# Bystanders Join In Cyberbullying on Social Networking Sites: The Deindividuation and Moral Disengagement Perspectives

**Tommy K. H. Chan**
Alliance Manchester Business School, The University of Manchester
Booth Street West, Manchester, M15 6PB, United Kingdom
tommy.chan@manchester.ac.uk

**Christy M. K. Cheung***
School of Business, Hong Kong Baptist University
34 Renfrew Road, Kowloon Tong, Hong Kong
ccheung@hkbu.edu.hk

**Izak Benbasat**
Sauder School of Business, The University of British Columbia
2053 Main Mall, Vancouver, BC V6T 1Z2, Canada
izak.benbasat@sauder.ubc.ca

**Bo Xiao**
Shidler College of Business, The University of Hawai‘i at Mānoa
2404 Maile Way, Honolulu, HI 96822, United States
boxiao@hawaii.edu

**Zach W. Y. Lee**
Durham University Business School, Durham University
Mill Hill Lane, Durham, DH1 3LB, United Kingdom
zach.lee@durham.ac.uk

*Corresponding author

[Accepted in **Information Systems Research** on 15th July 2022]

**Bystanders Join In Cyberbullying on Social Networking Sites: The Deindividuation and Moral Disengagement Perspectives**


## Abstract

Cyberbullying on social networking sites (SNSs) escalates when bystanders join in the bullying. Although researchers have recognized the devastating consequences of joining in cyberbullying behaviors, little is known about the role of information technology (IT) and its underlying mechanisms in fueling such negative group behavior on SNSs. To address this research gap, we develop and test an integrative model that explains bystanders' joining in cyberbullying behaviors on SNSs. Based on the theoretical premises of the *social identity model of deindividuation effects* (the SIDE model), we derive two *deindividuation experiences enabled by SNSs*, namely *experienced anonymity* and *experienced social identity*. We further use the *social network research framework* to gain insights into how IT features (i.e., digital profile, search and privacy, relational ties, and network transparency) enable these two deindividuation experiences. Considering the socially undesirable nature of joining in behaviors, we integrate the SIDE model with *moral disengagement theory* to explain how deindividuation experiences allow bystanders to bypass their psychological discomfort when engaging in such behaviors through the practice of moral disengagement mechanisms. Our research model is tested using a scenario survey, with two samples recruited from Amazon Mechanical Turk and Facebook. Our results support the influences of IT-enabled deindividuation experiences on bystanders' joining in cyberbullying and demonstrate the mediating effects of moral disengagement mechanisms in bridging the effects of deindividuation experiences on joining in behaviors. For researchers, the integrative view offers a conceptual bridge connecting IT features, deindividuation, moral disengagement, and negative online group behaviors on SNSs. For practitioners, our findings provide platform owners and governmental agencies with directions on how to mitigate cyberbullying on SNSs and other forms of deviant and undesirable online group behaviors.



***Keywords:*** online harms, cyberbullying, social networking sites (SNSs), bystanders, joining in, deindividuation, anonymity, social identity, moral disengagement

## 1. Introduction

As social networking sites (SNSs) have become an integral part of our personal and work lives, opportunities for and incidences of online harms have increased dramatically (Turel et al. 2019). *Cyberbullying on SNSs* refers to any form of aggressive, harmful behavior conducted by a group or an individual on an SNS repeatedly over time against targets who cannot easily defend themselves (Chan et al. 2021). Recent statistics show that 44% of the Internet users had experienced cyberbullying, 77% of which occurred on Facebook (ADL 2020). In particular, cyberbullying among individuals with purely virtual relations is increasing, with more SNS users being harassed by people they do not know offline. For example, in a 2018 survey, 53% of the respondents who stated that they had been harassed online reported not knowing their harasser in person (RAD 2018). Meanwhile, studies have identified a wide spectrum of negative consequences of cyberbullying-related victimization, which in extreme cases can lead to self-harming behaviors and suicide attempts (e.g., Longobardi et al. 2020).

Cyberbullying on SNSs typically involves three types of actors: perpetrator, victim, and bystanders. With the unbounded connectivity and public nature of SNSs, bystanders' joining in behaviors has become the invisible engine in the cycle of bullying. Bystanders are witnesses of cyberbullying incidences that they have not themselves perpetrated and by which they are not directly victimized (Twemlow et al. 2001). Bystanders' joining in behaviors in cyberbullying reinforces the abuse, exposes victims to a larger audience, and encourages further abuse by signaling their approval of the aggressive behavior. The recent tragedy of Hana Kimura, a cast member of the reality show *Terrace House*, has unfortunately morphed into an illustration of how bystanders' joining in behaviors aggravates cyberbullying on SNSs. Hateful and malicious comments disseminated via SNSs exposed Kimura's victimization to a larger audience and prolonged the abuse, leading to the victim's death (Nagumo and Imahashi 2020).

Cyberbullying on SNSs has attracted scholarly and public attention due to the devastating consequences of this behavior for individuals, society, and platform owners. Over the past decade, researchers from various disciplines (e.g., communication, education, psychology, computer science, and public health) have investigated cyberbullying on SNSs (e.g., Al-Garadi et al. 2019, Giumetti and

Kowalski 2022). Research on cyberbullying is also gaining attention in the information systems (IS) discipline (Chan et al. 2021), with a focus on perpetration (e.g., Lowry et al. 2016), victimization (e.g., Wright 2018), and bystanders' proactive reporting behavior (e.g., Wong et al. 2021). However, scant scholarly attention has been paid to bystanders' joining in behaviors in the IS literature. This is a critical omission because bystanders' joining in behaviors is a theoretically distinct phenomenon from cyberbullying perpetration. Although the behavioral manifestations of bystanders' joining in behaviors may appear similar to those of the perpetrators, a bystander who joins in cyberbullying does not initiate the perpetration and is usually driven by group pressure or expectation. More specifically, a bystander is not involved in creating the original cyberbullying post or identifying a target for perpetration; instead, the bystander only supports the perpetrators when the post appears on their newsfeed. Furthermore, while cyberbullying perpetration is primarily self-directed (i.e., perpetrators initiate abusive behaviors to pursue their agentic goals), bystanders' joining in behaviors is largely group-directed (i.e., bystanders join in the abuse to pursue their communal goals (Salmivalli 2010)). Therefore, existing findings concerning the antecedents (e.g., aggressive dispositions) and theoretical foundations (e.g., crime opportunity theory) of cyberbullying perpetration may not be appropriate for explaining bystanders' joining in behaviors.

*Bystanders' joining in cyberbullying on SNSs* refers to any behavior on an SNS by a bystander that builds upon the acts of the perpetrator and other aggressive bystanders and that intentionally or unintentionally hurts the victim or supports the bullies. Actions that manifest joining in cyberbullying on SNSs include, but are not limited to, giving a positive reaction (e.g., a Like) to a cyberbullying post or to other humiliating comments to support the bullies, sharing the post with other SNS users to hurt the victim, and leaving comments on the cyberbullying post that endorse the acts of the bullies or tease the victim. Understanding bystanders' joining in behavior requires an adaptation of an online group-based theory that captures the dynamics induced by the social environment and situational cues within a group bullying context.

In addition, research into bystanders' joining in behaviors has predominately focused on cyberbullying among children and adolescents who know each other in the classroom. Hence, these studies have viewed

cyberbullying as an extension of school or playground bullying and have explained joining in behaviors based on bystanders' offline relationship quality with the perpetrator and victim (Cao and Lin 2015). However, there is a need to derive alternative theoretical explanations for joining in behaviors among people with purely virtual relations on SNSs. Furthermore, studies have treated IT as a 'black box' or the research context without theorizing the effect of IT on bystanders' joining in cyberbullying behaviors. IT features are technical design choices made by platform owners that render different interpersonal interaction experiences (Kane et al. 2014). Therefore, understanding the roles of IT and its impacts on user behaviors is of prime importance, as these features may unintentionally afford and lead to undesirable SNS uses.

Against this backdrop, this study offers an integrative view of how IT influences bystanders' joining in cyberbullying behaviors on SNSs. First, building on the theoretical premises of the *social identity model of deindividuation effects* (the SIDE model) (Postmes et al. 1998, Reicher et al. 1995, Spears and Postmes 2015), we derive two *deindividuation experiences enabled by SNSs*, namely *experienced anonymity* and *experienced social identity*.[1] The SIDE model is well-suited for the current study because it was purposively developed to understand group dynamics and behaviors in computer-mediated communication settings. The model permits understanding how IT-enabled deindividuation experiences influence bystanders' behaviors in cyberbullying. We also draw on the *social network research framework* (Kane et al. 2014) to theorize how core IT features (including digital profile, search and privacy, relational ties, and network transparency) influence these deindividuation experiences. Finally, as joining in cyberbullying behaviors are socially undesirable, we incorporate *moral disengagement theory* (Bandura 2002, Bandura et al. 1996) into the SIDE model to capture the negative or undesirable aspects inherent in joining in behaviors. Given the paucity of research into bystanders' joining in behaviors among people with purely virtual relationships and little or no real-world connections, we choose a scenario in which online interest group members joined in to harass a new member who seemed deviant

[1] Hereafter we use 'experienced anonymity' for '*experienced anonymity enabled by SNSs'* and 'experienced social identity' for '*experienced social identity enabled by SNSs.'*

from the group's norms and values. The research model is tested using a scenario survey with two samples recruited from Amazon Mechanical Turk (MTurk) and Facebook, respectively.

This study contributes to IS research in novel ways and offers managerial insights for platform owners. Although IS researchers have recently begun to examine cyberbullying (e.g., Akram et al. 2022, Chan et al. 2019, Lowry et al. 2016, Marshall and Chan 2021, Wong et al. 2021), little effort has been made to investigate bystanders' joining in behaviors. Given the devastating consequences of such behaviors, we advance the IS literature by providing a rich theory of bystanders' joining in cyberbullying behaviors on SNSs through the lenses of the social network, deindividuation, and moral disengagement literature. This integrative model provides a conceptual bridge connecting social network features with deindividuation experiences, which in turn, directly influence bystanders' joining in behaviors or indirectly through moral disengagement mechanisms. Such a view echoes the sociotechnical perspective (Sarker et al. 2019) and enhances our theoretical understanding of the role of IT in fueling negative online group behaviors. It thereby generates interactional theoretical insights and makes a cross-disciplinary contribution to IS and cyberbullying literature (Tarafdar and Davison 2018). Furthermore, although SNS platform owners have taken initiatives to combat cyberbullying, the IS discipline has just started the conversation. Therefore, our integrative model enables IS researchers to comprehensively engage with their reference disciplines, address emerging societal issues, and advise platform owners regarding designs that can be implemented on SNSs to discourage cyberbullying behaviors. Thus, the findings of this study offer insights into the construction and maintenance of a safe, healthy, and sustainable online social environment.

The remainder of this paper is organized as follows. In the next section, we review the research on bystander responses to cyberbullying and introduce the theoretical foundations explaining bystanders' joining in behaviors on SNSs. We then discuss the proposed research model and develop the hypotheses. Next, we describe our research design and empirical results. Finally, we conclude the paper by discussing the implications for research and practice, the limitations of the study, and future research directions.

## 2. Research Background and Theoretical Foundations

### 2.1. Bystander Responses to Cyberbullying

Bystanders' responses to cyberbullying can be broadly categorized as (i) defending the victim and confronting the perpetrator, (ii) ignoring the incident, and (iii) joining in cyberbullying. Each response is driven by distinct factors and psychological mechanisms. Table A1 of Online Appendix A summarizes the literature on bystanders' responses to cyberbullying.

In studying bystanders' decision to defend the victim and confront the perpetrator, research has largely focused on the key steps specified in the bystander intervention model (Latané and Darley 1970). Perceived incident severity is one of the key factors influencing whether bystanders will take proactive actions to intervene (Bastiaensens et al. 2014, Koehler and Weber 2018). Specifically, the perception of severity promotes bystanders' intention to defend the victim through the mediating mechanisms of an increased perception of the situation's urgency and an increased feeling of personal responsibility (Obermaier et al. 2014). In addition, proactive personality traits (e.g., empathy; see Barlinska et al. (2018)) and positive social relationships (e.g., a close personal relationship between the bystander and victim; see Patterson et al. (2017)) increase the bystanders' tendency to defend the victim.

In contrast, research on bystanders' decision to ignore an incident has found that the presence of a higher number of bystanders increases bystanders' tendency to engage in passive observation (Brody and Vangelisti 2015), echoing the bystander apathy effect (Darley and Latane 1968). In addition, the characteristics of victims contribute to bystanders' inaction either through victim facilitation or through victim provocation, as prescribed in victim precipitation theory (Tepper et al. 2006). For instance, the victims' gender (Weber et al. 2019) and their information sharing behavior on SNSs (Schacter et al. 2016) have been shown to reduce bystanders' decisions to intervene in a cyberbullying situation due to victim blaming (Holfeld 2014).

Most published studies on bystanders' decision to join in cyberbullying, reinforce the bullies and attack the victim have found that the qualities of the bystander–victim relationship influence bystanders' joining in behaviors. This is not surprising given that most of this research has been conducted among

children in circumstances where the online relationship between the victim and bystanders was an extension of their offline relationship. For instance, Bastiaensens et al. (2014) found that bystanders were more likely to reinforce cyberbullying when they discovered that their good friends had reinforced the perpetrator's behavior. In addition, situational cues, such as bystanders' perception of injunctive norms that approve cyberbullying (Bastiaensens et al. 2016) and normative beliefs about aggression (Machackova and Pfetsch 2016), also predict joining in behaviors. For example, Jones et al. (2011) reported that identifying with the aggressive group members increased bystanders' tendency to join in. Finally, regarding individual characteristics, empathy produces an inhibitory effect on joining in behaviors (Barlinska et al. 2013), whereas previous victimization and perpetration experiences drive bystanders to join in cyberbullying (Bastiaensens et al. 2016). Nonetheless, in existing studies of bystanders' decision to join in cyberbullying, there is a lack of theoretical explanation of the drivers of bystanders' joining in behaviors among people having pure virtual relations. It is unclear why and how bystanders join in the perpetrators to bully a victim who have little real-world connections with them on SNSs. For this reason, we draw on the literature on social networks, deindividuation, and moral disengagement to develop an integrative model that explicates bystanders' joining in cyberbullying behaviors on SNSs.

### 2.2. Toward an Integrative View of Bystanders' Joining in Cyberbullying on SNSs

We draw on the SIDE model (Postmes et al. 1998, Reicher et al. 1995, Spears and Postmes 2015), the social network research framework (Kane et al. 2014), and moral disengagement theory (Bandura 2002, Bandura et al. 1996) to develop an integrative view (see Figure 1) that connects IT features with deindividuation experiences on SNSs, which then directly influence bystanders' joining in behaviors, or indirectly through the mediating effects of moral disengagement mechanisms.

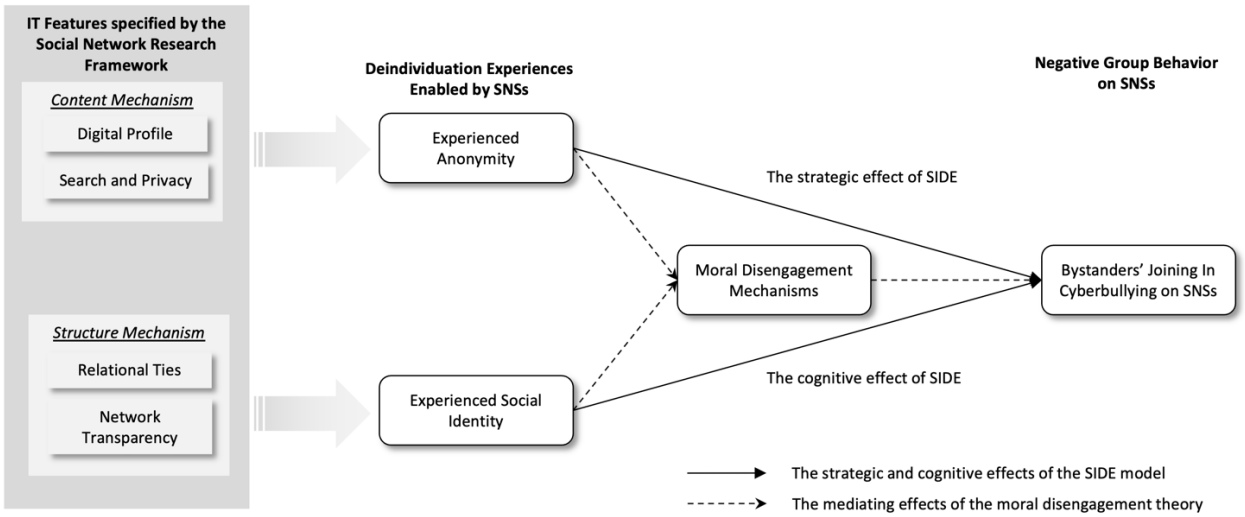


**Figure 1.** An Integrative View of Bystanders' Joining in Cyberbullying on SNSs

### 2.2.1. The Social Identity Model of Deindividuation Effects (The SIDE Model)

The SIDE model was specifically developed to understand media effects on online group behaviors and their underlying social psychological processes. Adopting a social identity and self-categorization approach (Tajfel and Turner 1986, Turner et al. 1987), the central premises of the SIDE model posit that the anonymous online environment suppresses the expression of individuality while accentuating the salience of social identity and supporting its expression, thus affecting group members' attitudes and behaviors (Walther 2011).

The SIDE model highlights two salient IT-enabled deindividuation experiences, namely experienced anonymity and experienced social identity (Spears and Lea 1994), that produce strategic and cognitive deindividuation effects and lead to online group behaviors. The *strategic effect* posits that online group behaviors are consequences of the opportunities enabled by *anonymity* to perform normative group behaviors or sometimes antisocial normative group behaviors that may otherwise be sanctioned or punished by authority figures. Hence, the strategic effect of the SIDE model explains "how [distinctive features of a communication technology] affect the ability to express identities in line with norms that might be sensitive to surveillance by the audience" (Spears and Postmes 2015, p. 33). Therefore, we argue

that experienced anonymity enabled by SNSs renders such strategic effect for bystanders' joining in behaviors. The *cognitive effect* presumes that online group behaviors are consequences of the salience of *social identity*, which shifts the perception of identity from an individual to a group level and regulates individual behaviors with the associated group norms (Spears and Postmes 2015). Hence, the cognitive effect of the SIDE model explains "how distinctive features of a communication technology affect the salience and operation of a particular identity" (Spears and Postmes 2015, p. 33). Accordingly, we argue that experienced social identity enabled by SNSs renders such cognitive effect for bystanders' joining in behaviors.

Based on the above theoretical premises, we use the SIDE model to guide our theorization. We argue that IT features (i.e., digital profile, search and privacy, relational ties, and network transparency) have the potential to enable the two deindividuation effects prescribed by the SIDE model, which lead to bystanders' joining in cyberbullying behaviors on SNSs. Specifically, we conceive that experienced anonymity allows bystanders to join in bullying even though such uninhibited and undesirable group behavior is subject to sanctions by authority; we also conceive that experienced social identity accentuates the values and beliefs exhibited by the group and focuses bystanders on salient norms manifest in the bullying situation. Hence, in an online interest group context, when members of the online group act as a group to bully one of the individuated members, bystanders (i.e., members of the group) will regulate their response by referencing the standard and expectation of the group and join in cyberbullying.

**2.2.2. The Social Network Research Framework and Deindividuation Experiences**

To derive richer insights into how IT enables the two deindividuation experiences, we draw on the social network research framework (Kane et al. 2014), which suggests that social interactions and outcomes on a social network (e.g., bystanders' joining in cyberbullying) could be influenced by four core IT features afforded by the social network, namely digital profile, search and privacy, relational ties, and network transparency (boyd and Ellison 2007). These four features are related to two mechanisms that explain interpersonal outcomes on a social network, namely *content* and *structure* (Borgatti and Foster 2003). The digital profile and search and privacy features, which represent the *content* explanatory mechanism,

enable users to determine how digital resources are shared and accessed through a social network. The relational ties and network transparency features, which represent the *structure* explanatory mechanism, enable users to establish and manage their connections with others in a social network.

We, thereby, argue that the digital profile and search and privacy features enable experienced anonymity. Experienced anonymity is defined as the extent to which bystanders perceive themselves and others as anonymous and non-identifiable in an online interest group on SNSs. For instance, the IT feature of '*manage activity logs'* allows users to manipulate, alter, and hide the content created and determine their availability. In addition, the '*search restriction'* feature enables users to determine who can view their profile and access the content they have contributed and allows users to also protect their content against discovery by search algorithms. These features allow users to manipulate their identities on the platform (i.e., the content explanatory mechanism). Specifically, when engaged in a particular behavior that potentially violates the publicly accepted norms, users could find these IT features enabling their engagement in deviant and sometimes harmful online group behaviors. The strategic advantages offered by experienced anonymity allow bystanders to join in cyberbullying—a socially unacceptable yet normative group behavior—without fearing potential sanctions from outgroup and governmental agencies. Empirical evidence showing the relationship between IT features of digital profile and search and privacy and experienced anonymity can be found in Table A2 of Online Appendix A.

We further argue that relational ties and network transparency features enable experienced social identity. Experienced social identity is defined as a bystander's self-concept that derives from their knowledge of membership in a social group, together with the value and emotional significance attached to that membership. For instance, the IT feature of '*joining* and *managing groups'* enables users to define a list of others with whom they can view and track their connections and followers on the platform (e.g., a Facebook group enables users with shared interests to participate in group communication). In addition, the '*people you may know'* feature suggests relationships that may be developed between individuals based on their network. These features allow users to join online groups with goals and values that match their own (i.e., the structure explanatory mechanism). While being part of a group, the experience of

social identity entices users to regulate their' attitudes and behaviors based on the salient norms, beliefs, and values exhibited by the group on a social network. The cognitive reconsideration triggered by experienced social identity renders bystanders to join in cyberbullying, even when the act of behavior is deviant and harmful. Empirical evidence showing the relationship between IT features of relational ties and network transparency and experienced social identity can be found in Table A2 of Online Appendix A.

**2.2.3. Moral Disengagement Theory**

To offer a contextualized explanation of bystanders' joining in behaviors on SNSs, we integrate moral disengagement theory (Bandura 2002, Bandura et al. 1996) with the SIDE model to explain how bystanders justify their detrimental joining in cyberbullying behaviors on SNSs through moral disengagement mechanisms. The SIDE model has been used to explain online group behaviors, independent of their positive or negative natures. For instance, the model has been used to explain posting supportive messages (Li and Zhang 2018) but also to explain cheating in online games (Chen and Wu 2015). As bystanders' joining in cyberbullying behaviors is detrimental, moral disengagement theory complements the SIDE model by accounting for the psychological tensions that bystanders may need to bypass when they are conscious of the negative impacts of joining in.

Moral disengagement theory explains the psychological maneuvers through which individuals' self-regulatory mechanisms of moral agency are selectively disengaged to enable engaging in detrimental conduct without experiencing psychological tensions (Bandura 2002, Bandura et al. 1996). Bandura (2002) argued that individuals care about acting morally and behaving ethically, and such actions are regulated through exercising moral agency. Individuals obtain a sense of satisfaction and self-worth by engaging in behaviors consistent with their moral standards, whereas the reverse brings self-sanctions and induces psychological discomfort and cognitive dissonance. Nevertheless, the regulation of moral action does not operate as an invariant internal control system and is not impervious to environmental and social conditions. Bandura (2016) also emphasized that "*the advent of the Internet ushered in a ubiquitous vehicle for disengaging moral self-sanctions from transgressive conduct. The Internet was designed as a*

*highly decentralized system that defies regulation. Anybody can get into the act, and nobody is in charge*" (p. 68). In particular, the regulation of moral action can be deactivated to justify negative behaviors through four major psychological mechanisms: (i) reconstruing the conduct, (ii) obscuring personal causal agency, (iii) misrepresenting or disregarding injurious consequences, and (iv) vilifying the recipients of maltreatment. By activating moral disengagement mechanisms, individuals can justify their inhumane conduct and avoid self-sanctioning. Hence, moral disengagement theory offers a theoretical lens to explain how bystanders perform detrimental joining in cyberbullying behaviors by exercising various psychological maneuvers.

**3. Research Model and Hypothesis Development**

Based on the theoretical premises of the SIDE model and moral disengagement theory, we posit that experienced anonymity and experienced social identity exert direct effects on bystanders' joining in cyberbullying. We also expect the two deindividuation experiences enabled by SNSs to exert indirect effects on bystanders' joining in cyberbullying via the four moral disengagement mechanisms listed above. Figure 2 depicts the proposed research model.

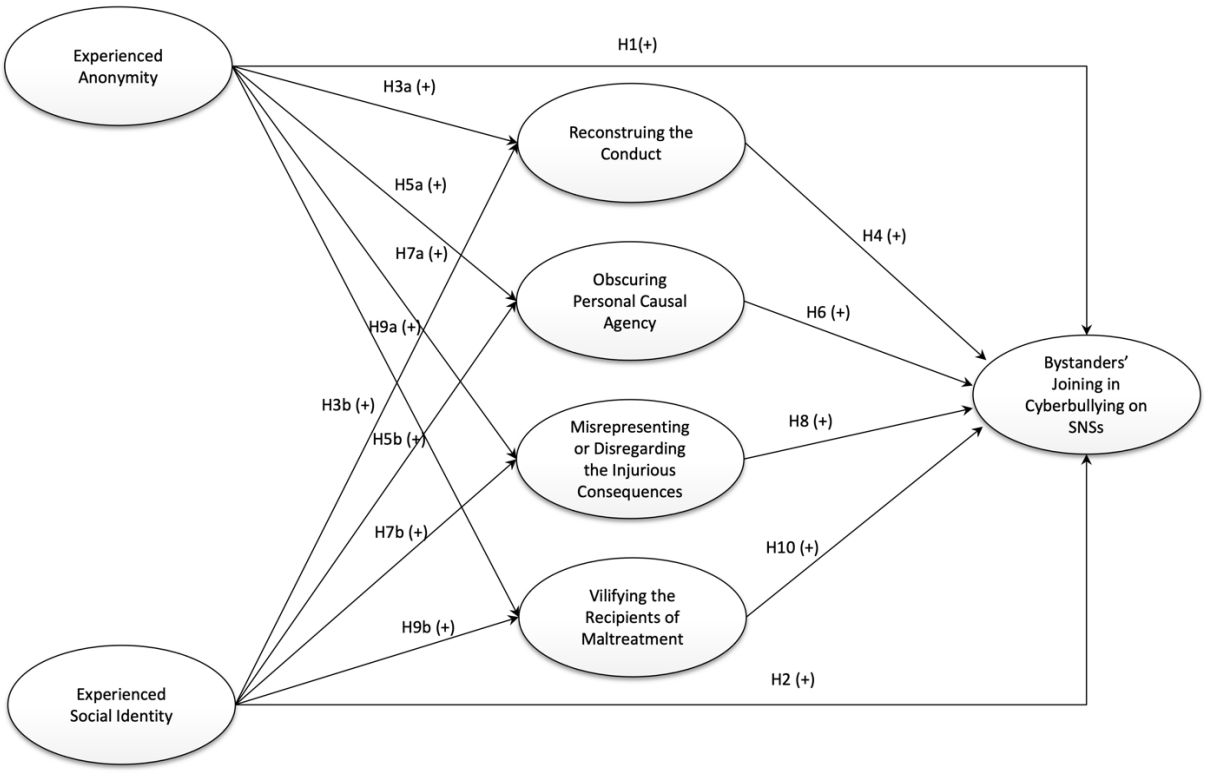

**Figure 2.** Proposed Research Model

### 3.1. Experienced Anonymity

Subscribing to the content explanatory mechanism of the social network research framework and the strategic effect of the SIDE model, experienced anonymity enabled by SNSs offers strategic advantages for joining in cyberbullying. The effect of experienced anonymity is akin to the theoretical premises specified in classical deindividuation theory (Prentice-Dunn and Rogers 1989). Both theories predict that individuals who believe they are anonymous are more likely to engage in deviant and sometimes harmful group behaviors (e.g., joining in cyberbullying) because they will not be identified and apprehended. Nevertheless, while deindividuation theory asserts people are mindless crowds, the SIDE model argues that individuals are aware that their actions could go against the laws and regulations and socially accepted norms. Experienced anonymity, however, offers strategic advantages for individuals to engage in these behaviors. This is because being unidentifiable to platform owners, governmental agencies, and other general SNS users implies possibilities to escape from formal and informal sanctions. For instance, Prentice-Dunn and Rogers (1989) found that students identifiable to researchers were less likely to engage in cheating behaviors. Reicher and Levine (1994) reported that anonymous students were more likely to endorse deviant group behaviors. In our study, experienced anonymity is expected to reduce identification cues and enable bystanders to behave in ways that may be against publicly accepted norms and the rules of platform owners and governmental agencies.

**H1:** Experienced anonymity is positively related to bystanders' joining in cyberbullying on SNSs.

### 3.2. Experienced Social Identity

Based on the structure explanatory mechanism of the social network research framework and the cognitive effect of the SIDE model, experienced social identity enabled by SNSs shifts one's attention from a personal to a social categorical level, leading to the prioritization of group beliefs and values and encouraging action consistent with the perceived stereotypic group norms. The cognitive considerations cause bystanders to regulate their behaviors by referencing the salient situational norm within the group (i.e., joining in cyberbullying) to meet the group expectation or avoid expulsion (Chen 2018). Studies

have shown that the online environment deters the expression and detection of individuality and accentuates group identity, especially when members are attached based on an interest in a particular topic and resource (Sassenberg 2002). For instance, Gaudette et al. (2021) found that members of the 4chan /b/ discussion board who engaged in trolling behaviors held a strong sense of shared identity. The sense of collective identity on 4chan is so strong that users do not refer actions on the board as something individuals performed but instead express agency in the form '4chan did' or '/b/ did.' Bastiaensens et al. (2014) also reported that bystanders had significantly higher behavioral intentions to join in cyberbullying when they noticed that their group members had reinforced cyberbullying. In our model, experienced social identity regulates bystanders' behavior such that they behave in ways that conform to the situational norms (i.e., joining in cyberbullying).

**H2:** Experienced social identity is positively related to bystanders' joining in cyberbullying on SNSs.

### 3.3. Moral Disengagement

#### 3.3.1. Reconstruing the Conduct

Reconstruing the conduct refers to the extent to which joining in is justified through the reconstrual of cyberbullying behavior. Such a disengagement practice includes euphemistic labeling, palliative comparison, and moral justification. In our study, we expect the two deindividuation experiences enabled by SNSs to influence bystanders' joining in behavior on SNSs through reconstruing the conduct.

Experienced anonymity enabled by SNSs lowers self-evaluation and evaluation apprehension (Postmes and Spears 1998). Bystanders with low self-awareness may believe that they are not accountable to other members and regard their joining in behaviors as not intended harm but 'a harmless joke' (Bauman and Newman 2013). For instance, SNS users tend to label cyberbullying euphemistically, and its reinforcement is an exchange of 'banter' (Steer et al. 2020). When bystanders are not aware of the seriousness of the joining in behaviors, they may justify their behavior by claiming that leaving an insulting comment online is less serious than insulting the individual in person.

Experienced social identity enabled by SNSs prompts members to behave in ways that they believe will benefit their group (Louis et al. 2005). Hence, bystanders may portray joining in behaviors as a

means to serve the goals and reinforce the values of the group (i.e., a positive anticipatory outcome thats benefit the group). Indeed, when people are paired with a partner or assigned to a group, they rate the enactment of antisocial behaviors as more acceptable and experience less guilt (Behnk et al. 2017). Thornberg (2015) showed that perpetrators reconstrued a favorable identity for bystanders who joined in cyberbullying. Furthermore, group members believed that the group's norms, rules, and moral standards, were superior to and more positive and correct than established social norms, eliciting the practice of moral justification (Alleyne et al. 2014).

Taken together, reconstruing the conduct allows bystanders to relabel their harmful actions as benign by using sanitizing language or convoluted concepts (i.e., euphemistic labeling) to contrast harmful behavior with even more reprehensible acts to make it seem more acceptable (i.e., palliative comparison) and portraying harmful behavior as serving a socially worthy or moral purpose (thus considering it personally and socially acceptable; i.e., moral justification). Once the harmful nature of reinforcement is reconstrued, bystanders can join in cyberbullying without experiencing psychological discomfort. Studies have found a positive association between moral justification and bullying behaviors (DeSmet et al. 2016, Robson and Witenberg 2013). For instance, cognitive reconstrual has been found to be positively associated with assisting (e.g., forwarding the post) or reinforcing (e.g., adding nasty comments) cyberbullying among adolescent bystanders (DeSmet et al. 2016).

**H3:** (a) Experienced anonymity and (b) experienced social identity is positively related to reconstruing the conduct.

**H4:** Reconstruing the conduct is positively related to bystanders' joining in cyberbullying on SNSs.

### 3.3.2. Obscuring Personal Causal Agency

Obscuring personal causal agency refers to the extent to which joining in is justified by denying personal agency for cyberbullying behaviors. Such a disengagement practice includes displacement of responsibility and diffusion of responsibility. In our study, we expect the two deindividuation experiences enabled by SNSs to influence joining in behavior on SNSs through the mechanism of obscuring personal causal agency.

Experienced anonymity enabled by SNSs reduces bystanders' sense of individuality and divests them of personal responsibility for their behavior in cyberbullying (Valkenburg and Peter 2011). Bystanders may believe that their behaviors are not accountable to other SNS users or platform owners. They may even argue that it is the platform owner's responsibility to maintain a healthy online space and moderate inappropriate posts. This displacement of responsibility reduces the negative affective state induced by cognitive dissonance (Gosling et al. 2006). Alternatively, bystanders may argue that it is the authorities' responsibility to draw a clear line between acceptable and unacceptable behaviors on SNSs. Indeed, Patterson et al. (2016) reported that bystanders perceived a lack of established rules, authority figures, and formal reporting mechanisms online. By transposing the cause of their harmful behavior to authority figures, bystanders are spared from self-censure. Furthermore, experienced anonymity makes it challenging to assign responsibility for harmful content to a specific group member (Runions and Bak 2015).

Experienced social identity enabled by SNSs leads bystanders to view the online interest group as a collective decision-maker and actor rather than an idiosyncratic personal identity (Walther 2011). Being part of a group is a major expedient for engaging in detrimental collective behaviors (Bandura et al. 1996). People behave more cruelly when they have the opportunity to attribute responsibility diffusely across a group rather than directly to one person (Bandura et al. 1975). In addition, the anticipated risk of formal sanctions and informal social costs associated with harmful behaviors decrease as group size increases (McGloin and Thomas 2016). Experienced social identity could lead bystanders to see themselves as interchangeable representatives of the online interest group on SNSs. This would help them deny their agentive role and attribute responsibility for harmful acts to the group (Anderson et al. 2014). A shared social identity also minimizes the psychological burden, such as regret, punishment, and stress (El Zein et al. 2019), resulting from any adverse outcomes of participating in negative collective behaviors.

Taken together, obscuring personal causal agency allows bystanders to view harmful joining in behaviors as the responsibility of the platform owners (i.e., displacement of responsibility) and distribute

the accountability for behavior across other people (i.e., diffusion of responsibility). When their agentive role in the harm caused by cyberbullying can be obscured or minimized, bystanders can join in harmful behaviors without self-condemnation. Runions and Bak (2015) contended that the 'social nature' of SNSs enables the diffusion and displacement of responsibility, with SNS users tending to view any pernicious act on SNSs as being spread across several agents who each play only a small role. Robson and Witenberg (2013) also showed that the diffusion of responsibility increased cyberbullying among students.

**H5:** (a) Experienced anonymity and (b) experienced social identity is positively related to obscuring personal causal agency.

**H6:** Obscuring personal causal agency is positively related to bystanders' joining in cyberbullying on SNSs.

### 3.3.3. Misrepresenting or Disregarding Injurious Consequences

Misrepresenting or disregarding injurious consequences refers to the extent to which joining in is justified through the ignorance, minimization, distortion, or disbelief of the harmful effects of cyberbullying behavior. In our study, we expect the two deindividuation experiences enabled by SNSs to influence joining in behavior on SNSs through the mechanism of misrepresenting or disregarding injurious consequences.

Experienced anonymity enabled by SNSs lowers the threshold of self-evaluation and the standard for self-regulatory behaviors because bystanders are no longer being seen or identified by their interaction partners (Postmes and Spears 1998). The possibility enabled by IT to be invisible and unidentifiable, such as by using a fake profile photograph instead of an actual photograph of their life with family members, allows bystanders to easily discredit evidence of any harm they may cause through selective inattention and cognitive distortion, ignoring any injurious consequences of their joining in actions.

Experienced social identity enabled by SNSs lures bystanders to focus on the social benefits of joining in cyberbullying (e.g., obtaining approval from and gaining prestige within the group) by avoiding or minimizing the recognition of harm. Roccas et al. (2004) showed that the feeling of guilt for misconduct

toward an outsider could be minimized when people are in a group. Social identity also reduces cognitive dissonance between personal norms and reflected group norms (Glasford et al. 2009). As a result, experienced social identity may encourage bystanders to misrepresent or disregard the victim's suffering and engage in collective violence that aligns with the group norms (Leidner et al. 2010). Any evidence of harm when they join in attacking a victim could be discredited or ignored.

Taken together, and consistent with DeSmet et al. (2016), we expect misrepresenting or disregarding injurious consequences to allow bystanders to join in cyberbullying by ignoring, minimizing, distorting, or disbelieving the harm that they are causing.

**H7:** (a) Experienced anonymity and (b) experienced social identity is positively related to misrepresenting or disregarding injurious consequences.

**H8:** Misrepresenting or disregarding injurious consequences is positively related to bystanders' joining in cyberbullying on SNSs.

### 3.3.4. Vilifying the Recipients of Maltreatment

Vilifying the recipients of maltreatment refers to the extent to which joining in is justified through the defamation of cyberbullying victims. Such a disengagement practice includes dehumanization and the attribution of blame. In our study, we expect the two deindividuation experiences enabled by SNSs to influence joining in behavior on SNSs through the mechanism of vilifying the recipients of maltreatment.

Experienced anonymity enabled by SNSs disinhibits bystanders from social expectations and standards and allows them to act apathetically toward the victim (Cheung et al. 2020, Diener and Wallbom 1976). Suler (2004) contended that users tend to engage in solipsistic introjection when they feel disinhibited; they consciously or unconsciously create and visualize communication partners within their intrapsychic world based on their expectations, wishes, and needs. Hence, it is reasonable to assume that bystanders who experience anonymity on SNSs can apathetically introject a negative image of the victim and attribute the blame for any harm suffered by the victim to their perceived or imagined characteristics. For example, research has shown that bystanders blame victims for being extroverted and disclosing too much information on SNSs (e.g., Schacter et al. 2016, Weber et al. 2013).

Experienced social identity enabled by SNSs encourages bystanders to delineate the boundaries of their social circle by excluding individuals who do not fit in. People tend to treat those in the same social circle (or those perceived as similar) with greater moral concern, thus empathizing with them if mistreated (Giner-Sorolla et al. 2012). In our study, we expect devaluing a victim as an outsider weaken bystanders' empathetic and vicarious emotional reactions to the victim's suffering (McHugo et al. 1982), thus justifying joining in behaviors. Indeed, dehumanizing people considered outsiders is among the most common forms of group dynamics (Harris and Fiske 2006). In the cyberbullying context, when bystanders observe members of the online interest group using derogatory stigmatizing labels to marginalize the victim and reinforce the group identity, they may have a greater tendency to blame the victim (Gini 2007) and consider it acceptable to treat the victim inhumanely (Waytz and Epley 2012).

Taken together, vilifying the recipients of maltreatment enables bystanders to (i) deprive the victim of human qualities or attribute bestial qualities to the victim (i.e., dehumanization) and (ii) blame the victim for bringing suffering upon themselves or ascribe harmful conduct to circumstances beyond the bystanders' control (i.e., attribution of blame). Studies have observed the attribution of the cause of bullying to deviant characteristics or behavior of the victims (e.g., Forsberg et al. 2014, Mooij 2011). For instance, Forsberg et al. (2014) found that bystanders defined victims as odd, stupid, or disturbing when justifying their bullying behavior through dehumanization and victim blaming.

**H9:** (a) Experienced anonymity and (b) experienced social identity is positively related to vilifying the recipients of maltreatment.

**H10:** Vilifying the recipients of maltreatment is positively related to bystanders' joining in cyberbullying on SNSs.

## 4. Research Method

### 4.1. Research Design

We developed an online survey to collect data for model testing using the scenario (vignette) technique. The technique facilitates discovering participants' responses to hypothetical situations (Wilks 2004); it is not restricted to written texts but can also consist of images, videos, or other media (Hughes and Huby

2002). In this study, the hypothetical cyberbullying scenario included a short descriptive text and an image about a cyberbullying incident in an online interest group on an SNS. Given that online harassment related to physical appearance is one of the most common forms of cyberbullying (Pew Research Center 2017), the descriptive text and image portrayed a scenario involving harassment of a target's body image.

The scenario involved the harassment of a new member by an anonymous perpetrator who is an existing member of an online interest group on Facebook and by other members of the same group. The group was originally set up for SNS users with shared interests to exchange information and engage in conversation related to fitness and exercise. The cyberbullying post concerned Pat (the victim), a new member who had just joined the fitness group, as shown in the bullying message 'Our newbie Pat' and the new member badge (i.e., a waving hand 👋) next to Pat's screen name. The harassment was carried out by the interest group members using language intended to upset and disturb the victim. The attacks were unsolicited and perceived as hurtful by the victim, who had asked the bullies to stop. Such harassment manifests as cyberbullying because (i) the group displays a clear norm and unified *intention to harm* the new member; (ii) there is a significant *power imbalance* between existing members and the new member, in terms of the number of bullies and social power; and (iii) the harassment of the victim with offensive language by existing group members constitutes *repetition* (Chan et al. 2021).

We created four posting scenarios by varying the gender of the victim (i.e., female or male) and the format of the post (i.e., text-only or text and image), while keeping the bullying message identical. This ensured a more generalizable representation of cyberbullying incidents occurring on SNSs. In addition, the four hypothetical posts were pre-tested with Facebook users to assess their naturalness and realism. The pre-test results indicated that the four posts were realistic and created arousal and that the messages to the victim were hurtful and negative. Online Appendix B provides the justification for using the scenario technique, and details of the development, validation, and pre-testing of the scenario.

### 4.2. Questionnaire Design and Measures

The questionnaire consisted of four parts. In the first part, after providing their consent to participate in the study, the participants were asked to answer three screening questions: (1) 'Which social networking

platforms do you usually use? (Select up to three)'; (2) 'Do you reside in the United States?'; and (3) 'Are you a member of a Facebook fitness group?' Screening Question 1 offered multiple social networking platforms. Only those participants who selected Facebook as one of their regular SNSs and answered 'Yes' to the other two screening questions were allowed to proceed.[2] Participants who did not meet the screening criteria were thanked and dismissed.

In the second part, the participants were asked to provide details of the actual Facebook fitness group to which they belonged, including the group name and their experience in the group. The group name was subsequently integrated into the scenario and questionnaire to enhance realism.

In the third part, the participants were given a textual description that set the scene for the cyberbullying incident. The scenario text portrayed a situation where the participants were browsing their Facebook newsfeed in their spare time and came across a post from their Facebook fitness group that caught their attention. The post was about Pat, a new member of the fitness group. The participants were told they had had no personal relationship or prior interaction with Pat since Pat had just joined the group. However, some fitness group members had already reacted to, left comments on, and shared the post. One of the four cyberbullying posts was then randomly presented to the participants. They were asked to pay attention to the post and read the comments attached to it. We did not hint at or use words related to cyberbullying in the instruction to minimize priming effects. Then, the participants were asked to describe what they had observed and how they felt in their own words. They were then asked to answer the questions related to the focal constructs. They were also asked whether they recognized that Pat (the victim) had asked the group members (the bullies) to stop and whether they felt that the post and interactions were hurtful, negative, arousing, and realistic.

In the last part, the participants were asked to provide demographic information and answer questions related to control variables, social desirability, and a marker variable. They were then debriefed and thanked.

[2] Since the laws and norms regarding social media use and cyberbullying vary across different countries, participants from the same country were recruited to maintain the consistency (Lowry et al. 2016).

All the measurement items were adapted from previous research, with slight modifications to fit the current research context. The items for all constructs were measured on a seven-point Likert scale. Multiple items were used to evaluate each construct to ensure validity and reliability. We conceptualized all of the constructs as reflective using the criteria from the methodological literature (MacKenzie et al. 2011, Polites et al. 2012). We also included three categories of control variables that may influence bystanders' behavioral decisions, which consist of (1) demographic characteristics and general SNS usage and self-efficacy, (2) personality traits (i.e., empathy and self-presentation) and experience with cyberbullying perpetration and victimization, and (3) variables related to the Facebook fitness group (i.e., the experience in the group, and frequency of interaction with other group members). Online Appendix C presents the list of measurement items and the details of the measurement model specification.

**4.3. Data Collection**

We pre-tested the preliminary version of the full survey instrument for clarity and validity with a panel of six social science researchers. The panel members' comments and feedback were considered when preparing the final version of the questionnaire. We then conducted a *pilot* test involving 255 SNS users recruited from MTurk to evaluate the flow and clarity of the revised questionnaire.

The main study comprised two rounds of data collection. Invitations to participate in an anonymous online survey were posted on MTurk and Facebook. The main survey was accessible via the two platforms for one week in November 2020. The invitation to participate on MTurk was set up as a Human Intelligence Task (HIT) and was made browsable and searchable on the MTurk website. Workers who completed the survey questionnaire were compensated by the specified reward for participation in a HIT. The invitation to participate on Facebook was delivered as a Boosted Post promoted to Facebook users in the United States aged 18 years or older. Interested Facebook users clicked on the survey link listed in the post to complete the questionnaire and were entered into a lucky draw to win an Amazon e-gift card.

During the survey period, 4,581 participants attempted the survey, and 2,272 passed the screening questions, of whom 1,577 participants completed the entire questionnaire. Responses were removed from the dataset if they (i) failed most of the attention-check questions or provided careless, random, or

haphazard responses; or (ii) failed to recognize that the victim had asked the bullies to stop. This left 1,179 valid and complete responses for the subsequent analysis; 696 from MTurk and 483 from Facebook. Table D1 of Online Appendix D shows a detailed breakdown of the characteristics of the two samples and describes the details of our data collection approaches and measures taken to ensure data quality.

## 5. Data Analysis and Results

### 5.1. Preliminary Analyses

We performed preliminary tests before assessing the measurement and structural models to detect Common Method Bias (CMB) and Social Desirability Bias (SDB). The results indicated that the two biases were not a major issue in this study (see Online Appendix E and F). We also tested whether there were any significant differences between the two samples. The results indicated that the two samples were comparable, with no significant differences between the two sets of participants (see Online Appendix G).

### 5.2. Model Testing

We validated the measurement and structural models using partial least squares (PLS) structural equation modeling. The SmartPLS 3.3.0 software package was used for the analysis. The PLS approach was chosen to test the research model based on methodological guidelines (see Online Appendix G). Following the two-step analytical approach (Hair et al. 2017), we performed a psychometric assessment of the measurement model, followed by an evaluation of the structural model, to ensure that the conclusions of the structural model were drawn from a set of measures with valid psychometric properties.

#### 5.2.1. Measurement Model

The test of the measurement model involved estimations of the reliability, convergent validity, and discriminant validity of the measurement items. The reliability and validity testing results indicate that the measurement model's psychometric properties were satisfactory (see Online Appendix G).

#### 5.2.2. Structural Model

We performed bootstrapping with 5,000 subsamples to test the significance levels of the path coefficients in the research model (Hair et al. 2017). Figure 3 shows the data analysis results.

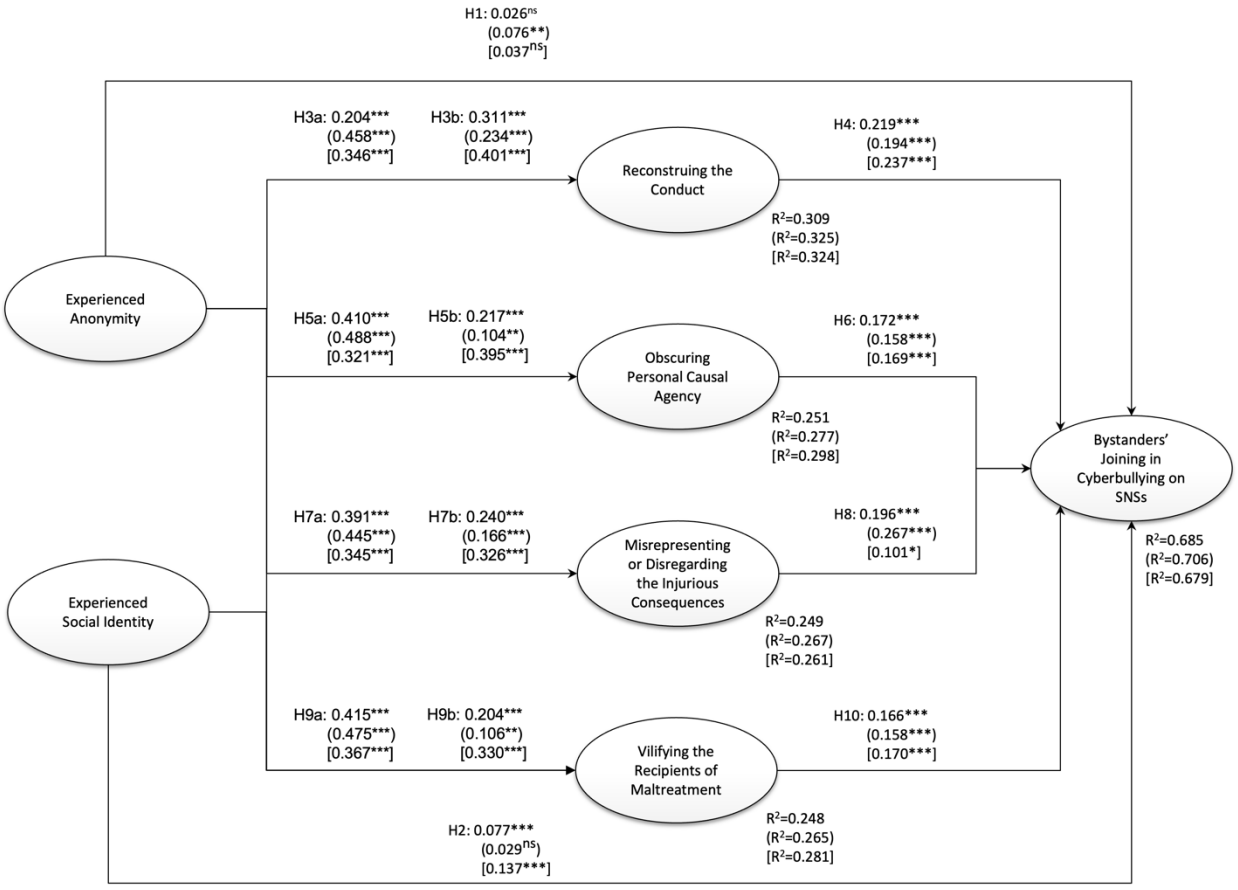


**Figure 3.** Results of Data Analysis
***Note 1.*** Non-significant$^{ns}$, $p < 0.05$*, $p < 0.01$**, $p < 0.001$***
***Note 2.*** Results based on the combined sample are shown without brackets; those with the MTurk sample are shown in parentheses; and those with the Facebook sample are shown in square brackets.

The coefficient of determination ($R^2$) returned high values for the key focal constructs. In the combined MTurk and Facebook sample, the model explained 68.5%, 70.6%, and 67.9% of the variance in bystanders' joining in cyberbullying, respectively. The results substantiate the predictive validity of the research model. The validity of the findings was also evaluated using the $Q^2$ value. After running the blindfolding procedure with an omission distance $D = 7$, we obtained $Q^2$ values well above zero for the dependent variables, indicating the predictive relevance of the research model. The path coefficients and

significance levels obtained indicate that most hypotheses were supported in the full model. In the combined and Facebook sample, only the relationship between experienced anonymity and bystanders' joining in cyberbullying was non-significant ($\beta_{\text{Combined}} = 0.026$, $p > .05$; $\beta_{\text{Facebook}} = 0.037$, $p > .05$), whereas in the MTurk sample, only the relationship between experienced social identity and bystanders' joining in cyberbullying was non-significant ($\beta_{\text{MTurk}} = 0.029$, $p > .05$). In general, the empirical results supported the validity of the deindividuation and moral disengagement perspectives in understanding bystanders' joining in behaviors. Self-presentation tendency was the only control variable that was significantly associated with bystanders' joining in cyberbullying, and this was only observed in the combined sample and MTurk sample ($\beta_{\text{Combined}} = 0.057$, $p < .05$; $\beta_{\text{MTurk}} = 0.073$, $p < .05$). We explain the above results in detail in the discussion section.

### 5.3. Post-Hoc Analysis

#### 5.3.1. Mediating Effects of Moral Disengagement Mechanisms

We compared the path model estimates with and without the moral disengagement constructs. Online Appendix H shows the data analysis results (see Table H1). For the baseline model, the results indicate that experienced anonymity and experienced social identity were positively and significantly related to bystanders' joining in cyberbullying. Regarding the full research model, we added the moral disengagement constructs as mediators of the relationship between experienced anonymity and experienced social identity and bystanders' joining in cyberbullying. In the combined sample, the $R^2$ value increased from 53.1% to 68.5%, indicating a large effect size of the moral disengagement mechanisms as predictors of bystanders' joining in cyberbullying ($f^2 = 0.489$). In the MTurk sample, the $R^2$ value increased from 54.2% to 70.6%, and in the Facebook sample, from 56.3% to 67.9%, also suggesting a large effect size of the moral disengagement mechanisms on bystanders' joining in cyberbullying ($f^2 = 0.558$ and $f^2 = 0.361$, repsecitvely).

We used the bootstrapping technique to assess the mediation effects of the moral disengagement mechanisms. Online Appendix H shows the results of the bootstrapping analyses (see Table H2a-c). The

results confirm the moral disengagement mechanisms as mediators of the effects of experienced anonymity and experienced social identity on the focal construct of bystanders' joining in cyberbullying. We identified two interesting patterns across the MTurk and Facebook samples. In both samples, experienced anonymity and experienced social identity had significant effects on the four moral disengagement mechanisms, which in turn had significant effects on the likelihood that bystanders would join in cyberbullying. The indirect effects of experienced anonymity and experienced social identity via these mediator constructs were significant. Furthermore, in the MTurk sample, the relationship between experienced anonymity and bystanders' joining in cyberbullying remained significant, suggesting partial (complementary) mediation of the moral disengagement mechanisms. However, the direct effect of experienced social identity on bystanders' joining in cyberbullying was not significant, suggesting full mediation. In the Facebook sample, the relationship between experienced social identity and bystanders' joining in cyberbullying remained significant, suggesting partial (complementary) mediation of the moral disengagement mechanisms. However, the direct effect of experienced anonymity on bystanders' joining in cyberbullying was not significant, suggesting full mediation.

Finally, we performed an additional analysis to compare the strengths of the moral disengagement mechanisms on bystanders' joining in cyberbullying across the four cyberbullying posts; i.e., according to the gender of the victim (i.e., female or male) and the format of the post (i.e., text-only or text and image). Table H3 of Online Appendix H shows the results. The effects of moral disengagement mechanisms on bystanders' joining in cyberbullying are consistent among the four posts. No significant difference was observed between the gender of the victim portrayed in the post or the format of the post.

**5.3.2. Role of IT on Deindividuation Experiences**

To further test our theorization of the influence of IT on deindividuation experiences, we examined the association between the four IT features and deindividuation experiences. Online Appendix H shows the measurement items (see Table H4) and their psychometric properties (see Table H5 and H6a-c). The data analyses indicated that digital profile ($\beta_{\text{Combined}} = 0.241$, $p < .001$; $\beta_{\text{MTurk}} = 0.300$, $p < .001$; $\beta_{\text{Facebook}} = 0.176$, $p < .01$) and search and privacy ($\beta_{\text{Combined}} = 0.225$, $p < .001$; $\beta_{\text{MTurk}} = 0.274$, $p < .001$; $\beta_{\text{Facebook}} =$

0.140, $p < .05$) had significant and positive relationships with experienced anonymity, whereas relational ties ($\beta_{Combined} = 0.348$, $p < .001$; $\beta_{MTurk} = 0.376$, $p < .001$; $\beta_{Facebook} = 0.282$, $p < .001$) and network transparency ($\beta_{Combined} = 0.256$, $p < .001$; $\beta_{MTurk} = 0.257$, $p < .001$; $\beta_{Facebook} = 0.183$, $p < .05$) had significant and positive relationships with experienced social identity. These results provided further evidence regarding how IT features enable deindividuation experiences.

## 6. Discussion

### 6.1. Key Findings

Our findings provide strong empirical support for our integrative view on bystanders' joining in cyberbullying on SNSs, particularly for users with purely virtual relations (i.e., with little or no real-world connections). We illustrate the relevance of the SIDE model in providing an useful explanation of online group behaviors on SNSs, where both experienced anonymity and experienced social identity significantly influence bystanders' joining in cyberbullying. The post-hoc analysis results also support our theory that IT enables the two deindividuation experiences. Specifically, digital profile and search and privacy, as presumed in the content explanatory mechanism in enabling users to determine the availability of information on the platform and to other users, lead to experienced anonymity. Relational ties and network transparency, as presumed in the structure explanatory mechanism in enabling users to establish and manage the connections between them and others in a social network, influence experienced social identity. The findings suggest that the social network research framework represents a parsimonious and legitimate framework for understanding the effects of IT on negative and undesirable interpersonal outcomes through connecting with the deindividuation perspective.

In addition, we demonstrate the power of moral disengagement as a set of psychological mechanisms justifying the engagement in detrimental online group behaviors. All four moral disengagement mechanisms partially or fully mediated the influence of experienced anonymity and experienced social identity on bystanders' joining in cyberbullying, indicating the deindividuation perspective is likely incomplete without the moral disengagement perspective. With partial mediation, the finding suggests that both the deindividuation experiences and moral disengagement mechanisms explain part of the

variance in bystanders' joining in cyberbullying; whereas with full mediation, the finding suggests that the moral disengagement mechanisms fully capture the effects of the deindividuation experiences on bystanders' joining in cyberbullying. The mediation effect suggests that the moral disengagement theory effectively complements the SIDE model by accounting for the psychological discomfort one might experience in participating in deviant and undesirable group behaviors when the deindividuation effect is insufficient to mobilize the enactment of such harmful behaviors. The results suggest that whilst IT plays a substantial role in enabling deindividuation, psychological mechanisms also contribute to the explanation of online harmful social behaviors, corroborating the sociotechnical perspective (Sarker et al. 2019).

The observed empirical findings were consistent across the four cyberbullying posts (i.e., between male or female cyberbullying victims and between text-only or text-and-image cyberbullying posts). These findings suggest that the conventional gender stereotype of female victims as more vulnerable than male victims may not hold in the digital space (Weber et al. 2013), especially in an online environment where bystanders have no prior interaction and have no knowledge of the victim. We, therefore, suggest that both male and female victims can be marginalized and disadvantaged in an anonymous online environment where social identity is salient (e.g., in an interest group or a political news thread setting). In addition, we find that regardless of their format, cyberbullying posts can be equally damaging by activating moral disengagement mechanisms. Specifically, for the text-only format, writing a harassing post in pure text involves minimal effort and might imply that it is an unintentional harmless joke. Bystanders might justify their joining in behaviors by reconstruing the conduct. For the text-and-image format, the image contained unfavorable cues for internal attribution (e.g., body image of the victim). Bystanders might justify their joining in behaviors by vilifying the recipients of maltreatment. With bystanders having no prior interaction with or knowledge of the victim, the victim could be easily discriminated against because of unfavorable visual cues that alienate them from the common basis of the online group. Based on these observations, we believe that moral disengagement theory offers a powerful theoretical explanation of bystanders' joining in behaviors across various cyberbullying situations.

Our analysis also revealed a surprising finding regarding the relative effects of the two deindividuation experiences. The strength and effect of the two deindividuation experiences were influenced by the platform on which the data were collected. Specifically, after adding the moral disengagement constructs into the baseline model, the influence of experienced anonymity on bystanders' joining in cyberbullying was significant only in the MTurk sample, whereas the influence of experienced social identity on bystanders' joining in cyberbullying was significant only in the Facebook sample. A plausible explanation is that the participants recruited from MTurk might have had stronger and more concrete notions of anonymity (and weaker and more abstract levels of social identity) than those recruited from Facebook. This could be attributed to MTurk offering an additional layer of anonymity because of its nature as a third-party platform. The researchers did not have access to the participants' personal information. In contrast, the participants recruited from Facebook were actively engaged in social networking activities before responding to the survey questionnaire. Accordingly, they produced a more concrete (and stronger) notion of social identity in response to the scenario.

### 6.2. Implications for Research

Our work is one of the first academic studies to provide a theoretical explanation of the drivers of bystanders' joining in behaviors among people with purely virtual relations. We also explored the role of IT and theorized how it influences deindividuation and moral disengagement. The results advance the theoretical understanding of harmful group behaviors on SNSs in general and have several implications for research on cyberbullying, moral disengagement, and broadly the societal implications of information technology use. We expect this study to have four significant implications for IS research.

First, this is one of the few studies that attempt to bridge the conceptual gaps between IT and deindividuation and delineate the underlying psychological mechanisms that drive bystanders' joining in behaviors. It thus produces interactional theoretical insights that contribute to both the IS and cyberbullying literature (Tarafdar and Davison 2018). Specifically, we show that the two deindividuation experiences derived from the SIDE model are powerful antecedents of bystanders' joining in cyberbullying on SNSs. As the SIDE model also opens to explaining prosocial online group behaviors,

we integrate it with moral disengagement theory to account for the underlying psychological tensions that may need to be bypassed for bystanders to justify their harmful behaviors. Furthermore, by bridging the IT features with the SIDE model, we provide preliminary evidence regarding how IT enables deindividuation. The integrative model, with its rich theorization, sheds light on factors influencing bystanders' joining in cyberbullying and provides insights for future IS research examining how IT inhibits or disinhibits harmful online group behaviors.

Second, we contribute to moral disengagement theory by improving its conceptualization and operationalization. Moral disengagement has often been conceptualized and operationalized as a unidimensional construct. It has been tested as such in prior cyberbullying research (e.g., Thornberg and Jungert 2013). However, our findings suggest that conceptualizing moral disengagement as a multidimensional construct could produce greater insights for formulating behavioral change interventions (DeSmet et al. 2016). Furthermore, an increasing number of studies have posited that the moral disengagement mechanism is triggered by specific circumstances or contextual factors (Moore et al. 2012). For instance, D'Arcy et al. (2014) found that security-related stress (i.e., overload, complexity, and uncertainty) activates moral disengagement to justify information security policy violations. We contribute to the moral disengagement literature by identifying the two deindividuation experiences as antecedents of moral disengagement in the social network context. As a result, it profoundly changes the current explanatory narrative of the antecedent, process, and causal mechanism of moral disengagement.

Third, we contribute to the cyberbullying literature by considering the overlooked scenario in which individuals have purely virtual relations. As information technology has become a central part of everyday life, there are ample opportunities for individuals to meet strangers in virtual spaces and encounter cyberbullying among people whom they do not know well, if at all. Although studies have illustrated the relevance of offline bystander–victim/perpetrator relationships in predicting joining in behaviors, we shed light on the same behavioral manifestation among individuals with purely virtual relations. Our findings, therefore, suggest that bystanders' joining in cyberbullying behaviors can be understood through the alternative lenses of IT-enabled deindividuation experiences and moral

disengagement, where the quality of relationships between the perpetrator, victim, and bystanders is non-exist.

Finally, we examine bystanders' joining in cyberbullying using a scenario-based online survey, in which one of four cyberbullying scenarios was randomly shown to the participants. Given the complex dynamics and interpersonal relationships embedded in a cyberbullying incident, a scenario-based design is well suited for studying bystander behaviors. Such a technique controls the biases and confounding effects associated with recalling past cyberbullying experiences. For instance, this may include the varying intra- and inter-personal characteristics between the perpetrator, victim, and bystanders and the setting and cause of the cyberbullying incident. The scenario technique is preferable to the recall technique for studying complex sociotechnical phenomena when researchers' interest is to delve into the effects of IT and the underlying social psychological mechanisms that drive specific individual and group behaviors. Our cyberbullying scenarios were carefully crafted to represent a common theme of cyberbullying incidents on SNSs and controlled for possible confounding effects. The key steps we took in developing, validating, and applying scenarios in our study could guide future cyberbullying research.

### 6.3. Implications for Practice

Our results suggest two pragmatic measures based on the deindividuation and moral disengagement perspectives that could be taken to reduce the negative behaviors of bystanders on SNSs.

First, our research findings show that two IT features—digital profile and search and privacy— enable experienced anonymity. Hence, platform owners who wish to discourage bystanders from joining in undesirable activities may consider regulating how users could share and access digital resources in a social network. For instance, they may implement measures to permit only users who have registered the service with their real identities to join interest groups. With a link established between their online and offline identities, users will perceive a higher level of accountability for their actions when participating in an online discussion. The emphasis on accountability for their actions should induce greater self-awareness and reduce the tendency to participate in an 'emotional hit and run' during online interactions. As a result, this potential measure prevents users from taking advantage of the platform's anonymous

nature to harass group members, mitigating the unintended impact of IT features in luring harmful social interactions and outcomes on the SNSs.

Second, our research findings show that the two other features—relational ties and network transparency—enable experienced social identity. Hence, platform owners should acknowledge the influence of social identity in igniting, driving, and prolonging detrimental online group behaviors. Although enhancing social interactions between different networks increases popularity and engagement across the site, not all interactions benefit users' well-being or a platform's sustainability. Hence, platform owners who wish to discourage bystanders from joining in undesirable activities may consider regulating how users could develop and maintain their connections in a social network. SNS users who ignite conflicts and abuses on the social network should be restricted by algorithms that limit their visibility or ability to develop connections in the social network. Reducing the spread of malicious content across these connections mitigates the harm to the victim by discouraging like-minded offenders from taking advantage of the platform's connectivity to form alliances and create an imbalance of power that disadvantages the victim. In addition, platform owners can reduce cyberbullying by monitoring what is mentioned and shared between connections and working closely with users who have reported inappropriate posts to moderate localized content.

Third, our findings reveal the detrimental role of moral disengagement mechanisms in justifying bystanders to join in cyberbullying on SNSs. Criminalizing cyberbullying would make it more difficult for bystanders to reconstrue the conduct by minimizing the gray areas in which they can apply euphemistic labeling or palliative comparisons to downplay the inappropriateness of joining in behaviors. Governmental agencies should indicate that users who assist and reinforce cyberbullies will be held accountable for inappropriate online group behaviors to reduce the use of obscuring personal causal agency. The certainty and severity of punishment for any inappropriate behaviors on SNSs should be emphatically articulated to reduce misrepresentation or disregard for injurious consequences. Less punitively, governmental agencies should develop educational programs to promote netiquette and individuals' knowledge of responsible IT use. Educational institutions could offer empathy training

programs to help individuals understand the suffering of cyberbullying victims and mitigate the practice of vilifying the recipients of maltreatment.

### 6.4. Limitations and Future Research Directions

Our study has some limitations that suggest promising research opportunities. First, although our results support an association between the two SNS deindividuation experiences and bystanders' joining in cyberbullying on SNSs, such effects may vary across social network platforms and cyberbullying behaviors. Specifically, the network structure and composition of a particular SNS may alter the relationships between anonymity, social identity, and bystanders' joining in cyberbullying and trigger different moral disengagement mechanisms in specific types of cyberbullying (e.g., harassment, denigration, outing, trickery, and exclusion). Future research is recommended to explore context-specific platform variables unique to each platform (e.g., public vs. private networks and networks with strong vs. weak ties) and examine their moderating effects on joining in cyberbullying behaviors to derive more fine-grained insights. Furthermore, we only measured the users' perceptions of IT features through a scenario study. Future research could adopt an experimental design to examine how features of SNSs induce deindividuation experiences in a more controlled laboratory setting.

Second, we used a scenario survey to test the research model. However, our scenarios only represented a particular cyberbullying situation, in which existing members of an online interest group showed their disapproval of a new member based on their body image. The scenarios also constrained the relationship between perpetrator, victim, and bystanders, such that the bystanders did not have prior interactions with or prior knowledge of the victim. Although such a setting is realistic and common on SNSs, it would be beneficial for future research to create other cyberbullying scenarios, test the bystanders' responses, and identify the boundary conditions of the deindividuation and moral disengagement perspectives. For instance, future research could consider creating a different group dynamic toward the victim among the existing members, introducing various levels of relationship among the perpetrator, victim, and bystanders (e.g., good, bad, and neutral), and manipulating the nature of the online group (e.g., a public group vs. a private group). Future research could also consider motivational factors that influence

bystanders' willingness to join in cyberbullying, such as perceived threats from the bullies if the bystanders decided to go against the group norms or remain silent, or perceived retaliation from the victim if the bystanders chose to join in. Factors that motivate or inhibit one from joining in cyberbullying may be worth further exploration to fully understand the dynamics that influence joining in decisions.

Third, future studies could benefit from using longitudinal surveys and ethnographies to triangulate our findings. A longitudinal survey would allow researchers to understand the long-term effects of IT use on deindividuation, moral disengagement and harmful IT use in general, and a participatory approach, such as ethnography, would allow researchers to understand how the prior and current relationships and interactions between the perpetrator, victim, and bystander affect joining in behaviors.

Fourth, the research model was tested using participants from the United States recruited via MTurk and Facebook. Although such a design ensures consistency in terms of the laws and norms regarding SNS use and cyberbullying, the generalizability of the findings beyond the United States adult population remains uncertain. Future research could test the research model in a broader population and other cultures. For example, given the prevalence of digital access, it is not uncommon for the elderly to experience cyberbullying. Therefore, future research is recommended to empirically test this research model on SNS users of other age groups to evaluate the extent to which the results obtained here can be generalized. Additionally, it would also be worth comparing the effects of anonymity, social identity, and moral disengagement on joining in behaviors in nations with different predominant cultural orientations (e.g., individualism vs. collectivism, masculinity vs. femininity, and indulgence vs. restraint).

## 7. Conclusion

Drawing on the SIDE model, moral disengagement theory, and the social network research framework, we developed and tested a research model explaining what makes bystanders join bullies in attacking a victim in an online interest group on an SNS. Using two samples recruited from MTurk and Facebook, our research model explained a substantial variance in bystanders joining in cyberbullying. Our findings highlight the key role of IT features, deindividuation experiences, and moral disengagement in shaping such harmful online group behavior. The results have significant implications for research on

the adverse and unintended use of information technology and provide pragmatic guidance for platform owners and governmental agencies in enhancing platform sustainability and formulating intervention measures.

**Acknowledgment**

We gratefully appreciate the comments and suggestions of the senior editor, Yulin Fang, the associate editor, J.J. Po-An Hsieh, and the anonymous reviewers throughout this paper's review process. Their excellent inputs have significantly improved the quality of our work. This work was substantially supported by a fellowship award (Project No. HKBU SRFS2021-2H03) and partially supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Nos. HKBU12511016, HKBU12500020).

**References**

ADL. (2020) Online Hate and 2020 Harassment: The American Experience 2020. Accessed 26 July, 2021, https://www.adl.org/media/14643/download.

Akram Z, Khan AG, Akram U, et al. (2022) A Contemporary View of Interpersonal Aggression and Cyberbullying through Ict: Multilevel Insights from Lmx Differentiation. *Internet Research* ahead-of-print(ahead-of-print).

Al-Garadi MA, Hussain MR, Khan N, et al. (2019) Predicting Cyberbullying on Social Media in the Big Data Era Using Machine Learning Algorithms: Review of Literature and Open Challenges. *IEEE Access* 7:70701-70718.

Alleyne E, Fernandes I, and Pritchard E (2014) Denying Humanness to Victims: How Gang Members Justify Violent Behavior. *Group Processes & Intergroup Relations* 17(6):750-762.

Anderson J, Bresnahan M, and Musatics C (2014) Combating Weight-Based Cyberbullying on Facebook with the Dissenter Effect. *Cyberpsychology, Behavior, and Social Networking* 17(5):281-286.

Bandura A (2002) Selective Moral Disengagement in the Exercise of Moral Agency. *Journal of Moral Education* 31(2):101-119.

Bandura A (2016) *Moral Disengagement: How People Do Harm and Live with Themselves* (Worth Publishers).

Bandura A, Barbaranelli C, Caprara GV, et al. (1996) Mechanisms of Moral Disengagement in the Exercise of Moral Agency. *Journal of Personality and Social Psychology* 71(2):364-374.

Bandura A, Underwood B, and Fromson ME (1975) Disinhibition of Aggression through Diffusion of Responsibility and Dehumanization of Victims. *Journal of Research in Personality* 9(4):253-269.

Barlinska J, Szuster A, and Winiewski M (2013) Cyberbullying among Adolescent Bystanders: Role of the Communication Medium, Form of Violence, and Empathy. *Journal of Community & Applied Social Psychology* 23(1):37-51.

Barlinska J, Szuster A, and Winiewski M (2018) Cyberbullying among Adolescent Bystanders: Role of Affective Versus Cognitive Empathy in Increasing Prosocial Cyberbystander Behavior. *Frontiers in Psychology* 9:799-843.

Bastiaensens S, Pabian S, Vandebosch H, et al. (2016) From Normative Influence to Social Pressure: How Relevant Others Affect Whether Bystanders Join in Cyberbullying. *Social Development* 25(1):193-211.

Bastiaensens S, Vandebosch H, Poels K, et al. (2014) Cyberbullying on Social Network Sites: An Experimental Study into Bystanders' Behavioural Intentions to Help the Victim or Reinforce the Bully. *Computers in Human Behavior* 31(February 2014):259-271.
Bauman S, and Newman ML (2013) Testing Assumptions About Cyberbullying: Perceived Distress Associated with Acts of Conventional and Cyber Bullying. *Psychology of Violence* 3(1):27-38.
Behnk S, Hao L, and Reuben E (2017) Partners in Crime: Diffusion of Responsibility in Antisocial Behaviors. *IZA Discussion Papers*:11031.
Borgatti SP, and Foster PC (2003) The Network Paradigm in Organizational Research: A Review and Typology. *Journal of Management* 29(6):991-1013.
boyd DM, and Ellison NB (2007) Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication* 13(1):210-230.
Brody N, and Vangelisti AL (2015) Bystander Intervention in Cyberbullying. *Communication Monographs* 83(1):1-26.
Cao B, and Lin W-Y (2015) How Do Victims React to Cyberbullying on Social Networking Sites? The Influence of Previous Cyberbullying Victimization Experiences. *Computers in Human Behavior* 52:458-465.
Chan TKH, Cheung CMK, and Lee ZWY (2021) Cyberbullying on Social Networking Sites: A Literature Review and Future Research Directions. *Information & Management* 58(2):103411.
Chan TKH, Cheung CMK, and Wong RYM (2019) Cyberbullying on Social Networking Sites: The Crime Opportunity and Affordance Perspectives. *Journal of Management Information Systems* 36(2):574-609.
Chen H-T (2018) Spiral of Silence on Social Media and the Moderating Role of Disagreement and Publicness in the Network: Analyzing Expressive and Withdrawal Behaviors. *New Media & Society* 20(10):3917-3936.
Chen VHH, and Wu Y (2015) Group Identification as a Mediator of the Effect of Players' Anonymity on Cheating in Online Games. *Behaviour & Information Technology* 34(7):658-667.
Cheung CMK, Wong RYM, and Chan TKH (2020) Online Disinhibition: Conceptualization, Measurement, and Implications for Online Deviant Behavior. *Industrial Management & Data Systems*.
D'Arcy J, Herath T, and Shoss MK (2014) Understanding Employee Responses to Stressful Information Security Requirements: A Coping Perspective. *Journal of Management Information Systems* 31(2):285-318.
Darley JM, and Latane B (1968) Bystander Intervention in Emergencies: Diffusion of Responsibility. *Journal of Personality and Social Psychology* 8(4):377-383.
DeSmet A, Bastiaensens S, Van Cleemput K, et al. (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. *Computers in Human Behavior* 57(April 2016):398-415.
Diener E, and Wallbom M (1976) Effects of Self-Awareness on Antinormative Behavior. *Journal of Research in Personality* 10(1):107-111.
El Zein M, Bahrami B, and Hertwig R (2019) Shared Responsibility in Collective Decisions. *Nature Human Behaviour* 3(6):554-559.
Forsberg C, Thornberg R, and Samuelsson M (2014) Bystanders to Bullying: Fourth-to Seventh-Grade Students' Perspectives on Their Reactions. *Research Papers in Education* 29(5):557-576.
Gaudette T, Scrivens R, Davies G, et al. (2021) Upvoting Extremism: Collective Identity Formation and the Extreme Right on Reddit. *New Media & Society* 23(12):3491-3508.
Giner-Sorolla R, Leidner B, and Castano E (2012) Dehumanization, Demonization, and Morality Shifting: Paths to Moral Certainty in Extremist Violence. Hogg MA, Blaylock DL, eds. *Extremism and the Psychology of Uncertainty* (Wiley-Blackwell Boston, MA), 165-182.
Gini G (2007) Who Is Blameworthy?: Social Identity and Inter-Group Bullying. *School Psychology International* 28(1):77-89.

Giumetti GW, and Kowalski RM (2022) Cyberbullying Via Social Media and Well-Being. *Current Opinion in Psychology*:101314.
Glasford DE, Dovidio JF, and Pratto F (2009) I Continue to Feel So Good About Us: In-Group Identification and the Use of Social Identity—Enhancing Strategies to Reduce Intragroup Dissonance. *Personality and Social Psychology Bulletin* 35(4):415-427.
Gosling P, Denizeau M, and Oberlé D (2006) Denial of Responsibility: A New Mode of Dissonance Reduction. *Journal of Personality and Social Psychology* 90(5):722-733.
Hair JJF, Hult GTM, Ringle C, et al. (2017) *A Primer on Partial Least Squares Structural Equation Modeling (Pls-Sem)*, 2nd ed. (Sage Publications, Thousand Oaks).
Harris LT, and Fiske ST (2006) Dehumanizing the Lowest of the Low: Neuroimaging Responses to Extreme out-Groups. *Psychological Science* 17(10):847-853.
Holfeld B (2014) Perceptions and Attributions of Bystanders to Cyber Bullying. *Computers in Human Behavior* 38:1-7.
Hughes R, and Huby M (2002) The Application of Vignettes in Social and Nursing Research. *Journal of Advanced Nursing* 37(4):382-386.
Jones SE, Manstead ASR, and Livingstone AG (2011) Ganging up or Sticking Together? Group Processes and Children's Responses to Text-Message Bullying. *British Journal of Psychology* 102(1):71-71.
Kane GC, Alavi M, Labianca G, et al. (2014) What's Different About Social Media Networks? A Framework and Research Agenda. *MIS Quarterly* 38(1):274-304.
Koehler C, and Weber M (2018) ” Do I Really Need to Help?!” Perceived Severity of Cyberbullying, Victim Blaming, and Bystanders’ Willingness to Help the Victim. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 12(4):Article 4.
Latané B, and Darley JM (1970) *The Unresponsive Bystander: Why Doesn't He Help?* (Appleton-Century-Crofts).
Leidner B, Castano E, Zaiser E, et al. (2010) Ingroup Glorification, Moral Disengagement, and Justice in the Context of Collective Violence. *Personality and Social Psychology Bulletin* 36(8):1115-1129.
Li S, and Zhang G (2018) Intergroup Communication in Online Forums: The Effect of Group Identification on Online Support Provision. *Communication Research* 48(6):874-894.
Longobardi C, Settanni M, Fabris MA, et al. (2020) Follow or Be Followed: Exploring the Links between Instagram Popularity, Social Media Addiction, Cyber Victimization, and Subjective Happiness in Italian Adolescents. *Children and Youth Services Review* 113(June 2020):104955.
Louis WR, Taylor DM, and Douglas RL (2005) Normative Influence and Rational Conflict Decisions: Group Norms and Cost-Benefit Analyses for Intergroup Behavior. *Group Processes & Intergroup Relations* 8(4):355-374.
Lowry PB, Zhang J, Wang C, et al. (2016) 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. *Information Systems Research* 27(4):962-986.
Machackova H, and Pfetsch J (2016) Bystanders’ Responses to Offline Bullying and Cyberbullying: The Role of Empathy and Normative Beliefs About Aggression. *Scandinavian Journal of Psychology* 57(2):169-176.
MacKenzie SB, Podsakoff PM, and Podsakoff NP (2011) Construct Measurement and Validation Procedures in Mis and Behavioral Research: Integrating New and Existing Techniques. *MIS Quarterly* 35(2):293.
Marshall A, and Chan TKH (2021) Cybermobbing Behaviors on Social Networking Sites: A Social Impact and Dual Social Influences Perspective.
McGloin JM, and Thomas KJ (2016) Incentives for Collective Deviance: Group Size and Changes in Perceived Risk, Cost, and Reward. *Criminology* 54(3):459-486.
McHugo GJ, Smith CA, and Lanzetta JT (1982) The Structure of Self-Reports of Emotional Responses to Film Segments. *Motivation and Emotion* 6(4):365-385.

Mooij T (2011) Differences in Pupil Characteristics and Motives in Being a Victim, Perpetrator and Witness of Violence in Secondary Education. *Research Papers in Education* 26(1):105-128.
Moore C, Detert JR, Klebe Treviño L, et al. (2012) Why Employees Do Bad Things: Moral Disengagement and Unethical Organizational Behavior. *Personnel Psychology* 65(1):1-48.
Nagumo J, and Imahashi R. (2020) Hana Kimura's Death Fuels Debate over How to Stop Cyberbullying. *Nikkei Asian Review*, Accessed June 16, 2020, https://asia.nikkei.com/Spotlight/Asia-Insight/Hana-Kimura-s-death-fuels-debate-over-how-to-stop-cyberbullying.
Obermaier M, Fawzi N, and Koch T (2014) Bystanding or Standing By? How the Number of Bystanders Affects the Intention to Intervene in Cyberbullying. *New Media & Society* 18(8):1491-1507.
Patterson LJ, Allan A, and Cross D (2016) Adolescent Bystanders' Perspectives of Aggression in the Online Versus School Environments. *Journal of Adolescence* 49(June 2016):60-67.
Patterson LJ, Allan A, and Cross D (2017) Adolescent Bystander Behavior in the School and Online Environments and the Implications for Interventions Targeting Cyberbullying. *Journal of School Violence* 16(4):361-375.
Pew_Research_Center. (2017) Online Harassment 2017. Accessed 2 September, 2019, https://www.pewinternet.org/2017/07/11/online-harassment-2017/.
Polites GL, Roberts N, and Thatcher J (2012) Conceptualizing Models Using Multidimensional Constructs: A Review and Guidelines for Their Use. *European Journal of Information Systems* 21(1):22-48.
Postmes T, and Spears R (1998) Deindividuation and Antinormative Behavior: A Meta-Analysis. *Psychological Bulletin* 123(3):238-259.
Postmes T, Spears R, and Lea M (1998) Breaching or Building Social Boundaries? Side-Effects of Computer-Mediated Communication. *Communication Research* 25(6):689-715.
Prentice-Dunn S, and Rogers RW (1989) Deindividuation and the Self-Regulation of Behavior. Paulus PB, ed. *Psychology of Group Influence* (Lawrence Erlbaum Associates, Inc, London), 87-109.
RAD. (2018) Online Harassment: Still a Big Problem and Getting Worse. Accessed 1 September, 2019, http://onlineharassmentdata.org.
Reicher S, and Levine M (1994) Deindividuation, Power Relations between Groups and the Expression of Social Identity: The Effects of Visibility to the out-Group. *British Journal of Social Psychology* 33(2):145-163.
Reicher SD, Spears R, and Postmes T (1995) A Social Identity Model of Deindividuation Phenomena. *European Review of Social Psychology* 6(1):161-198.
Robson C, and Witenberg RT (2013) The Influence of Moral Disengagement, Morally Based Self-Esteem, Age, and Gender on Traditional Bullying and Cyberbullying. *Journal of School Violence* 12(2):211-211.
Roccas S, Klar Y, and Liviatan I (2004) *Exonerating Cognitions, Group Identification, and Personal Values as Predictors of Collective Guilt among Jewish-Israelis* (Cambridge University Press, UK).
Runions KC, and Bak M (2015) Online Moral Disengagement, Cyberbullying, and Cyber-Aggression. *Cyberpsychology Behavior and Social Networking* 18(7):400-405.
Salmivalli C (2010) Bullying and the Peer Group: A Review. *Aggression and Violent Behavior* 15(2):112-120.
Sarker S, Chatterjee S, Xiao X, et al. (2019) The Sociotechnical Axis of Cohesion for the Is Discipline: Its Historical Legacy and Its Continued Relevance. *MIS Quarterly* 43(3):695-719.
Sassenberg K (2002) Common Bond and Common Identity Groups on the Internet: Attachment and Normative Behavior in on-Topic and Off-Topic Chats. *Group Dynamics: Theory, Research, and Practice* 6(1):27-37.
Schacter HL, Greenberg S, and Juvonen J (2016) Who's to Blame?: The Effects of Victim Disclosure on Bystander Reactions to Cyberbullying. *Computers in Human Behavior* 57(April 2016):115-121.
Spears R, and Lea M (1994) Panacea or Panopticon? The Hidden Power in Computer-Mediated Communication. *Communication Research* 21(4):427-459.

Spears R, and Postmes T (2015) Group Identity, Social Influence, and Collective Action Online: Extensions and Applications of the Side Model. Sundar SS, ed. *The Handbook of the Psychology of Communication Technology*, vol. 32 (Wiley Online Library, UK), 23-46.
Steer OL, Betts LR, Baguley T, et al. (2020) "I Feel Like Everyone Does It"- Adolescents' Perceptions and Awareness of the Association between Humour, Banter, and Cyberbullying. *Computers in Human Behavior* 108(July 2020):106297.
Suler J (2004) The Online Disinhibition Effect. *CyberPsychology & Behavior* 7(3):321-326.
Tajfel H, and Turner JC (1986) The Social Identity Theory of Intergroup Behaviour. Worchel S, Austin WG, eds. *Psychology of Intergroup Relations* (Nelson-Hall, Chicago, IL), 7-24.
Tarafdar M, and Davison RM (2018) Research in Information Systems: Intra-Disciplinary and Inter-Disciplinary Approaches. *Journal of the Association for Information Systems* 19(6):523-551.
Tepper BJ, Duffy MK, Henle CA, et al. (2006) Procedural Injustice, Victim Precipitation, and Abusive Supervision. *Personnel Psychology* 59(1):101-123.
Thornberg R (2015) School Bullying as a Collective Action: Stigma Processes and Identity Struggling. *Children & Society* 29(4):310-320.
Thornberg R, and Jungert T (2013) Bystander Behavior in Bullying Situations: Basic Moral Sensitivity, Moral Disengagement and Defender Self-Efficacy. *Journal of Adolescence* 36(3):475-483.
Turel O, Matt C, Trenz M, et al. (2019) Panel Report: The Dark Side of the Digitization of the Individual. *Internet Research* 29(2):274-288.
Turner JC, Hogg MA, Oakes PJ, et al. (1987) *Rediscovering the Social Group: A Self-Categorization Theory* (Basil Blackwell).
Twemlow SW, Fonagy P, Sacco FC, et al. (2001) Improving the Social and Intellectual Climate in Elementary Schools by Addressing the Bully-Victim-Bystander Power Struggles. Cohen J, ed. *Caring Classrooms, Intelligent Schools: The Social Emotional Education of Young Children* (Teachers College Press, New York, NY), 162-181.
Valkenburg PM, and Peter J (2011) Online Communication among Adolescents: An Integrated Model of Its Attraction, Opportunities, and Risks. *Journal of Adolescent Health* 48(2):121-127.
Walther JB (2011) Theories of Computer-Mediated Communication and Interpersonal Relations. *The Handbook of Interpersonal Communication* 4:443-479.
Waytz A, and Epley N (2012) Social Connection Enables Dehumanization. *Journal of Experimental Social Psychology* 48(1):70-76.
Weber M, Koehler C, and Schnauber-Stockmann A (2019) Why Should I Help You? Man Up! Bystanders' Gender Stereotypic Perceptions of a Cyberbullying Incident. *Deviant Behavior* 40(5):585-601.
Weber M, Ziegele M, and Schnauber A (2013) Blaming the Victim: The Effects of Extraversion and Information Disclosure on Guilt Attributions in Cyberbullying. *Cyberpsychology, Behavior, and Social Networking* 16(4):254-259.
Wilks T (2004) The Use of Vignettes in Qualitative Research into Social Work Values. *Qualitative Social Work* 3(1):78-87.
Wong RYM, Cheung CMK, Xiao B, et al. (2021) Standing up or Standing By: Understanding Bystanders' Proactive Reporting Responses to Social Media Harassment. *Information Systems Research* 32(2):561-581.
Wright M (2018) Cyberbullying Victimization through Social Networking Sites and Adjustment Difficulties: The Role of Parental Mediation. *Journal of the Association for Information Systems* 19(2):113-123.