An Investigation on the Portrayal of Blue Whale Challenge on YouTube and Twitter

Amro Khasawneh[1], Kapil Chalil Madathil[1,2], Emma Dixon[4], Pamela Wisniewski[5], Heidi Zinzow[3], Rebecca Roth[3]

[1]Department of Industrial Engineering, [2]Department of Civil Engineering, & [3]Department of Psychology, Clemson University
[4]Department of Information Studies, University of Maryland, College Park
[5]Department of Computer Science, University of Central Florida

INTRODUCTION

Social media has created opportunities for children to gather social support online (Blackwell et al., 2016; Gonzales, 2017; Jackson, Bailey, & Foucault Welles, 2018; Khasawneh, Rogers, Bertrand, Madathil, & Gramopadhye, 2019; Ponathil, Agnisarman, Khasawneh, Narasimha, & Madathil, 2017). However, social media also has the potential to expose children and adolescents to undesirable behaviors. Research showed that social media can be used to harass, discriminate (Fritz & Gonzales, 2018), dox (Wood, Rose, & Thompson, 2018), and socially disenfranchise children (Page, Wisniewski, Knijnenburg, & Namara, 2018). Other research proposes that social media use might be correlated to the significant increase in suicide rates and depressive symptoms among children and adolescents in the past ten years (Mitchell, Wells, Priebe, & Ybarra, 2014). Evidence based research suggests that suicidal and unwanted behaviors can be promulgated through social contagion effects, which model, normalize, and reinforce self-harming behavior (Hilton, 2017). These harmful behaviors and social contagion effects may occur more frequently through repetitive exposure and modelling via social media, especially when such content goes "viral" (Hilton, 2017).

One example of viral self-harming behavior that has generated significant media attention is the Blue Whale Challenge (BWC). The hearsay about this challenge is that individuals at all ages are persuaded to participate in self-harm and eventually kill themselves (Mukhra, Baryah, Krishan, & Kanchan, 2017). Research is needed specifically concerning BWC ethical concerns, the effects the game may have on teenagers, and potential governmental interventions. To address this gap in the literature, the current study uses qualitative and content analysis research techniques to illustrate the risk of self-harm and suicide contagion through the portrayal of BWC on YouTube and Twitter Posts. The purpose of this study is to analyze the portrayal of BWC on YouTube and Twitter in order to identify the themes that are presented on YouTube and Twitter posts that share and discuss BWC. In addition, we want to explore to what extent are YouTube videos compliant with safe and effective suicide messaging guidelines proposed by the Suicide Prevention Resource Center (SPRC).

METHOD

Two social media websites were used to gather the data: 60 videos and 1,112 comments from YouTube and 150 posts from Twitter. The common themes of the YouTube videos, comments on those videos, and the Twitter posts were identified using grounded, thematic content analysis on the collected data (Padgett, 2001). Three codebooks were built, one for each type of data. The data for each site were analyzed, and the common themes were identified. A deductive coding analysis was conducted on the YouTube videos based on the nine SPRC safe and effective messaging guidelines (Suicide Prevention Resource Center, 2006). The analysis explored the number of videos that violated these guidelines and which guidelines were violated the most. The inter-rater reliabilities between the coders ranged from 0.61 – 0.81 based on Cohen's kappa. Then the coders conducted consensus coding.

RESULTS & FINDINGS

Three common themes were identified among all the posts in the three social media platforms included in this study. The first theme included posts where social media users were trying to raise awareness and warning parents about this dangerous phenomenon in order to reduce the risk of any potential participation in BWC. This was the most common theme in the videos and posts. Additionally, the posts claimed that there are more than 100 people who have played BWC worldwide and provided detailed description of what each individual did while playing the game. These videos also described the tasks and different names of the game. Only few videos provided recommendations to teenagers who might be playing or thinking of playing the game and fewer videos mentioned that the provided statistics were not confirmed by reliable sources.

The second theme included posts of people that either criticized the teenagers who participated in BWC or made fun of them for a couple of reasons: they agreed with the purpose of BWC of *"cleaning the society of people with mental issues,"* or they misunderstood why teenagers participate in these kind of challenges, such as thinking they mainly participate due to peer pressure or to "show off".

The last theme we identified was that most of these users tend to speak in detail about someone who already participated in BWC. These videos and posts provided information about their demographics and interviews with their parents or acquaintances, who also provide more details about the participant's personal life.

The evaluation of the videos based on the SPRC safe messaging guidelines showed that 37% of the YouTube videos met fewer than 3 of the 9 safe messaging guidelines. Around

50% of them met only 4 to 6 of the guidelines, while the remaining 13% met 7 or more of the guidelines.

DISCUSSION

This study is the first to systematically investigate the quality, portrayal, and reach of BWC on social media. Based on our findings from the emerging themes and the evaluation of the SPRC safe messaging guidelines we suggest that these videos could contribute to the spread of these deadly challenges (or suicide in general since the game might be a hoax) instead of raising awareness. Our suggestion is parallel with similar studies conducted on the portrait of suicide in traditional media (Fekete & Macsai, 1990; Fekete & Schmidtke, 1995). Most posts on social media romanticized people who have died by following this challenge, and younger vulnerable teens may see the victims as role models, leading them to end their lives in the same way (Fekete & Schmidtke, 1995). The videos presented statistics about the number of suicides believed to be related to this challenge in a way that made suicide seem common (Cialdini, 2003). In addition, the videos presented extensive personal information about the people who have died by suicide while playing the BWC. These videos also provided detailed descriptions of the final task, including pictures of self-harm, material that may encourage vulnerable teens to consider ending their lives and provide them with methods on how to do so (Fekete & Macsai, 1990). On the other hand, these videos both failed to emphasize prevention by highlighting effective treatments for mental health problems and failed to encourage teenagers with mental health problems to seek help and providing information on where to find it.

YouTube and Twitter are capable of influencing a large number of teenagers (Khasawneh, Ponathil, Firat Ozkan, & Chalil Madathil, 2018; Pater & Mynatt, 2017). We suggest that it is urgent to monitor social media posts related to BWC and similar self-harm challenges (e.g., the Momo Challenge). Additionally, the SPRC should properly educate social media users, particularly those with more influence (e.g., celebrities) on elements that boost negative contagion effects. While the veracity of these challenges is doubted by some, posting about the challenges in unsafe manners can contribute to contagion regardless of the challlenges' true nature.

REFERENCES

Blackwell, L., Hardy, J., Ammari, T., Veinot, T., Lampe, C., & Schoenebeck, S. (2016). LGBT Parents and Social Media: Advocacy, Privacy, and Disclosure During Shifting Social Movements. *In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 610–622). New York, NY, USA: ACM.

Suicide Prevention Resource Center. (2006). Safe and effective messaging for suicide prevention. Suicide Prevention Resource Center Newton, MA.

Cialdini, R. B. (2003). Crafting Normative Messages to Protect the Environment. *Current Directions in Psychological Science*, 12(4), 105–109.

Gonzales, A. L. (2017). Disadvantaged minorities' use of the Internet to expand their social networks. *Communication Research*, 44(4), 467–486.

Fekete, S., & Macsai, E. (1990). Hungarian suicide models, past and present. *Suicidal Behavior and Risk Factors*, 149–156.

Fekete, S., & Schmidtke, A. (1995). *PsycNET*. Retrieved August 24, 2018, from http://psycnet.apa.org/record/1995-98656-010

Fritz, N., & Gonzales, A. (2018). Privacy at the Margins| Not the Normal Trans Story: Negotiating Trans Narratives While Crowdfunding at the Margins. *International Journal of Communication Systems*, 12(0), 20.

Hilton, C. (2017). Unveiling self-harm behavior: what can social media site Twitter tell us about self-harm? A qualitative exploration. *Journal of Clinical Nursing*, 26(11-12), 1690–1704.

Khasawneh, A., Ponathil, A., Firat Ozkan, N., & Chalil Madathil, K. (2018). How Should I Choose My Dentist? A Preliminary Study Investigating the Effectiveness of Decision Aids on Healthcare Online Review Portals. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 62, No. 1, pp. 1694-1698). Sage CA: Los Angeles, CA: SAGE Publications.

Khasawneh, A., Rogers, H., Bertrand, J., Madathil, K. C., & Gramopadhye, A. (2019). Human adaptation to latency in teleoperated multi-robot human-agent search and rescue teams. Automation in Construction, 99, 265-277.

Mitchell, K. J., Wells, M., Priebe, G., & Ybarra, M. L. (2014). Exposure to websites that encourage self-harm and suicide: prevalence rates and association with actual thoughts of self-harm and thoughts of suicide in the United States. *Journal of Adolescence*, 37(8), 1335–1344

Mukhra, R., Baryah, N., Krishan, K., & Kanchan, T. (2017). "Blue Whale Challenge": A Game or Crime? *Science and Engineering Ethics*.

Padgett, D. K. (2011). Qualitative and Mixed Methods in Public Health. SAGE Publications.

Page, X., Wisniewski, P., Knijnenburg, B. P., & Namara, M. (2018). Social media's have-nots: an era of social disenfranchisement. *Internet Research*, 28(5), 1253–1274.

Pater, J., & Mynatt, E. D. (2017). Defining Digital Self-Harm. *In CSCW* (pp. 1501–1513).

Ponathil, A., Agnisarman, S., Khasawneh, A., Narasimha, S., & Madathil, K. C. (2017). An empirical study investigating the effectiveness of decision aids in supporting the sensemaking process on anonymous social media. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 61, No. 1, pp. 798-802). Sage CA: Los Angeles, CA: SAGE Publications.