April 13), https://www.theguardian.com/technology/2010/apr/13/facebook-safety.

Koedinger KR, Aleven V (2007) Exploring the assistance dilemma in experiments with cognitive tutors. *Ed. Psych. Rev.* 19(3):239–264.

Kowalski RM, Limber SP, Agatston PW (2008) *Cyberbullying: Bullying in the Digital Age* (Blackwell, Malden, MA).

Krasnova H, Widjaja T, Buxmann P, Wenninger H, Benbasat I (2015) Research note—Why following friends can hurt you: An exploratory investigation of the effects of envy on social networking sites among college-age users. *Inform. Systems Res.* 26(3):585–605.

La Greca AM, Lopez N (1998) Social anxiety among adolescents: Linkages with peer relations and friendships. *J. Abnormal Child Psych.* 26(2):83–94.

Laible D, Eye J, Carlo G (2008) Dimensions of conscience in mid-adolescence: Links with social behavior, parenting, and temperament. *J. Youth Adolescence* 37(7):875–887.

Latané B, Darley JM (1968) Group inhibition of bystander intervention in emergencies. *J. Personality Soc. Psych.* 10(3):215–221.

Latané B, Darley JM (1970) *The Unresponsive Bystander: Why Doesn't He Help?* (Appleton-Century-Crofts, New York).

Latané B, Nida S (1981) Ten years of research on group size and helping. *Psych. Bull.* 89(2):308–324.

Lee AS (2004) *Thinking About Social Theory and Philosophy for Information Systems*, Mingers J, Willcocks L, eds. (John Wiley and Sons, England), 1–26.

Lee Y, Larsen KR (2009) Threat or coping appraisal: Determinants of SMB executives' decision to adopt anti-malware software. *Eur. J. Inform. Systems* 18(2):177–187.

Leonardi PM (2014) Social media, knowledge sharing, and innovation: Toward a theory of communication visibility. *Inform. Systems Res.* 25(4):796–816.

Leonardi PM (2015) Ambient awareness and knowledge acquisition: Using social media to learn" who knows what" and" who knows whom. *MIS Quart.* 39(4):747–762.

Leonardi PM, Vaast E (2017) Social media and their affordances for organizing: A review and agenda for research. *Acad. Management Ann.* 11(1):150–188.

Leung AN, Wong N, Farver JM (2018) You are what you read: The belief systems of cyber-bystanders on social networking sites. *Frontiers Psych.* 9(April):1–11. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00365/full.

Lewis D (2011) Whistleblowing in a changing legal climate: Is it time to revisit our approach to trust and loyalty at the workplace? *Bus. Ethics Eur. Rev.* 20(1):71–87.

Liang H, Xue Y (2010) Understanding security behaviors in personal computer usage: A threat avoidance Perspective. *J. Assoc. Inform. Systems* 11(7):394–413.

Lindsay M, Booth JM, Messing JT, Thaller J (2016) Experiences of online harassment among emerging adults emotional reactions and the mediating role of fear. *J. Interpersonal Violence* 31(19):3174–3195.

Loewenstein G, Small DA (2007) The Scarecrow and the Tin Man: The vicissitudes of human sympathy and caring. *Rev. General Psych.* 11(2):112–126.

Lowry PB, D'Arcy J, Hammer B, Moody GD (2016a) "Cargo Cult" science in traditional organization and information systems survey research: A case for using nontraditional methods of data collection, including Mechanical Turk and online panels. *J. Strategic Inform. Systems* 25(3):232–240.

Lowry PB, Moody GD, Galletta DF, Vance A (2013) The drivers in the use of online whistle-blowing reporting systems. *J. Management Inform. Systems* 30(1):153–190.

Lowry PB, Zhang J, Wang C, Siponen M (2016b) Why do adults engage in cyberbullying on social media? An integration of online disinhibition and deindividuation effects with the social structure and social learning model. *Inform. Systems Res.* 27(4):962–986.

Macháčková H, Dedkova L, Sevcikova A, Cerna A (2013) Bystanders' support of cyberbullied schoolmates. *J. Community Appl. Soc. Psych.* 23(1):25–36.

MacKenzie SB, Podsakoff PM, Podsakoff NP (2011) Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quart.* 35(2):293–334.

Manstead ASR, Fischer AH (2001) Social appraisal: The social world as object of and influence on appraisal processes. Scherer KR, Schorr A, Johnstone T, eds. *Appraisal Processes in Emotion: Theory, Methods, Research* (Oxford University Press, New York), 221–232.

Mason W, Suri S (2012) Conducting behavioral research on Amazon's Mechanical Turk. *Behav. Res. Methods* 44(1):1–23.

Matook S, Cummings J, Bala H (2015) Are you feeling lonely? The impact of relationship characteristics and online social network features on loneliness. *J. Management Inform. Systems* 31(4):278–310.

McFarland LA, Ployhart RE (2015) Social media: A contextual framework to guide research and practice. *J. Appl. Psych.* 100(6):1653–1677.

McKnight DH, Carter M, Thatcher JB, Clay PF (2011) Trust in a specific technology: An investigation of its components and measures. *ACM Trans. Inform. Systems* 2(2):1–25.

Miceli MP, Near JP (1985) Characteristics of organizational climate and perceived wrongdoing associated with whistle-blowing decisions. *Personnel Psych.* 38(3):525–544.

Miceli MP, Near JP (1992) *Blowing the Whistle: The Organizational and Legal Implications for Companies and Employees* (Lexington Books, New York).

Mishna F, Khoury-Kassabri M, Gadalla T, Daciuk J (2012) Risk factors for involvement in cyber bullying: Victims, bullies and bully–victims. *Children Youth Services Rev.* 34(1):63–70.

Moore MJ, Nakano T, Enomoto A, Suda T (2012) Anonymity and roles associated with aggressive posts in an online forum. *Comput. Human Behav.* 28(3):861–867.

Near JP, Miceli MP (1995) Effective-whistle blowing. *Acad. Management Rev.* 20(3):679–708.

Nickerson AB, Aloe AM, Livingston JA, Feeley TH (2014) Measurement of the bystander intervention model for bullying and sexual harassment. *J. Adolescence* 37(4):391–400.

Obermaier M, Fawzi N, Koch T (2016) Bystanding or standing by? How the number of bystanders affects the intention to intervene in cyberbullying. *New Media Soc.* 18(8):1491–1507.

Park C, Im G, Keil M (2008) Overcoming the mum effect in IT project reporting: Impacts of fault responsibility and time urgency. *J. Assoc. Inform. Systems* 9(7):409–431.

Park C, Keil M (2009) Organizational silence and whistle-blowing on IT projects: An integrated model. *Decision Sci.* 40(4):901–918.

Paul S, Smith PK, Blumberg HH (2012) Comparing student perceptions of coping strategies and school interventions in managing bullying and cyberbullying incidents. *Pastoral Care Ed.* 30(2):127–146.

Perry JL (1993) *Whistleblowing, Organizational Performance, and Organizational Control* (M. E. Sharpe, Armonk, NY).

Petter S, Straub D, Rai A (2007) Specifying formative constructs in information systems research. *MIS Quart.* 31(4):623–656.

Plan International UK (2017) Almost half of girls aged 11–18 have experienced harassment or bullying online. Accessed August 18, https://plan-uk.org/media-centre/almost-half-of-girls-aged-11-18-have-experienced-harassment-or-bullying-online.

Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP (2003) Common method biases in behavioral research: A critical review of the literature and recommended remedies. *J. Appl. Psych.* 88(5): 879–903.

Polites GL, Roberts N, Thatcher J (2012) Conceptualizing models using multidimensional constructs: A review and guidelines for their use. *Eur. J. Inform. Systems* 21(1):22–48.

Pöyhönen V, Juvonen J, Salmivalli C (2012) Standing up for the victim, siding with the bully or standing by? Bystander responses in bullying situations. *Soc. Development* 21(4):722–741.

Pozzoli T, Gini G (2013a) Friend similarity in attitudes toward bullying and sense of responsibility to intervene. *Soc. Influence* 8(2–3):161–176.

Pozzoli T, Gini G (2013b) Why do bystanders of bullying help or not? A multidimensional model. *J. Early Adolescence* 33(3): 315–340.

Prentice DA, Miller DT (1993) Pluralistic ignorance and alcohol use on campus: Some consequences of misperceiving the social norm. *J. Personality Soc. Psych.* 64(2):243–256.

Price M, Dalgleish J (2010) Cyberbullying: Experiences, impacts and coping strategies as described by Australian young people. *Youth Stud. Australia* 29(2):51–59.

Quirk R, Campbell M (2015) On standby? A comparison of online and offline witnesses to bullying and their bystander behaviour. *Ed. Psych.* 35(4):430–448.

Roberts N, Thatcher J (2009) Conceptualizing and testing formative constructs: tutorial and annotated example. *ACM SIGMIS Database: DATABASE Adv. Inform. Systems* 40(3):9–39.

Runions K, Shapka JD, Dooley J, Modecki K (2013) Cyber-aggression and victimization and social information processing: Integrating the medium and the message. *Psych. Violence* 3(1):9–26.

Salmivalli C (2010) Bullying and the peer group: A review. *Aggression Violent Behav.* 15(2):112–120.

Sarker S, Chatterjee S, Xiao X, Elbanna A (2019) The sociotechnical axis of cohesion for the IS discipline: Its historical legacy and its continued relevance. *MIS Quart.* 43(3):695–719.

Schacter HL, Greenberg S, Juvonen J (2016) Who's to blame?: The effects of victim disclosure on bystander reactions to cyberbullying. *Comput. Human Behav.* 57(April):115–121.

Schwarz A, Rizzuto T, Carraher-Wolverton C, Roldán JL, Barrera-Barrera R (2017) Examining the impact and detection of the urban legend of common method bias. *ACM SIGMIS Database: DATABASE Adv. Inform. Systems* 48(1):93–119.

Shultz E, Heilman R, Hart KJ (2014) Cyber-bullying: An exploration of bystander behavior and motivation. *Cyberpsych.: J. Psychosocial Res. Cyberspace* 8(4):53–70.

Skinner WF, Fream AM (1997) A social learning theory analysis of computer crime among college students. *J. Res. Crime Delinquency* 34(4):495–518.

Slonje R, Smith PK, Frisén A (2013) The nature of cyberbullying, and strategies for prevention. *Comput. Human Behav.* 29(1):26–32.

Smith HJ, Keil M (2003) The reluctance to report bad news on troubled software projects: A theoretical model. *Inform. Systems J.* 13(1):69–95.

Song J, Oh I (2018) Factors influencing bystanders' behavioral reactions in cyberbullying situations. *Comput. Human Behav.* 78-(January):273–282.

Suler J (2004) The online disinhibition effect. *Cyberpsych. Behav.* 7(3):321–326.

Sulkowski ML (2011) An investigation of students' willingness to report threats of violence in campus communities. *Psych. Violence* 1(1):53–65.

Sumter SR, Baumgartner SE, Valkenburg PM, Peter J (2012) Developmental trajectories of peer victimization: Off-line and online experiences during adolescence. *J. Adolescent Health* 50(6): 607–613.

Sun H, Fang Y (2016) Choosing a fit technology: Understanding mindfulness in technology adoption and continuance. *J. Assoc. Inform. Systems* 17(6):377–411.

Tams S, Thatcher JB, Craig K (2018) How and why trust matters in post-adoptive usage: The mediating roles of internal and external self-efficacy. *J. Strategic Inform. Systems* 27(2):170–190.

Tarling R, Morris K (2010) Reporting crime to the police. *British J. Criminology* 50(3):474–490.

Tavakoli AA, Keenan JP, Cranjak-Karanovic B (2003) Culture and whistleblowing an empirical study of Croatian and United States managers utilizing Hofstede's cultural dimensions. *J. Bus. Ethics* 43(1–2):49–64.

Thatcher J, Wright R, Sun H, Zagenczyk T, Klein R (2018) Mindfulness in information technology use: Definitions, distinctions, and a new measure. *MIS Quart.* 42(3):831–847.

The Children's Society, YoungMinds (2018) Safety net: The impact of cyberbullying on children and young people's mental health. Accessed February 1, https://youngminds.org.uk/media/2189/pcr144b_social_media_cyberbullying_inquiry_full_report.pdf.

Thornberg R (2007) A classmate in distress: Schoolchildren as bystanders and their reasons for how they act. *Soc. Psych. Ed.* 10(1):5–28.

Tu Z, Turel O, Yuan Y, Archer N (2015) Learning to cope with information security risks regarding mobile device loss or theft: An empirical examination. *Inform. Management* 52(4):506–517.

Turel O, Matt C, Trenz M, Cheung CM, D'Arcy J, Qahri-Saremi H, Tarafdar M (2019) Panel report: The dark side of the digitization of the individual. *Internet Res.* 29(2):274–288.

Twemlow SW, Fonagy P, Sacco FC (2004) The role of the bystander in the social architecture of bullying and violence in schools and communities. *Ann. N. Y. Acad. Sci.* 1036(1):215–232.

Van Cleemput K, Vandebosch H, Pabian S (2014) Personal characteristics and contextual factors that determine "helping," "joining in," and "doing nothing" when witnessing cyberbullying. *Aggressive Behav.* 40(5):383–396.

Van Royen K, Poels K, Vandebosch H, Adam P (2017) "Thinking before posting?" Reducing cyber harassment on social networking sites through a reflective message. *Comput. Human Behav.* 66(January):345–352.

Vance A, Lowry P, Eggett D (2015) A new approach to the problem of access policy violations: Increasing perceptions of accountability through the user interface. *MIS Quart.* 39(2):345–366.

Veletsianos G, Houlden S, Hodson J, Gosse C (2018) Women scholars' experiences with online harassment and abuse: Self-protection, resistance, acceptance, and self-blame. *New Media Soc.* 20(12): 4689–4708.

Ward MK, Broniarczyk SM (2011) It's not me, it's you: How gift giving creates giver identity threat as a function of social closeness. *J. Consumer Res.* 38(1):164–181.

Wei X, Zhang ZX, Chen XP (2015) I will speak up if my voice is socially desirable: A moderated mediating process of promotive vs. prohibitive voice. *J. Appl. Psych.* 100(5):1641–1652.

Wenninger H, Krasnova H, Buxmann P (2019) Understanding the role of social networking sites in the subjective well-being of users: A diary study. *Eur. J. Inform. Systems* 28(2):126–148.

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

Wong RYM, Cheung CMK, Xiao B (2018) Does gender matter in cyberbullying perpetration? An empirical investigation. *Comput. Human Behav.* 79(February):247–257.

Wong-Lo M, Bullock LM (2014) Digital metamorphosis: Examination of the bystander culture in cyberbullying. *Aggression Violent Behav.* 19(4):418–422.

Wright RT, Campbell DE, Thatcher JB, Roberts NH (2012) Operationalizing multidimensional constructs in structural equation modeling: Recommendations for IS research. *Comm. Assoc. Inform. Systems* 30(23):367–412.

Ybarra ML, Espelage DL, Mitchell KJ (2007) The co-occurrence of Internet harassment and unwanted sexual solicitation victimization and perpetration: Associations with psychosocial indicators. *J. Adolescent Health* 41(6):S31–S41.

Zhao WX, Jiang J, Weng J, He J, Lim EP, Yan H, Li X (2011) Comparing twitter and traditional media using topic models. Clough P, Foley C, Gurrin C, Jones G, Kraaij W, Lee H, Murdock V, eds. *Advances in Information Retrieval* (Springer, Berlin), 338–349.