Online Moral Disengagement, Cyberbullying, and Cyber-Aggression

Kevin C. Runions, BSc(Hon), BEd, MA, PhD,1,2 and Michal Bak, BSc(Hon)1

Abstract

The study of moral disengagement has greatly informed research on aggression and bullying. There has been some debate on whether cyberbullies and other cyber-aggressors show more or less of a tendency for moral disengagement than traditional aggressors and bullies. However, according to the triadic model of reciprocal determinism, an individual’s behavior influences and is influenced by both personal factors and his/her social environment. This article reviews the literature to propose a new conceptual framework addressing how features of the online context may enable specific mechanisms that facilitate moral disengagement. Specific affordances for moral disengagement proposed here include the paucity of social–emotional cues, the ease of disseminating communication via social networks, and the media attention on cyberbullying, which may elicit moral justification, euphemistic labeling, palliative comparison, diffusion and displacement of responsibility, minimizing and disregarding the consequences for others, dehumanization, and attribution of blame. These ideas suggest that by providing affordances for these mechanisms of moral disengagement, online settings may facilitate cyber-aggression and cyberbullying.

Introduction

When people act maliciously, it is rare for them to acknowledge their actions as bad or evil. Instead, they may offer an assortment of circumstances and contexts beyond their control to explain or justify their behavior. Albert Bandura’s social cognitive theory of moral agency posits that moral behaviors are guided by one’s standards of right and wrong, developed as a part of one’s identity as a moral agent.1 As moral values become integrated and give rise to a moral self, emotional responses act to guide the agent in generating moral behaviors and avoiding those that deviate from his/her own moral standards.2 When individuals’ actions are at odds with their own moral standards, moral disengagement allows them to distance themselves cognitively and avoid aversive emotions, such as guilt and shame.3–5 By switching off their tendencies to judge themselves as morally wrong, specific moral disengagement mechanisms enable individuals to engage in immoral behavior without incurring the negative emotional consequences associated with those violations.

Moral disengagement is coming of age as an important framework for understanding the enactment of morality in behavior and, in particular, aggressive behavior,4,5 traditional bullying,6,7 and cyber-aggressive behavior.8 This research shows that relative to nonaggressive individuals, bullies,9 as well as bystanders who assist bullies or reinforce their behavior by laughing and cheering, display higher levels of moral disengagement. Individuals who victimize others tend to reframe their actions to fit with their own moral codes. Accordingly, bystanders who defend and help victims of bullying have lower levels of moral disengagement.10 Instead of justifying immoral behaviors that threaten an individual’s perception of the self as a moral agent, research suggests that defenders are more likely to act in ways that are consistent with their views of right or wrong. A recent meta-analysis makes clear that the effect size of the relation of moral disengagement to aggressive behavior is at least as large as that of empathy, hostile attribution biases, emotion knowledge, or social competence.11

This article reviews the literature on moral disengagement and the online context to develop a new conceptual framework that aims to account for the differential recruitment of moral disengagement online due to contextual features afforded by online communication technologies. Motivated by recent studies in this area, the framework addresses some important issues encountered by researchers. For example, in their recent meta-analysis, Gini et al.11 highlighted some conceptual confusion of the role that moral disengagement might play in cyber settings. A few studies provide evidence to show that moral disengagement may be less relevant to

1Educational Psychology and Leadership Studies, University of Victoria, Victoria, Canada.
2Telethon Kids Institute, University of Western Australia, Perth, Australia.
cyber-aggression than to offline aggression. However, Pornari and Wood^{12} speculate that online settings may enable aggression due to the ease of acting anonymously, and the perception of greater distance between the aggressor and the target. It is possible that cyber-aggression may not require the same degree of moral disengagement at the individual level, as features of computer mediated communication may play a facilitative role in the use of moral disengagement mechanisms.^{13} The aim is to present a thorough conceptual treatment of how the features and affordances provided by online settings might promote moral disengagement.

The proposed conceptual framework builds upon Bandura’s social cognitive theory, which emphasizes triadic reciprocal determinism—the co-action of behavior, cognition and other personal factors, as well as the social environmental conditions and influences as bidirectional determinants of development.^{14} As Bandura postulated, behaviors vary meaningfully across contexts in ways that can provide insight into understanding the origins and maintenance of those behaviors.^{14,15} Following this line of thinking, examining the features of specific contexts in which problematic behaviors arise^{16} and how those contexts interact with personal predispositions is essential to conceptualizing and addressing those problems fully.

Bauman argues that “the technological world in which youth socialize may be a social context that promotes moral disengagement.”^{17} A handful of recent articles examined the links between moral disengagement and cyber-aggressive behavior using measures of moral disengagement that generalize across contexts.^{12,18,19} In their recent meta-analysis, Kowalski et al.^{20} found that moral disengagement measured in this way shows a moderate association with cyberbullying measures ($r = 0.27$), but this tells us little about how the online social context may promote moral disengagement. Furthermore, there is little discussion about the particular mechanisms of moral disengagement that may be afforded by specific features of cyber-settings.

As a situated process, moral disengagement is seen as dependent not only on the characteristics of the individual, but also on the context in which an individual is acting. Moral disengagement scales have been tailored to specific social settings or activities, such as bullying^{21} and antisocial behaviors in sports.^{22} This article discusses recent conceptual developments that address how the specific structural and functional features of online media provide distinct opportunities for cyber-aggression. Furthermore, it considers how the online setting may pose a set of affordances for moral agency and moral disengagement that is distinct from face-to-face settings. First, Bandura’s specific mechanisms of moral disengagement and the research on moral disengagement in cyberbullying and online aggression are reviewed. Then some of the affordances provided online for moral disengagement and how the mechanisms manifest in the digital world that young people inhabit are discussed.

**Mechanisms of Moral Disengagement**

According to Bandura,^{3} moral disengagement mechanisms allow individuals to justify immoral behavior by reframing any or all four aspects of the situation. In order for individuals to engage in immoral behaviors yet still act in a way that is consistent with their moral standards, they must change the way they view (a) the behavior, (b) the agent’s responsibility, (c) the target of the behavior, and/or (d) the outcomes. Reframing the behavior is accomplished through moral justifications, euphemistic labeling, and/or advantageous comparisons, which enable the individual to view their immoral behavior as ultimately moral or benign. Individuals could also morally disengage by changing their perceptions of their own personal responsibility for the behavior via displacement or diffusion of responsibility to others. Moral disengagement mechanisms that blame the target of the behavior for their own suffering, or dehumanize the target, act to reframe the individual’s perceptions about the target’s role for the behavior. Finally, the harmful outcomes of the immoral behavior can be minimized or viewed as providing benefits for the victim.

These four sets of mechanisms provide a model for understanding how people could frame their negative behavior positively to avoid emotions of guilt and shame, which are associated with moral transgressions. With regard to traditional bullying, Thornberg and Jungert^{21} found that victim blame and moral justification predict bullying behaviors. This work complements Bandura’s experimental manipulations of mechanisms to assess their influence in evoking aggression, demonstrating the role of dehumanization and diffusion of responsibility.^{23} More recently, White-Ajmani and Bursik^{24} found that when diffusion of responsibility, displacement of responsibility, or revenge was primed in a provocative situation, aggressive behaviors were more commonly observed. Of these, victim blame operated uniquely to amplify the association between trait-level moral disengagement and aggressive behavior, suggesting that it is a particularly powerful situational moderator.

As Bandura’s focus on triadic reciprocal determinism would suggest, the role of the social context in which the behavior arises needs to be examined, and the ways that the social context influences the specific moral disengagement mechanisms need to be considered. Pornari and Wood^{12} have begun the process of examining how the online context might support moral disengagement in suggesting that the distance to the victim and the inability to see the victim’s reaction may create an illusion that no harm was inflicted. The online context provides structural affordances for moral disengagement that can potentially increase the use of specific moral disengagement mechanisms, thereby enabling cyber-aggressive behavior. The next section examines the nature of these technological affordances.

**Technological Affordances for Specific Mechanisms of Moral Disengagement**

New portable information and communication technology (ICT)—technologies such as smartphones, which integrate telecommunication and computing, and the platforms for communication and social networking that they enable—has rapidly and thoroughly altered the way people engage with one another. Recently, Runions et al.^{25} provided an overview of how ICT might influence the way children and youth process social information. They proposed ways in which the features of commonly used ICT devices might provide distinct affordances for cyberbullying that may alter the norms for particular behaviors when online. For example, in the offline world, if a student were engaged in surveilling and harassing another student in the middle of the night, this...
would be considered extreme, disturbing, and likely warrant intervention by the authorities. Online, the 24/7 nature of the medium makes it a simple and not exceptional thing to be reading someone’s social media posting, and posting harassing messages to them, in the middle of the night. These affordances may be important for understanding how social context may impact on moral behavior via processes of selectively activating moral disengagement. These distinct aspects of online communication constitute important features that interact with the agent to facilitate acts of cyber-aggression and cyberbullying via mechanisms of moral disengagement.

When users are online, specific features of the ICT context offer opportunities for actions that are different from the opportunities afforded in other social contexts. For example, Halpern and Gibbs demonstrated that in providing easier options for anonymous comments, political discourse (e.g., comments) on YouTube is less polite than the less-anonymous Facebook platform. But what are the specific affordances for moral disengagement that exist for current ICT devices and platforms? How might these new social contexts support moral disengagement differently from offline social contexts? The next section examine the structural and functional features specific to online interactions that may provide unique opportunities for a situated view of moral disengagement.

**Online moral disengagement: disregard and distortion of consequences, and dehumanization**

A key insight that Bandura provided was how much easier it may be to engage in harming others when the harm is invisible to the perpetrator due to distance or time between the harmful act and the consequent harm. A primary feature of most ICT that sharply differentiates it from face-to-face modes of communication is the paucity of social–emotional cues. Although online streaming of video (e.g., Skype) is possible, text-based communication is still the norm. Shifts toward shorter messages and instantaneous correspondences have contributed to a preference for ICT-based communication. But the reliance on text-only communication results in the absence of nonverbal and paralinguistic (e.g., tone of voice, hesitation, etc.) cues. Removing these cues eliminates one of the foundational conditions for the elicitation of empathy in human communication. The physical and temporal distance afforded by ICT communication creates an emotional gap that enables cyber-aggressors to disregard the emotional consequences of aggressive acts.

To understand the role that ICT media may have on this moral disengagement mechanism better, it is important to understand how empathy is activated. Empathy appears to operate via two distinct but related processes, one affective and the other cognitive. Affective empathy mechanisms develop in infancy and operate via affective mirroring and emotional resonance between child and other social agents. There is evidence to show that the brain regions activated during vicarious experiences of pain are similar to those activated when experiencing pain first hand. Similarly, the observation of emotions in others activates the same neuropsychological regions that would be activated during an experienced emotional response. Empathic responding also involves cognitive processes of mentalizing and perspective taking, which develop later in childhood (e.g., theory of mind). Mature empathic responses require the capacity to adopt the perspective of others in a social context.

Under normal conditions, affective and cognitive empathy processes co-act to engage empathic responses. Ruions has argued that the paucity of social cues in ICT media may result in blocking the activation of affective empathic processes. This absence of online triggers for affective empathy may leave potential cyber-aggressors with only the cognitive perspective-taking route available. This may require more cognitive computation on the part of the empathizer than does the affective route. For children and youth who are still developing their perspective-taking abilities, this demand on their fledgling perspective-taking skills may be too great, and they may be more likely to fail in taking perspective, and thus in experiencing empathy.

Building on Pornari and Wood’s insight, in the absence of social cues, cyber-aggressors may not have adequate social information for accurate estimations of harm. Instead, they project their own interpretation of their action onto the victim, thereby distorting the consequences of their acts. Many youth report that they interpret cyber-aggression as a joke. In cases where intent to harm and repetition can be established, the ease of the “joking” label would appear to clearly indicate a process of active moral disengagement via euphemism.

In discussing this, however, it is important to note that moral disengagement may not be at play when young people cite joking as a reason for their cyber-aggressive behavior. It is possible, under certain online circumstances, that reports of cyber-aggression as jokes reflect the perpetrator’s genuine lack of intent to harm, and are not simply excuses for bad behavior. In the absence of emotional cue information resulting from one’s own aggression, cyber-aggressors may believe that the victim also accepts the aggressive act as a joke. In this sense, the cyber-aggressor may be literally unable to engage with the social information required to recognize the unintended harm. As Bandura has argued, external triggers provide a crucial rout for activation of self-regulatory mechanisms and can only become active if they are triggered. Moral self-regulation in Bandura’s model requires people to be able to monitor their behavior in relation to the their setting and circumstances. Online settings provide little or no nonverbal or paralinguistic cues, and, by this aspect of their structure, may prevent this self-monitoring, such that cyber-victimization takes place in the absence of intent to harm on the aggressor’s part.

Moreover, by stripping communication of emotional content, use of ICT media may be necessarily structurally dehumanizing. As Bandura wrote, “[t]o perceive another as human activates empathic reactions through perceived similarity.” By removing emotion cues that operate via evolutionarily old pathways, there is a risk of creating contexts where dehumanization can occur by design. This may help account for the normativity of hostile communications online.

**Online moral disengagement: attribution of blame online**

The absence of nonverbal or paralinguistic social or emotional cues to guide intention interpretations may have other consequences for online situated moral disengagement. From
social information processing research, it is known that ambiguous social cues are prone to being misinterpreted as reflecting malicious intent, and that such an interpretation can motivate aggression. These tendencies to interpret ambiguous social cues as hostile may affect children generally across their social relationships or may be specific to particular relationships. The absence of clear intent cues in e-mail and other ICT media appears to increase interpretations of aggressive intent. Thus, cyber-communication may promote increased hostile attributions of intent, triggering attributions of blame. Ambiguous communications, common on social media, may provoke self-justifications of cyber-aggression as retaliatory responses, with responsibility for the perceived provocation attributed to the other person; one’s own behavior is perceived as merely a reaction. These processes may be particularly pertinent to aggressive-victims (i.e., bully-victims), whose responses to perceived provocation fuel cycles of bullying.

Another feature of ICT-based communication that may further promote moral disengagement is the fact that digital data are potentially permanent. Messages or images shared online have the potential to exist in perpetuity without any degradation or alteration from time or memory effects. This feature of online communication may fuel rumination, as a (self-perceived) victim may revisit the provocative communication at their leisure, maintaining and potentially exacerbating their anger and rage-induced cyber-aggression. Thus, moral disengagement via attribution of blame may operate not just as an immediate reaction, but also over the long term, and structural affordances of ICT-media may help to sustain that moral disengagement.

Finally, it is worth noting that blame may override empathic response. An IMRI study shows that attitudes of blame for a victim have been shown to decrease feelings of empathy for the victim’s physical pain. Although a young research field, this evidence for blame being sufficient to override empathic responses provides an initial mechanism by which to account for the moral disengagement that results from blame. By creating conditions of ambiguity that increase the likelihood of (mis)attributions of blame, ICT-media may fuel moral disengagement that may afford cyber-aggression.

Online moral disengagement: advantageous comparison in an age of viral media

The final process examined in this article involves the media window into the lives of young people. It is widely acknowledged that the current generation of young people is growing up immersed in a technologically advanced world that is quite unlike that of previous generations, and thereby likely exposed to more media than previous generations.

Meanwhile, media attention to bullying has grown dramatically in recent years. Several high-profile suicide cases of recent years have drawn attention to issues of bullying and cyberbullying. It is not unreasonable to hope that the media attention paid to cyberbullying, trolling, and other forms of cyber-aggression are providing a level of exposure to the potential for harm that would make any moral justification for cyber-aggression untenable. With the media coverage of suicides by Rebecca Ann Sedwick in the United States and Rehtaeh Parsons in Canada, for example, one might hope that this would serve a valuable educational function and warning to young people who might otherwise engage in cyberbullying.

There is a risk that this media attention backfires. By highlighting extreme cases, the media environment in which young people are living may afford moral disengagement via advantageous comparison. Bandura has noted that one’s own self-evaluation depends in part on what that behavior is compared. With social media and traditional media outlets paying increased attention to extreme and tragic cases of cyberbullying, the vividness of those behaviors may increase for young people, potentially providing an easy comparison point against which to minimize the severity of one’s own behaviors. In the absence of other sources of information about less extreme cases, the vividness of the media might afford some young people engaging in online aggression to frame their own cyber-aggression as benign, relative to the extreme cases in the media. In effect, the focus on extreme cases might make it easy for children to conclude that their online behavior does not resemble the media’s portrayal of cyberbullying. To the authors’ knowledge, no research to date has focused on this admittedly speculative hypothesis and asked youth about their
behaviors in relation to extreme cases. Qualitative studies may be the best place to start in investigating this potential mechanism of moral disengagement.

We are living in the midst of a natural experiment around cyber-aggression. Can the eradication of one mechanism for moral disengagement (i.e., moral justification) override the affordances provided by ICT and the current media environment to result in a societal reduction in bullying? An empirical response to this question will require a consideration of mechanisms of moral disengagement. This article has aimed to provide some preliminary conceptual directions for addressing this question. In outlining how moral disengagement may be differentially recruited by affordances available within the cyber-context, compared to other contexts, the aim is to spur research on at a process level.

Conclusions

Perren and Gutzwiller-Helfenfinger suggest that research address situation-specific features of cyber-aggression to examine more closely how moral disengagement might be promoted. This article aims to provide a conceptual framework for such research. Furthermore, the collective understanding of the role of moral disengagement in cyber-aggression and cyberbullying requires specific attention to what affordances are provided by the specific social context in which aggression and/or bullying arises. Furthermore, to understand MD, it is necessary to understand how it is that moral behavior occurs online in spite of these affordances and the social influence they constitute. This article is a small step toward interrogating the nature of the social influences in the online context, itself only one facet of Bandura’s concern for studying the “reciprocal interplay of cognitive, affective and social influences” that help determine whether people engage in cyber-aggression and cyberbullying. For cyber-aggression and cyberbullying, this means focusing on how young people interact with the technologies to interact with one another.

By examining the socio-technological context of cyber-aggression, the hope is to inspire new research that can study the affordances that new technologies may provide for moral disengagement. The ideas introduced here may inform the development of online context-specific measures of moral disengagement. Perhaps even more fruitful would be for these ideas to inform experimental studies following the tradition of Bandura et al. that manipulate aspects of the online setting to determine the impact on moral decision making and behavior. Identifying whether and how the setting may afford selective activation of mechanisms of moral disengagement should also be of value for cyber-safety programs. From such research, it may be possible to be better positioned to plan cyberbullying prevention programs that address context-specific processes. Understanding how the online setting may differentially afford moral disengagement could be a key step in planning and testing intervention and prevention approaches to support cyber-safety and reduce cyber-aggression.

Notes

a. Recent work has emphasized the need to distinguish bullying conducted online from offline aggression. Features of online communication such as the capacity for others to share a single act of aggression easily complicates some of these definitional criteria of bullying. In general, this article address how aspects of the online medium may recruit moral disengagement to produce cyber-aggression.

b. The moral disengagement mechanisms are presented in an order that diverges from Bandura’s usual sequence. The specific online processes that are introduced here sometimes operate upon multiples of these processes, and analysis is presented based on the relative influence by the online features on moral disengagement mechanisms.

Author Disclosure Statement

No competing financial interests exist.

References


18. Perren S, Sticca F. (2011) Bullying and morality: are there differences between traditional bullies and cyberbullies? Society for Research in Child Development Biennial Meeting, Montreal, QC.


41. Friedman RA, Currall SC. Conflict escalation: dispute exacerbating elements of e-mail communication. Human Relations 2003; 56:1325–1347.


Address correspondence to:
Dr. Kevin Runions
Telethon Kids Institute
100 Roberts Rd.
University of Western Australia
Subiaco
Perth
Western Australia 6008
Australia

E-mail: kevin.runions@telethonkids.org.au
Copyright of CyberPsychology, Behavior & Social Networking is the property of Mary Ann Liebert, Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.