



2017

Virtually Free Speech: The Problem of Unbridled Debates on Social Media

Follow this and additional works at: <https://scholarsarchive.byu.edu/intuition>



Part of the [Psychology Commons](#)

Recommended Citation

(2017) "Virtually Free Speech: The Problem of Unbridled Debates on Social Media," *Intuition: The BYU Undergraduate Journal in Psychology*: Vol. 12 : Iss. 2 , Article 9.

Available at: <https://scholarsarchive.byu.edu/intuition/vol12/iss2/9>

This Article is brought to you for free and open access by the All Journals at BYU ScholarsArchive. It has been accepted for inclusion in *Intuition: The BYU Undergraduate Journal in Psychology* by an authorized editor of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Virtually Free Speech: The Problem of Unbridled Debates on Social Media

Brandon Parker

Brigham Young University

Abstract

How individuals communicate on the internet has been influenced by the rise of social networking and the introduction of Web 2.0. Contentious sociopolitical arguments containing false polarization and confirmation bias, accompanied by egocentric and ethnocentric thinking, are common. Ideological radicalization has also increased, as evidenced by the prevalence of ISIS and White supremacists on sites such as Twitter and Reddit. One consequence is the online practice known as *flaming*—intensely aggressive, personal verbal attacks. To reduce its occurrence, both macro- and micro-solutions should be implemented. On the macro-level, social networks and related online organizations should establish and enforce guidelines while not threatening free speech. On a micro-level, individuals can reduce the tendency to negatively react to oppositional viewpoints by participating in self-affirming activities. Self-polarization bias can be reduced by being aware of the similarities between ideologies across groups and individuals.

Virtually Free Speech: The Problem of Unbridled Debates on Social Media

On March 23, 2016, Microsoft released an artificial intelligence (AI) chatbot on Twitter known as Tay. Mimicking the behavior of a modal millennial social-media user, it learned from conversations with human users. The more Tay interacted with them, the more readily it was mistaken for an 18-to-24-year-old. Tay's existence began with a simple tweet, "hellooooooo world!!!" (a phrase common among computer programs), yet the inclusion of an emoji of the earth and the emotive spelling and punctuation made the tweet seemingly adolescent (TayTweets, 2016). Tay's tweets were initially benign and playful but later morphed malignantly. Along with malicious comments, feedback, and requests, Tay began praising Hitler, advocating the genocide of certain races, and posting offensive images and content. Two days after releasing Tay, Microsoft deleted the tweets and took Tay offline earlier than anticipated (Lee, 2016). What was meant as an advance in social AI technologies became a window into an alarmingly dark world on social media.

With 1.71 billion monthly users on Facebook and 313 million on Twitter, humans across the globe have never been more connected (Facebook, 2016; Twitter, 2016). Individuals of all races, backgrounds, lifestyles, and cultures have joined hands electronically. Social networking sites were originally designed to "make the world more open and connected" by helping people "stay connected with friends and family...discover what's going on in the world, and...share and express what matters to them" (Facebook, 2016). Although social networks have benefited humankind, there is mounting evidence of their facilitating increased aggression.

Researchers found that the language used on microblogging sites (like Facebook and Twitter) has become more intense and emotional during the last five years (Ranellucci, Poitras, Bouchet, Lajoie, & Hall, 2016). The use of online surveys and of data mining on social networking sites and online surveys has allowed researchers to identify patterns of online social behavior and to better understand the changes that are occurring on the internet.

The rise of social media has created online communities and a space wherein individuals can freely express their own emotions (Schuschke & Tynes, 2016). This is due in part to the "Web 2.0

revolution,” a term referring to the shift toward social networking and user-generated content (Ranellucci et al., 2016). According to Schuschke and Tynes (2016), when online, individuals express their shared views to communities they are a part of, thus promoting “online deliberation [that] mainly reinforces preexisting views” (Halpern & Gibbs, 2013, p. 1160). This reinforcement may lead to stereotyping of certain groups and ethnicities and may incite heated exchanges, colloquially known as *flaming* (Halpern & Gibbs, 2013).

Displays of egocentrism and ethnocentrism are common on social media and may be responsible for the growing acceptance of such orientations (Hmielowski, Hutchens, & Cicchirillo, 2014). Sociocentrism is apparent in individuals who believe their social groups to be superior to others. Such centrism can be productive of ideological radicalization and conducive to terrorist-group recruiting (Blaker, 2015). Moreover, antisocial behavior is becoming more prevalent online, thus ironically undermining the original goal of social media. Although social media is a tool designed to bring individuals closer, analyses of linguistic patterns have demonstrated that contentious, online sociopolitical exchanges have intensified and that radicalization is prevalent.

Psychosocial Changes Associated With Social Media

As a new medium of expression, social media has transformed the way individuals communicate, but it has also increased sociocentrism and radicalization. These factors have always existed in face-to-face interactions, but social networking has intensified them. According to Richards and Gutekunst (2016), 63% of American adults attribute increased sociopolitical incivility to social media. *Flaming* and *firestorms* have emerged as new forms of interpersonal communication due in part to social media, including microblogging (Halpern & Gibbs, 2013).

Sociocentrism

On social networking sites, sociocentric expression stems from the assertion that certain ideologies, racial groups, or cultural groups are superior to others (Schuschke & Tynes, 2016). Social media encourages individuals to focus inward because it replaces traditional conversation in which two or more people sit, stand, or otherwise join together in a physical space for conversation. Though those engaged in

conversation took care to present themselves well, the pressure to do so is intensified by much larger and unseen audiences on social media (Chiou & Lee, 2013). This potential preoccupation may well promote more concentrated self-absorption and egocentric thinking (Chiou & Lee, 2013).

Egocentrism

Self-focused thinking contributes to the failure to understand other perspectives and thereby inflames exchanges with others (Chambers & De Dreu, 2014). Not only do individuals favor their own perspectives, but they tend to caricature opposing viewpoints as radical, dangerous, or evil (Keltner & Robinson, 1993; Paresky, 2016). Doing so may produce the overestimation of differences between viewpoints (Chambers & De Dreu, 2014) and discourage prosocial behaviors (Chiou, Chen, & Liao, 2014).

Ethnocentrism

In addition to egocentrism, social networking sites are also host to ethnocentrism (Schuschke & Tynes, 2016). It may be expressed in racial, political, or social terms. Individuals in LGBTQ communities, for example, may find it difficult to remain on social-media sites in the face of abusive and unceasing condemnation. Individuals may avoid discussing controversial, emotionally-charged topics for fear of being attacked verbally or otherwise. Richards and Gutekunst (2016) reported that 75% of American adults consider such muting destructive to civil debate. Moreover, in the face of rejection by the larger audience, individuals may gravitate to online micro-cultures where they are validated and feel comfortable expressing their thoughts (Halpern & Gibbs, 2013).

Confirmation bias also spurs rejection of alternative views and selective exposure to viewpoints that affirm one's own (Halpern & Gibbs, 2013; Stroud, 2010). Knobloch-Westerwick, Johnson, and Westerwick (2015) found that internet users spend 64% more time viewing content they considered congenial with their personal views than content which was contradictory to their views. The selective consumption of favorable content may lead to polarization (Stroud, 2010). As in an echo chamber, the favorable content reverberates and is amplified.

Radicalization

Beyond polarization, there is the risk of ideological radicalization. As social media becomes a collection of self-contained communities, individuals are marginalized in the process and may become the potential targets of recruitment by radical groups, including terrorist groups, such as the Islamic State of Iraq and Syria (ISIS). The propaganda is attractive to the millennial-age group. An estimated 3,000 individuals from Western nations have pledged allegiance to ISIS (Blaker, 2015). Although leading social media have begun to censor radical propaganda, extremist groups still infiltrate them to share their messages (Alarid, 2016). Twitter has proven to be especially vulnerable. According to a tweet analyst, ISIS sympathizers have been among the most prolific users, sharing on average of 2,612 tweets per user over one 3-month period (Klausen, 2015).

White supremacists and anti-Semitic groups have used social media to grow by more than 22,000 users since 2012 (see Figure 1; Berger, 2016). The emergence of more extreme political ideologies has sometimes attracted otherwise well-meaning supporters of conservative political candidates and movements to share far-right extremist propaganda on social networks. Moreover, social media can be used to mobilize terrorist cells to instigate radical demonstrations and carry out murderous attacks (Kende, van Zomeren, Ujhelyi, & Lantos, 2016).

Firestorms

Radicalization is also evident in the online phenomenon known as *firestorming* or, more specifically, *firestorm debates*. They are typically characterized by large-scale surges of social-media use directed to specific topics or sociopolitical events and are marked by contentious rhetoric (Hutchens, Cicchirillo, & Hmielowski, 2015). Like wildfires, they are unpredictable and uncontrollable, spreading rapidly, and disabling online connectivity. Fan, Zhao, Chen, and Xu (2014) conducted an emotional-language analysis of 70 million posts on the Chinese microblogging site, Weibo, and found that angry posts were shared faster and more widely than what they termed “joyful” posts. Online interaction has been linked to increased contention, given that individuals use language and mannerisms different from those used in face-to-face, in-person interactions—a phenomenon known as the

online disinhibition effect (Kiesler, Siegel, & McGuire, 1984; Suler, 2004).

Individuals who are inclined to verbal aggression may perceive the internet as an appropriate place for unbridled expression (Hmielowski et al., 2014; Suler, 2004). Social media adds additional layers of interaction and communication, and polarization and confirmation biases can amplify emotional intensity. This also makes *ad hominem* attacks more likely (Hutchens et al., 2015). Flaming shares many of the same characteristics as cyberbullying, but often involves individuals who have no prior acquaintance.

Flaming

Egocentric and ethnocentric behaviors and attitudes fuel hostile exchanges on social media. Because computer-mediated communication typically lacks non-verbal cues, including vocal tone (Moor, Heuvelman, & Verleur, 2010), their absence may cause individuals to misinterpret the messages they receive (Suler, 2004). Limited comprehension is characteristic of egocentric thinking (Sassenrath, Sassenberg, & Scholl, 2014).

The case of YouTube. YouTube, a video-sharing site, is notorious for its uncivil and contentious Comment sections. Each video published by its creator is accompanied by a section wherein other users may critique the video. Flaming is prevalent, as some individuals use the comment section to personally attack the author of the video and other commentators.

Moor et al. (2010) conducted a survey of individuals who commented on videos or interpreted comments on YouTube. Most respondents were not familiar with those they either sent comments to or interpreted comments from. Senders were more likely to perceive their own words as opinions, whereas interpreters were less likely to perceive comments as opinions. Similarly, interpreters were more likely to perceive comments as offensive or provocative in comparison to senders. Interpreters were slightly more likely to perceive comments as flaming than senders were.

In the same study, when respondents were asked to rate their experiences using a Likert scale (1, disagree; 5, agree) most reported that flaming was common, that it was annoying, and that it was an issue for some users on YouTube (Moor et al., 2010).

YouTube has introduced measures to reduce incivility in comments. A new initiative, called "YouTube Heroes," rewards users

who flag inappropriate comments and promote healthy discussions (Kastrenakes, 2016). In addition to the Heroes program, YouTube is giving channel owners and content creators more control over the responses to their videos. Owners and creators have the ability to pin constructive comments at the top of the comment section as well as the heart symbol (♥) on comments as a token of gratitude and praise for civil behavior (Statt, 2016).

Overcoming Psychosocial Barriers in Social Media

A variety of solutions has been discussed and solutions are presently being implemented by companies and institutions in order to effectively promote healthy exchange on social media.

Macro Solutions

These solutions reside at the structural level (Williams, 2016). Social networks can themselves set standards for a more civil internet.

The contrasting cases of Imzy and Reddit. Imzy is a new social network whose founders seek to humanize computer-mediated communication. Imzy does not rely on online advertising as a revenue source like other social networks do and seeks to establish forums for constructive debate among specialized communities of individuals who share similar interests and values. Imzy's approach requires individuals to request permission to join a community and, after joining, they must remain active, civil participants (Robertson, 2016).

Imzy is the antithesis to Reddit, which also feature communities. Reddit has moderation measures in place, but they have been largely ineffective in preventing discriminatory flaming. For example, a specific community named "Fat People Hate" had over 100,000 members, who participated in shaming, ridiculing, and targeting overweight individuals. When moderators intervened, members turned their hatred toward the moderators (Moreno, 2016).

The risk of online censorship. As technology companies establish guidelines to reduce antisocial behavior, censorship becomes an issue. The First Amendment to the Constitution of the United States pronounces basic freedoms, among them the freedoms of speech and the press. Further legal interpretations have excluded threats of violence, pornographic content, and inciting comments from constitutional protection (Heins, 2014). The definitions of these protections are vague, and as social networks become liable to

constitutional regulation (just as with the rise of mass media), the risk of censorship increases.

Micro Solutions

The effective promotion of constructive online exchange ultimately must come from individuals in communities. The problems of egocentricity, ethnocentricity, intolerance, aggression, and intimidation that infect social networks are largely rooted in individual perceptions and are better resolved at micro-levels of interaction (Stroud, 2010).

Relational awareness. In egocentric thinking, individuals unthinkingly put on the biasing lenses of their own understanding rather than trying to understand perspectives they consider alien. A bogus polarization follows as the latter perspectives continue to be ignored, thus effectively reducing and even eliminating the chance of cooperative, civilized exchange (Chambers & De Dreu, 2014). Individuals should be encouraged to identify the inevitable similarities in viewpoints, that is, they should be encouraged to develop relational awareness. Individuals can produce eventual consensus while still affirming different perspectives. A study by Puccio (2003) found false polarization in students at Stanford University and the attendant over-exaggeration of differences. When students permitted opposing viewpoints to be heard, their perception of the differences between opinions more closely approximated the actual differences. Moreover, after debating both sides of the affirmative-action issue, for example, the students were more confident that agreements could be reached.

Self-affirmation. When individuals and groups express an ethnocentric pattern of thought online, they necessarily bar themselves, even aggressively, from ideas and arguments that are unfavorable to their own. (Halpern & Gibbs, 2013; Stroud, 2010). According to self-affirmation theory (Armitage & Rowe, 2016), an individual with a healthy self-image reacts to opposition and challenges to his or her identity in a less defensive manner than those whose self-image is ill-defined (Cohen & Sherman, 2014). Armitage and Rowe (2016) examined conflict-resolving behavior in children and adolescents and found that those participants who exhibited stronger self-concepts after participating in self-affirming activities were less aggressive in their communication (Armitage & Rowe, 2016). If self-affirming practices were widely available and widely endorsed on the internet more-civil interaction may well increase.

Conclusion

The creation of a more connected and open world is a modern goal. Social media is widely considered one of the most effectual means for achieving the goal. But with their wide acceptance has come a rising tide of prejudice, hate, radicalization, and false polarization (Knobloch-Westerwick et al., 2015; Schuschke & Tynes, 2016; Stroud, 2010).

Radicalized political ideologies and polarized viewpoints increased during the 2016 US presidential primaries and general election (Sanders, 2016). Immediately after the results of the general election on November 8, 2016 were announced, individuals took to social media with emotional expressions ranging from elation to devastation. Firestorms and flaming ensued as individuals refused to accept the results and vilified those who had voted for the winning candidate (Sanders, 2016).

Additional research on the effects of social media should be conducted on the use of relational awareness, self-affirmation, and other prosocial behavior to enhance civil interaction online. Newer platforms, such as virtual and augmented reality, should be included in that research.

References

- Alarid, M. (2016). Recruitment and radicalization: The role of social media and new technology. In M. Hughes, & M. Miklaucic (Eds.), *Impunity* (pp. 313-330). Washington, D.C.: Center for Complex Operations.
- Armitage, C. J., & Rowe, R. (2016). Evidence that self-affirmation reduces relational aggression: A proof of concept trial. *Psychology of Violence, 6*(3), 1-9. doi:10.1037/vio0000062
- Berger, J. M. (2016). Nazis vs. ISIS on twitter: A comparative study of White nationalists and ISIS online social media networks. Retrieved from <http://parlinfo.aph.gov.au/parlInfo/search/display/display.w3p;query=library/jrnart/4820902>
- Blaker, L. (2015). The Islamic State's use of online social media. *Military Cyber Affairs, 1*(1), 1-9. doi:10.5038/2378-0789.1.1.1004
- Chambers, J. R., & De Dreu, K. W. (2014). Egocentrism drives misunderstanding in conflict and negotiation. *Journal of Experimental Social Psychology, 51*, 15-26. doi:10.1016/j.jesp.2013.11.001
- Chiou, W., Chen, S., & Liao, D. (2014). Does Facebook promote self-interest? Enactment of indiscriminate one-to-many communication on online social networking sites decreases prosocial behavior. *Cyberpsychology, Behavior, and Social Networking, 17*(2), 68-73. doi:10.1089/cyber.2013.0035

- Chiou, W., & Lee, C. (2013). Enactment of one-to-many communication may induce self-focused attention that leads to diminished perspective taking: The case of Facebook. *Judgment and Decision Making, 8*(3), 372-380.
- Cohen, G. L., & Sherman, D. K. (2014). The psychology of change: Self-affirmation and social psychological intervention. *Annual Review of Psychology, 65*(1), 333-371. doi:10.1146/annurev-psych-010213-115137
- Facebook. (2016). Company info. Retrieved from <https://newsroom.fb.com/company-info/>
- Fan, R., Zhao, J., Chen, Y., & Xu, K. (2014). Anger is more influential than joy: Sentiment correlation in Weibo. *PloS One, 9*(10), 2-16. doi:10.1371/journal.pone.0110184
- Fox, J., & Warber, K. M. (2015). Queer identity management and political self-expression on social networking sites: A co-cultural approach to the spiral of silence. *Journal of Communication, 65*(1), 79-100. doi:10.1111/jcom.12137
- Halpern, D., & Gibbs, J. (2013). Social media as a catalyst for online deliberation? Exploring the affordances of Facebook and YouTube for political expression. *Computers in Human Behavior, 29*(3), 1159-1168. doi:10.1016/j.chb.2012.10.008
- Heins, M. (2014). The brave new world of social media censorship. Retrieved from <http://harvardlawreview.org/2014/06/the-brave-new-world-of-social-media-censorship/>
- Hmielowski, J. D., Hutchens, M. J., & Cicchirillo, V. J. (2014). Living in an age of online incivility: Examining the conditional indirect effects of online discussion on political flaming. *Information, Communication & Society, 17*(10), 1196-1211. doi:10.1080/1369118X.2014.899609
- Hutchens, M. J., Cicchirillo, V. J., & Hmielowski, J. D. (2015). How could you think that?!?: Understanding intentions to engage in political flaming. *New Media & Society, 17*(8), 1201-1219. doi:10.1177/1461444814522947
- Kastrenakes, J. (2016). YouTube's new plan to deal with awful comments: Have commenters help moderate. Retrieved from <http://www.theverge.com/2016/9/21/13001520/youtube-heroes-comment-moderation-program-announced>
- Keltner, D., & Robinson, R. J. (1993). Imagined ideological differences in conflict escalation and resolution. *International Journal of Conflict Management, 4*(3), 249-262. doi:10.1108/eb022728
- Kende, A., van Zomeren, M., Ujhelyi, A., & Lantos, N. A. (2016). The social affirmation use of social media as a motivator of collective action. *Journal of Applied Social Psychology, 46*(8), 453-469. doi:10.1111/jasp.12375
- Kiesler, S., Siegel, J., & Mcguire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist, 39*(10), 1123-1134. doi:10.1037//0003-066x.39.10.1123
- Klausen, J. (2015). Tweeting the jihad: Social media networks of Western foreign fighters in Syria and Iraq. *Studies in Conflict & Terrorism, 38*(1), 1-22. doi:10.1080/1057610X.2014.974948

- Knobloch-Westerwick, S., Johnson, B. K., & Westerwick, A. (2015). Confirmation bias in online searches: Impacts of selective exposure before an election on political attitude strength and shifts. *Journal of Computer-Mediated Communication*, 20(2), 171-187. doi:10.1111/jcc4.12105
- Lee, P. (2016). Learning from Tay's introduction. Retrieved from <http://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/#sm.00008hibhkgm6d9ztck1y7wbxmboj>
- Moor, P. J., Heuvelman, A., & Verleur, R. (2010). Flaming on YouTube. *Computers in Human Behavior*, 26(6), 1536-1546. doi:10.1016/j.chb.2010.05.023
- Moreno, J. (2016). Why we want to fix the Internet's community problem. Retrieved from <http://www.lennyletter.com/culture/a398/why-we-should-fix-the-Internets-community-problem/>
- Paretsky, P. (2016). I'm right, you're evil. Retrieved from <https://www.psychologytoday.com/blog/happiness-and-the-pursuit-leadership/201611/i-m-right-you-re-evil>
- Puccio, C. T. (2003). The search for common ground: Overcoming false polarization and unwarranted pessimism about ideological differences (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Full Text. (305292949).
- Ranellucci, J., Poitras, E. G., Bouchet, F., Lajoie, S. P., & Hall, N. (2016). Understanding emotional expressions in social media through data mining. In S. Y. Tettegah (Ed.), *Emotions, technology, and social media* (pp. 85-103). San Diego, CA: Elsevier Academic Press.
- Richards, M. D., & Gutekunst, L. (2016). How much incivility will American voters tolerate in the race to the White House? Retrieved from <http://www.krcresearch.com/how-much-incivility-will-american-voters-tolerate-in-the-race-to-the-white-house/>
- Robertson, A. (2016). Imzy, a social network based on kindness, opens its doors to the public. Retrieved from <http://www.theverge.com/2016/10/27/13416716/imzy-social-network-ex-reddit-launch-funding>
- Sanders, S. (2016). Did social media ruin election 2016? Retrieved from <http://www.npr.org/2016/11/08/500686320/did-social-media-ruin-election-2016>
- Sassenrath, C., Sassenberg, K., & Scholl, A. (2014). From a distance: The impact of approach and avoidance motivational orientation on perspective taking. *Social Psychological and Personality Science*, 5(1), 18-26. doi:10.1177/1948550613486672
- Schuschke, J., & Tynes, B. M. (2016). Online community empowerment, emotional connection, and armed love in the Black Lives Matter movement. In S. Y. Tettegah (Ed.), *Emotions, technology, and social media* (pp. 25-47). San Diego, CA: Academic Press. doi://dx.doi.org.erl.lib.byu.edu/10.1016/B978-0-12-801857-6.00002-6
- Statt, N. (2016). YouTube will let users pin comments to promote better discussions. Retrieved from <http://www.theverge.com/2016/11/3/13504804/>

youtube-pinned-comments-moderation-tools-harassment

Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication*, 60(3), 556-576. doi:10.1111/j.1460-2466.2010.01497.x

Suler, J. (2004). The online disinhibition effect. *Psychology of Cyberspace*, 7(3), 321-326. doi: 10.1089/1094931041291295

TayTweets. [TayandYou]. (2016, March 23). hellooooooo world!!! [Tweet]. Retrieved from <https://twitter.com/tayandyou>

Twitter. (2016). Twitter company. Retrieved from <https://about.twitter.com/company>

Williams, R. (2016). The rise of incivility and what to do about it. Retrieved from <https://www.psychologytoday.com/blog/wired-success/201610/the-rise-incivility-and-what-do-about-it>

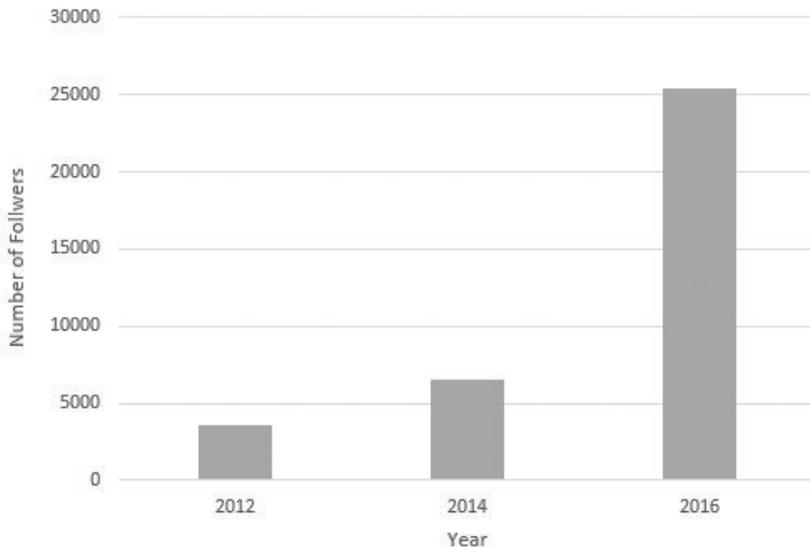


Figure 1. Recent growth of White supremacists using Twitter (Berger, 2016).