

Can the 'Dark Triad' traits be a predictor of cyber hate speech?

Hemamali Tennakoon

Antisocial behaviours have escalated to the cyberspace. Yet, the views on identification, mitigation and interventions relevant to minimising such behaviours lack consensus. Also, there are a multitude of antisocial behaviours studied in various online settings but not with clear demarcations between behaviours and their predictors. Hence, there is room for further research on cyber hate speech (a variant of online antisocial behaviour), particularly in Social Networking Sites (SNS). Hence, this article presents a brief review of literature exploring the triggers of antisocial behaviour, broadly identified as 'cyber aggression'. Then, the question is raised whether cyber hate speech, a sub-set of cyber aggression is triggered by personality factors such as the 'Dark Triad' traits. Consequently, this paper contributes to knowledge by proposing to test personality traits as an important predictor of computer-mediated anti-social behaviour in SNS and provide directions for future research.

Human aggression can be succinctly defined as "any behaviour directed towards another individual that is carried out with proximate (immediate) intent to cause harm (Anderson & Bushman, 2002). Further, cyber aggression is defined as "...intentional harm delivered by the use of electronic means to a person or a group of people irrespective of their age, who perceive(s) such acts as offensive, derogatory, harmful, or unwanted" (Grigg, 2010). The terms 'intent' and 'causing harm' stands out in both definitions, irrespective of the context in which aggression is defined. Based on the definitions, one can also be argue that cyber hate speech is a sub-set of the broader concept of cyber aggression.

Findings of the 'Italian Hate Map' project (Lingiardi et al., 2019) found that most hate speech on Twitter is directed at women, followed by immigrants, gay and lesbian persons, Muslims, Jews and disabled persons. Legislations across the globe have not yet evolved to successfully govern and criminalise offences such as cyber hate speech, dissemination of racist

and xenophobic material online (Banks, 2010, 2011, Laverick & Joyce, 2019; Quintel & Ullrich, 2019). In such a context, the consequences of online hate speech can have detrimental consequences not only from an individual (victims') point of view but also from a societal perspective. For instance, it was found that individuals or groups exposed to hate speech could experience emotional distress (Lee-Won et al., 2019). It is further argued that in the absence of intervention, hate movements could "extend its collective identity internationally, thereby facilitating a potential 'global racist subculture'" (Perry & Olsson, 2009).

Given the consequences of cyber hate speech, researchers have attempted to identify factors that trigger cyber hate (online and offline). Findings lack harmony and point to multiple factors. The naming or labelling of these factors vary. For instance, some have identified social triggers (Schultze-Krumbholz & Scheithauer, 2009; Zimmerman & Ybarra, 2016) while others label similar factors as 'situational' (Engelhardt & Bartholow, 2013; Lapidot-Lefler & Barak, 2015). Further, individual factors have been identified and labelled as 'personality related' (Robinson & Wilkowski, 2010), 'biological' (Silver & Nedelec, 2019; Yeudall, 1978) and 'personal' (Norris, 2004; Runions & Bak, 2015).

Theoretically speaking, General Theory of Crime (Gottfredson & Hirschi, 1990) suggests that individual factors such as self-control is directly linked to criminal acts. General Aggression Model (Anderson & Bushman, 2002) extend this view to include individual differences and situational factors that could affect the level of self-control. In further identification of personal factors, Social Ecology Theory argue that person variables such as biological traits (age, sex), past history (traditional bullying involvement), personality attributes (empathy, self-control) can lead to cyberbullying perpetration (Barlett, 2019). Therefore, one could argue that personality attributes are a determinant of online aggressive behaviours including engaging in online hate speech.

Based on this theoretical background and other recent studies (Webster et al., 2014; Pabian et al., 2015; Kokkinos et al., 2016; Tang and Fox, 2016), the argument that individual personalities could influence intention to act aggressively online can be reinforced. However, there are various personality traits that can be considered. A popular choice among researchers is the study of Dark Triad traits in studying cyberbullying (Arntfield, 2015; Brown et al., 2019; Goodboy & Martin Van, 2015; Van Geel et al., 2017; Kurek et al., 2019). Some studies focus on the Dark Triad of personality and cyber aggression in general (Furnham et al., 2013; Peluchette et al., 2015; Withers et al., 2017) while others focus specific types of aggressive behaviours in SNS settings including online hating behaviour (Sorokowski et al., 2020), harmful online behaviors and language (Bogolyubova et al., 2018), celebrity abuse (Scott et al., 2020) etc.

While these studies focus on different aspects of online aggression and abuse, there is a

lack of strong, empirical evidence explaining the relationship between Dark Triad personality traits and online hate speech, specific to SNS. Further, some studies use secondary data where software is used to detect hate speech among existing social media posts (Räsänen et al., 2016). This is due to methodological issues in collecting data directly from perpetrators. Researchers have also pointed out issues in quantifying hate speech (Guberman et al., 2016) which call for methodological improvements to hate speech related research in SNS. Similarly, there is room for a comprehensive definition of ‘hate speech’. For instance, it is not clear whether the terms such as ‘hate speech- trolling’, ‘anti-Muslim sentiments’ (Jakubowicz et al., 2017), verbal aggression in Trolling (Veszelszki, 2017) can be used synonymous with cyber hate speech. Therefore, the following future research directions are proposed to close the above research gaps:

1. Lack of empirical evidence on ‘Dark Triad’ personality traits and online hate speech in SNS can be remedied by conducting studies involving multiple research approaches, both quantitative and qualitative as well as primary and secondary (e.g. using online surveys involving volunteer panel of online SNS users, experimental studies involving scenario- based data collection).
2. Currently, researchers use the short Dark Triad Questionnaire by Jones and Paulhus (2014) to measure Dark Triad of personality consisting of Machiavellianism, narcissism, and psychopathy. Further, Buss-Perry Aggression Questionnaire (BPAQ) (Buss and Perry, 1992) can be adapted to measure intention to engage in online hate speech. However, the earlier discussed issue of quantifying ‘hate speech’ and measuring ‘intention’ rather than actual behaviour still prevail. Therefore, future studies could focus on developing new measures and validating them through replication to address methodological concerns.
3. Hate speech on SNS can be directed at various groups based on individual and societal prejudices, stereotyping, and/or discrimination (e.g. racism and xenophobia, antisemitism and Islamophobia, and sexism). Future research involving hate speech on SNS should focus on these specific aspects rather than group them all under one study. This would allow for specific hate speech related definitions to emerge that are more meaningful and comprehensive.

References

- Anderson, C.A. and Bushman, B.J., 2002. Human aggression. *Annual Review of Psychology*, 53.
- Arntfield, M., 2015. Towards a cybervictimology: Cyberbullying, routine activities theory, and the anti-sociality of social media. *Canadian Journal of Communication*, 40(3).
- Banks, J., 2010. Regulating hate speech online. *International Review of Law, Computers & Technology*, 24(3), 233-239.
- Banks, J., 2011. European regulation of cross-border hate speech in cyberspace: The limits of legislation. *European Journal of Crime, Criminal Law and Criminal Justice*, 19, p.1.
- Barlett, C.P., 2019. *Predicting Cyberbullying: Research, Theory, and Intervention*. Academic Press.
- Bogolyubova, O., Panicheva, P., Tikhonov, R., Ivanov, V. and Ledovaya, Y., 2018. Dark personalities on Facebook: Harmful online behaviors and language. *Computers in Human Behavior*, 78, 151-159.
- Brown, W.M., Hazraty, S. and Palasinski, M., 2019. Examining the dark tetrad and its links to cyberbullying. *Cyberpsychology, Behavior, and Social Networking*, 22(8), 552- 557.
- Buss, A. H., & Perry, M. 1992. The Aggression Questionnaire. *Journal of Personality and Social Psychology*, 63, 452–459.
- Engelhardt, C.R. and Bartholow, B.D., 2013. Effects of situational cues on aggressive behavior. *Social and Personality Psychology Compass*, 7(10), 762-774.
- Furnham, A., Richards, S.C. and Paulhus, D.L., 2013. The Dark Triad of personality: A 10 year review. *Social and Personality Psychology Compass*, 7(3), 199-216.
- Goodboy, A.K. and Martin, M.M., 2015. The personality profile of a cyberbully: Examining the Dark Triad. *Computers in Human Behavior*, 49, 1-4.
- Gottfredson, M.R. and Hirschi, T., 1990. *A general theory of crime*. Stanford University Press.
- Grigg, D.W., 2010. Cyber-aggression: Definition and concept of cyberbullying. *Journal of Psychologists and Counsellors in Schools*, 20(2),143-156.

- Guberman, J., Schmitz, C. and Hemphill, L., 2016, February. Quantifying toxicity and verbal violence on Twitter. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion* (pp. 277-280).
- Jakubowicz, A., Dunn, K., Mason, G., Paradies, Y., Bliuc, A.M., Bahfen, N., Oboler, A., Atie, R. and Connelly, K., 2017. Racism and the Affordances of the Internet. In *Cyber Racism and Community Resilience* (pp. 95-145). Palgrave Macmillan, Cham.
- Jones, D.N. and Paulhus, D.L., 2014. Introducing the short dark triad (SD3) a brief measure of dark personality traits. *Assessment*, 21(1), 28-41.
- Kokkinos, C.M., Baltzidis, E. and Xynogala, D., 2016. Prevalence and personality correlates of Facebook bullying among university undergraduates. *Computers in Human Behavior*, 55, 840-850.
- Kurek, A., Jose, P.E. and Stuart, J., 2019. 'I did it for the LULZ': How the dark personality predicts online disinhibition and aggressive online behavior in adolescence. *Computers in Human Behavior*, 98, 31-40.
- Lapidot-Lefler, N. and Barak, A., 2015. The benign online disinhibition effect: Could situational factors induce self-disclosure and prosocial behaviors?. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 9(2).
- Laverick, W. and Joyce, P., 2019. Conclusion: Hate Crime—Brexit and Beyond. In *Racial and Religious Hate Crime* (pp. 325-365). Palgrave Macmillan, Cham.
- Lee-Won, R.J., White, T.N., Song, H., Lee, J.Y. and Smith, M.R., 2019. Source magnification of cyberhate: affective and cognitive effects of multiple-source hate messages on target group members. *Media Psychology*, 1-22.
- Lingiardi, V., Carone, N., Semeraro, G., Musto, C., D'Amico, M. and Brena, S., 2019. Mapping Twitter hate speech towards social and sexual minorities: a lexicon-based approach to semantic content analysis. *Behaviour & Information Technology*, 1-11.
- Norris, K.O., 2004. Gender stereotypes, aggression, and computer games: An online survey of women. *Cyberpsychology & Behavior*, 7(6), 714-727.
- Pabian, S., De Backer, C.J. and Vandebosch, H., 2015. Dark Triad personality traits and adolescent cyber-aggression. *Personality and Individual Differences*, 75, 41-46.

- Peluchette, J.V., Karl, K., Wood, C. and Williams, J., 2015. Cyberbullying victimization: Do victims' personality and risky social network behaviors contribute to the problem?. *Computers in Human Behavior*, 52, 424-435.
- Perry, B. and Olsson, P., 2009. Cyberhate: the globalization of hate. *Information & Communications Technology Law*, 18(2), 185-199.
- Quintel, T. and Ullrich, C., 2019. *Self-Regulation of Fundamental Rights? The EU Code of Conduct on Hate Speech, Related Initiatives and Beyond. Fundamental Rights Protection Online: The Future Regulation Of Intermediaries*, Edward Elgar Publishing, Summer/Autumn.
- Räsänen, P., Hawdon, J., Holkeri, E., Keipi, T., Näsi, M. and Oksanen, A., 2016. Targets of online hate: Examining determinants of victimization among young Finnish Facebook users. *Violence and Victims*, 31(4), 708-725.
- Robinson, M.D. and Wilkowski, B.M., 2010. Personality processes in anger and reactive aggression: An introduction. *Journal of Personality*, 78(1), 1-8.
- Runions, K.C. and Bak, M., 2015. Online moral disengagement, cyberbullying, and cyber-aggression. *Cyberpsychology, Behavior, and Social Networking*, 18(7), 400- 405.
- Schultze-Krumbholz, A. and Scheithauer, H., 2009. Social-behavioral correlates of cyberbullying in a German student sample. *Zeitschrift für psychologie/Journal of Psychology*, 217(4), 224-226.
- Scott, G.G., Brodie, Z.P., Wilson, M.J., Ivory, L., Hand, C.J. and Sereno, S.C., 2020. Celebrity abuse on Twitter: the impact of tweet valence, volume of abuse, and dark triad personality factors on victim blaming and perceptions of severity. *Computers in Human Behavior*, 103, 109-119.
- Silver, I. and Nedelec, J.L., 2019. The Anatomy of Violence: The Biological Roots of Crime. *Journal of Criminal Justice Education*, 30 (1), 159-161.
- Sorokowski, P., Kowal, M., Zdybek, P. and Oleszkiewicz, A., 2020. Are online haters psychopaths? Psychological predictors of online hating behavior. *Frontiers in psychology*, 11, 553.
- Tang, W.Y. and Fox, J., 2016. Men's harassment behavior in online video games: Personality traits and game factors. *Aggressive Behavior*, 42(6), 513-521.

- Van Geel, M., Goemans, A., Toprak, F. and Vedder, P., 2017. Which personality traits are related to traditional bullying and cyberbullying? A study with the Big Five, Dark Triad and sadism. *Personality and Individual Differences*, 106, 231-235.
- Veszelszki, Á., 2017. Verbal and Visual Aggression in Trolling. *Virtual Reality-Real Visuality. Frankfurt am Main et al.: Peter Lang*, 141-155.
- Webster, G.D., DeWall, C.N., Pond Jr, R.S., Deckman, T., Jonason, P.K., Le, B.M., Nichols, A.L., Schember, T.O., Crysel, L.C., Crosier, B.S. and Smith, C.V., 2014. The brief aggression questionnaire: Psychometric and behavioral evidence for an efficient measure of trait aggression. *Aggressive Behavior*, 40(2), 120-139.
- Withers, K.L., Parrish, J.L., Terrell, S. and Ellis, T.J., 2017. The relationship between the “dark triad” personality traits and deviant behavior on social networking sites. *Social Computing, AMCIS 2017 Proceedings*.
- Yeudall, L.T., 1978. The neuropsychology of aggression. *Clarence M. Hincks Memorial Lecture: Psychobiological Approaches to Aggression in Mental Illness and Mental Retardation. Alberta Hospital, Edmonton, Canada*.
- Zimmerman, A.G. and Ybarra, G.J., 2016. Online aggression: The influences of anonymity and social modeling. *Psychology of Popular Media Culture*, 5(2), 181.