The Role of Nudges in Mitigating and Preventing Cyberbullying on Social Media

Jiayue Mao*

Rice University, Houston, TX 77005
**Corresponding author. Email: Jiayue.mao@rice.edu*

ABSTRACT

The prevalence of social media (e.g., Instagram, TikTok) heavily influences the social networks of teenagers and students. Recently, the popularity of “spill the tea” accounts, where account owners publish gossip about their classmates and friends, leads to a new form of cyberbullying. This manner, in particular, increases the victim’s vulnerability as social media penetrates their personal lives profoundly. Thus, this paper aims to examine methods to design and alter features of social media or web applications to help avoid such new forms of cyber-harassment among adolescents. The paper employs a qualitative approach by examining previous studies and cases to investigate the rationales behind creating, following, and witnessing gossip accounts. Insights from behavioral economics are provided for practitioners and platforms in hope of preventing cyberbullying. The article emphasizes concepts such as priming effect, status quo bias, intertemporal choice, framing effect, and the Prospect Theory. This paper will hopefully shed light on ways to create an innocuous and friendly online environment for teens’ growth and mental health.

Keywords: *Nudges, Cyberbullying, Teenage, Online behaviors, Behavioral economics*

1. INTRODUCTION

With the launching of smartphones swaying new trends among schoolers, the usage of social media seems like an innate skill that teenagers naturally gain when they receive their first iPhone. Based on a 2018 survey, 95% of the teens from the United States had access to a smartphone; 72% and 69% of the teens used Instagram and Snapchat respectively [1]. However, besides the advantages of connecting with family members and finding new friends, 24% of the teens surveyed expressed negative emotions towards social media; 27% of them claimed bullying or rumor spreading as the main reason for the uneasiness brought by social media [1]. Indeed, cyberbullying is never a half-formulated word. Defined as “[involving] electronic communications directed from the perpetrator to the victim only” [2], cyberbullying used to be conducted through manners like phone calls, emails, or text messages in the early decades of the 21st century.

Yet, in the context of social media, like Instagram and TikTok, cyberbullying is shown in distinct forms of hatred comments under posts or the creation of gossip accounts, which are steadily gaining popularity during 2021. These accounts often have noticeable school names followed by “tea” to attract followers from the institution, and account owners (usually also teenage students from the school) publish gossip about their classmates [3]. These accounts are not an innocuous space for students to exchange intimacy but rather become the new shelter for attacking attires of disliked classmates or spreading fake rumors about a friend’s personal life. Therefore, this paper uses insights from behavioral economics, a subfield of economics that accounts for psychological, cognitive, and emotional factors in human behaviors [4], to understand users’ motivations. Possible nudges, or “any aspect of the choice architecture that alters people’s behavior in a predictable way without forbidding any options or significantly changing their economic incentives” [8], are examined to investigate how one can design or alter social media to avoid such recent forms of cyberbullying. Prosocial behaviors can be elicited through nudges, which successfully reduce risky leftover medications at homes and the amount of downloaded malware [18, 19].

Much attention has been given to cyberbullying mitigation through regulatory acts, yet few have examined the use of nudges in mitigating adolescents’ online harassment, especially this new form of cyberbullying through gossip accounts. This paper aims to fill this gap by applying qualitative analysis on survey data and previous studies. Section II discusses the motivations behind creating such accounts and investigates reasons for following the gossip accounts and strategies to discourage such behaviors. Section III focuses on potential solutions for encouraging bystander intervention, including students, parents, and the social media platforms, and their abilities to maintain affected students' mental health.

2. RATIONALES BEHIND ACCOUNT CREATORS AND FOLLOWERS THROUGH PSYCHOLOGY

Compared with traditional bullying, cyberbullying through social media accounts involves a more intense level of publicity since any user who is aware of the gossip accounts has access to the content. Thus, online harassment has greater damage because it affects the victims' entire social environment and reduces victims' ability to control, and the sense of helplessness makes victims more vulnerable [5]. It is, therefore, crucial to scrutinize both account creators' and followers' motivations to mitigate this effect.

2.1 Account owners' abuse of online platforms' publicity, anonymity, and medium

As punishment and revenge constitute two of the commonest reasons for cyberbullying (students who are annoyed by the victims or have tiffs with the victims previously tend to extend their detestation to online platforms) [6], account owners take advantage of the publicity aspect to create more powerful and persisting harm to the victim. In this case, mechanisms that increase decision points in making a negative choice (i.e., post a hatred comment or publish fake information) help reduce unwanted behaviors [7]. Once the platform has identified a gossip account, it may ask the account creator to provide additional verification steps when sending a post. For instance, users need to provide a security code sent to their phone number or email address before publishing a post. The platform can also ask account owners to undergo a robot check, usually dragging a piece of the puzzle to complete the picture or selecting all the chimneys in the set of given pictures. In this case, users have more time to reconsider the content of their posts and have a higher chance to edit or delete the unfriendly content.

Many perpetrators also take advantage of the anonymity of the online environment. As some owners create gossip accounts solely for fun [6], the anonymity feature reduces the fear of receiving punishment or judgment from their surroundings. This feature is especially attractive for spill-the-tea accounts. Thus, it is vital to remind owners to take responsibility for their speeches and help them recognize the negative influence of their actions. The priming effect, which claims that "subtle influences can increase the ease with which certain information comes to mind" [8], can be used to increase perpetrators' upbeat emotions and awareness of social transparency, the feeling of being "watched" by others. For example, in the draft section, a positive message prompt, like "sharing a delightful moment of your day", may nudge users towards optimism and be less likely to share injurious messages. Showing the number of potential viewers can also remind users of the harmful influence of their hatred posts (i.e., providing a message of "You are sharing this message with 72 people"). Several studies show a strong positive causal relationship between view notification and social transparency, as well as the indirect effect on accountability [9].

The third motivation for creating a gossip account is exploiting the specialty of the medium. As over 70% of teenage schoolers use Instagram, the penetration of the platforms in students' social circles is immense. Yet only 48% of adults between age 30 and 49 ever used Instagram, and such percentage decreases to 29% for ages 50 to 64, based on a 2021 survey [12]. Such discrepancy may make seeking help from adults less probable, and thus perpetrators sense greater control and more freedom in the online sphere. Perpetrators may feel more confident in getting away with the harassment because people tend to assign less weight to intermediate and high probability events according to the Prospect Theory [14]. Displaying a warning message in terms of frequency (e.g., "one out of four offenders is suspended from the school") may help users correctly identify the risk since it is easier to understand and interpret frequency terms [15]. Therefore, perpetrators are less inclined to post humiliating texts or photos online.

2.2 Followers' appeal to publicity, conformity, and anonymity

Account followers are also a crucial part of reducing cyberbullying since creators, for fun or punishment, need public attention and support when judging or degrading the victim, and the publicity aspect fulfills this perfectly. Thus, by decreasing exposure of the posts and the number of viewers by discouraging followers, creators will be unable to gain the level of popularity and publicity as they expected. Furthermore, the harm induced by cyberbullying could also be minimized as the influence of the post is restrained to a smaller group of victims' social networks.

One motivation to follow the gossip accounts is peer pressure, where students yield to friends or classmates who participate in gossip spreading. Based on the Prospect Theory, loss frames that focus on costs trigger response more effectively than gain frames [10]. Thus, a message framed in terms of the number of friends who do not follow the gossip (e.g., "You have 37 connections who do NOT follow this account") can reduce the chance of following. Especially for platforms like Instagram, where people use real names to connect with friends, students are empowered to find allies and thus become less inclined to conform.

Like the motivation behind creating the account, the anonymity aspect of certain platforms attracts followers since they are free to express thoughts or comments on victims' behaviors and appearances. Hatred comments under the posts are another feature of cyberbullying [11]. Strategies to increase accountability can be applied here. For instance, the platform may encourage users to upload a profile photo by default setting (i.e., status quo bias) or link their accounts to other social media platforms. As users recognize their comments may be viewed by their close friends or family members, they will be more reserved and thoughtful in expressing ideas. Indeed, including a profile picture is associated with better accountability since people feel identified with the account [10]. As users are reminded of potential judgment from their acquaintances, they become more aware of their actions and thus less likely to spread hatred comments.

3. IMPLICATIONS AND APPLICATIONS FROM BEHAVIORAL ECONOMICS THOUGH BYSTANDERS

Bystanders, or parties who do not actively participate in cyberbullying activities, play a vital role in reducing online harassment. Bystanders often include students, parents, teachers, and moderators. For students who witness the incident, two common reasons for not reporting to parents or the school are fear of getting into trouble (e.g., becoming the new target of perpetrators) and disbelief in the effectiveness of reporting [13]. Thus, the social platform can be a useful tool where students seek help. The first concern can be addressed through the platform design, including providing or emphasizing the anonymity of the report function, as studies show that anonymity directly influences users' perceived risks and trust in the platform so that users are more likely to flag sensitive content [16]. Visual cues, including designing the flag button (i.e., report button) with vibrant color to make it more visible, may also encourage students to flag offensive photos. Providing specific multiple choices for reasons for the report instead of asking users to write the reasons could make the reporting process simpler. By decreasing the decision points, consumers' rate of successfully filing the report can be increased. For the second concern regarding outcome effectiveness, it is vital to help students recognize the social media policies and entitled rights to build confidence in reporting [17]. The social media platform can display tips that show the number and the time of processing the reports. Platforms can also send confirmation messages and emails to show their active endeavor in handling the information.

For students without such two concerns, nudges that provoke empathy can be implemented to increase bystander intervention. For hatred comments, view notification (i.e., users recognize that their view of the post is received by the creator) is an effective tool in reducing bystander apathy as people feel more responsibility to help [20]. Building upon previous cases, increasing social transparency is the key component in designing nudges since awareness of one's action increases accountability [9, 20]. Similarly, for gossip accounts, this strategy can be adapted so that users' followers can see users' view history so that the users feel more involved as part of the problem.

Parental monitoring of youth's computer use and online navigation can be a useful tool in controlling students who intend to engage in cyberbullying. Yet, for parents, the main obstacle for intervention may be contributing to inattentiveness and unawareness [21]. Some parents may simply be too busy to heed youths' online activities, and others are unfamiliar with the features of different social media platforms. For the first case, nudges can be implemented by changing the default setting so that parents could monitor automatically and with ease. For instance, when students register for a new account, their account will be linked to the parent's account, or weekly reports will be sent to the parents' work email address by default. The status quo bias indicates that people tend to stick with the default setting [22]. Thus, parents can receive recent updates from their children promptly and choose to intervene if necessary. For parents who are unfamiliar with social media use, it is vital to investigate which part of the platform is confusing. For example, if parents are unfamiliar with the popular memes or trending vocabularies, regular exposure to new terms through news notification or feeds may be useful in learning [23].

For moderators and the platform, many strategies have already been implemented. For example, Instagram applies artificial intelligence to capture sensitive words and provides warning messages to users [24]. Such a strategy is particularly useful since delaying actions gives users a chance to reconsider the content of the message instead of directly prohibiting them from sending the message [25]. This method can be more specifically targeted at teenage students by expanding the list of sensitive words by incorporating vines and memes and checking for prerogative content in pictures and hashtags. Moreover, adding decision points can decrease unwanted behavior [7]. Specially targeted at gossip accounts, the process of creating posts or pictures can be complicated by adding extra verification steps as previously mentioned.

4. CONCLUSION

Cyberbullying has emerged as one of many troubling social issues in recent decades and plagues many teenagers in school. It influences students' social networks and peer groups. In particular, the advent of the spill-the-tea accounts not only exposes privacy threats to youths' personal lives but also facilitates cyberbullying and enlarges its scope and effect. Thus, this paper analyzes the underlying factors for this phenomenon through three different lenses: motivations behind account creators, motivations behind account followers, and the role of bystanders. Account creators often use gossip accounts to entertain, take revenge, or punish peers they dislike. Thus, several nudges that increase their accountability and awareness of potential punishment are provided in the article. Account followers often yield to peer pressure or curiosity, and their action of following the account enables account owners to extend their influence and further add pressure to victims' social groups. Finally, the role of bystander intervention is also crucial in mitigating cyberbullying. Suggestions for encouraging actions of witnessing students, parents, and moderators are modified with behavioral economic insights.

The paper provides new angles on how platforms may adopt prevention strategies to improve user experience and create a benign ambient for teens' growth to maintain healthy online behaviors. Although in the article, solutions targeting different parties' psychological and social motivations of involvement in gossip spreading are provided, there has not been extensive empirical evidence in support of the effectiveness of the suggestions. In the future, more experiments to analyze the impact of delayed responses, extended decision points, and framing of warning messages are strongly needed. Especially, the implementation of nudges should be checked preferably with teenage participants since they may have different ways of processing information and are experts in using social media. As more data is gathered, other motivations may be unfolded to revise and improve the nudge models and theories.

REFERENCES

[1] Anderson, Monica, and Jingjing Jiang. "Teens, Social Media & Technology 2018." Pew Research Center: Internet, Science & Tech, Pew Research Center, 31 May 2018, https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/.

[2] Colette Langos.Cyberpsychology, Behavior, and Social Networking.Jun 2012.285-289.http://doi.org/10.1089/cyber.2011.0588

[3] Jargon, Julie. "Spilling the Tea, the Cyberbullying Tactic Plaguing Schools, Parents and Students." The Wall Street Journal, Dow Jones & Company, 18 Dec. 2021, https://www.wsj.com/articles/spilling-the-tea-the-cyberbullying-tactic-plaguing-schools-parents-and-students-11639836002.

[4] Teitelbaum, Joshua and Zeiler, Kathryn, "Research Handbook on Behavioral Law and Economics" (2018). Books. 35. https://scholarship.law.bu.edu/books/35

[5] Sticca F, Perren S. Is cyberbullying worse than traditional bullying? Examining the differential roles of medium, publicity, and anonymity for the perceived severity of bullying. J Youth Adolesc. 2013 May;42(5):739-50. doi: 10.1007/s10964-012-9867-3. Epub 2012 Nov 27. PMID: 23184483.

[6] Wang, CW., Musumari, P.M., Techasrivichien, T. et al. "I felt angry, but I couldn't do anything about it": a qualitative study of cyberbullying among Taiwanese high school students. BMC Public Health 19, 654 (2019). https://doi.org/10.1186/s12889-019-7005-9

[7] Cheema, Amar, and Dilip Soman. "The Effect of Partitions on Controlling Consumption." Journal of Marketing Research, vol. 45, no. 6, Dec. 2008, pp. 665–675, doi:10.1509/jmkr.45.6.665.

[8] Thaler, Richard H., and Cass R. Sunstein. Nudge: Improving Decisions Using the Architecture of Choice. Yale University Press, 2008.

[9] Samuel Hardman Taylor, Dominic DiFranzo, Yoon Hyung Choi, Shruti Sannon, and Natalya N. Bazarova. 2019. Accountability and Empathy by Design: * Encouraging Bystander Intervention to Cyberbullying on Social Media. Proc. ACM Hum.-Comput. Interact. 3, CSCW, Article 118 (November 2019), 26 pages. https: //doi.org/10.1145/3359220

[10] Tversky, Amos, and Daniel Kahneman. "Rational Choice and the Framing of Decisions." The Journal of Business, vol. 59, no. 4, University of Chicago Press, 1986, pp. S251–78, http://www.jstor.org/stable/2352759.

[11] Indrawan, Fani. "IMPOLITENESS STRATEGY IN INSTAGRAM CYBERBULLYING: A CASE STUDY OF JENNIFER DUNN POSTED BY @LAMBE_TURAH." ETNOLINGUAL [Online], 2.1 (2018): 1-19. Web. 10 Jan. 2022

[12] Auxier, Brooke, and Monica Anderson. "Social Media Use in 2021." Pew Research Center: Internet, Science & Tech, Pew Research Center, 9 Apr. 2021, https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/.

[13] Huang, Yun-yin, and Chien Chou. "An Analysis of Multiple Factors of Cyberbullying among Junior High School Students in Taiwan." Computers in Human Behavior, vol. 26, no. 6, 19 July 2010, pp. 1581–1590., https://doi.org/10.1016/j.chb.2010.06.005.

[14] Tversky, A., Kahneman, D. Advances in prospect theory: Cumulative representation of uncertainty. J Risk Uncertainty 5, 297–323 (1992). https://doi.org/10.1007/BF00122574

[15] Burkell, Jacquelyn. "What are the chances? Evaluating risk and benefit information in consumer health materials." Journal of the Medical Library Association : JMLA vol. 92,2 (2004): 200-8.

[16] Lowry, Paul Benjamin, et al. "The Drivers in the Use of Online Whistle-Blowing Reporting Systems." Journal of Management Information Systems, vol. 30, no. 1, 2013, pp. 153–190., https://doi.org/10.2753/mis0742-1222300105.

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

[18] Terri Voepel-Lewis, Frances A. Farley, John Grant, Alan R. Tait, Carol J. Boyd, Sean Esteban McCabe, Monica Weber, Calista M. Harbagh, Brian J. Zikmund-Fisher; Behavioral Intervention and Disposal of Leftover Opioids: A Randomized Trial. Pediatrics January 2020; 145 (1): e20191431. 10.1542/peds.2019-1431

[19] Petrykina, Yelena, et al. "Nudging Users towards Online Safety Using Gamified Environments." Computers & Security, vol. 108, 29 May 2021, p. 102270., https://doi.org/10.1016/j.cose.2021.102270.

[20] Dominic DiFranzo, Samuel Hardman Taylor, Franccesca Kazerooni, Olivia D. Wherry, and Natalya N. Bazarova. 2018. Upstanding by Design: Bystander Intervention in Cyberbullying. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, Paper 211, 1–12. DOI:https://doi.org/10.1145/3193574.3173785

[21] David-Ferdon C, Hertz MF. Electronic media, violence, and adolescents: an emerging public health problem. J Adolesc Health. 2007 Dec;41(6 Suppl 1):S1-5. doi: 10.1016/j.jadohealth.2007.08.020. PMID: 18047940.

[22] Samuelson, W., Zeckhauser, R. Status quo bias in decision making. J Risk Uncertainty 1, 7–59 (1988). https://doi.org/10.1007/BF00055564

[23] Beale, Andrew & Hall, Kimberly. (2007). Cyberbullying: What School Administrators (and Parents) Can Do. The Clearing House. 81. 8-12. 10.3200/TCHS.81.1.8-12.

[24] Steinmetz, Katy. "Inside Instagram's Ambitious Plan to Fight Bullying." Time, Time, 30 Apr. 2021, https://time.com/5619999/instagram-mosseri-bullying-artificial-intelligence/.

[25] Levit, T., Cismaru, M. Marketing social marketing theory to practitioners. Int Rev Public Nonprofit Mark 17, 237–252 (2020). https://doi.org/10.1007/s12208-020-00245-4