Intersect, Vol 10, No 3 (2017)

Moderating the Uncontrollable

Jenny Kim
Stanford University

Abstract

Traditional forms of media have shifted online, transforming methods of engagement. Information sharing, which once required time and various means of communication, is now global and instant. Particularly, news organizations have integrated technology into their online platforms. The most dynamic technology of news organizations is comment threads, in which readers are free to publicly express their opinion. Comment threads increase reader engagement, but their lack of moderation blurs boundaries of free speech. Moderating comment threads in online news changed the relationship between readers and journalists, shaping how they view each other as contributors of public information. By identifying readers and journalists as consumers and producers, respectively, of news as an information demand, this paper examines the successes and failures of existing moderation technologies of comment threads and proposes an alternative method of moderation. The proposed moderation technique leads to better discussions facilitated directly by journalists and ultimately fosters a sense of community that technology alone cannot provide.

Introduction

Online news changed the way consumers read and communicate. It is a new sphere of communication that allows consumers to publicly and instantly express their thoughts, giving rise to participatory forums that redefine the relationship between readers and journalists (Meyer et al., 2013). Once passive, online newspapers are now an active medium eliminating the "gatekeeping" (Domingo, 2008) that print newspapers strictly complied with. Comment threads, a digital forum of public opinion, replaced letters to the editor in print news, narrowing the gap between readers and journalists. Despite their tremendous popularity among readers, comment threads do not fulfill information demand for news organizations and cause problems for journalists, creating a gap in the information market. To close this gap, moderation technologies must not only address revenue and readership, but also create stronger relational ties between readers and journalists. First, the current state of moderation in online newspapers and its effects on readers and journalists is considered. Then, three examples of current moderation technologies that help meet the producer information demand are discussed. Lastly, a new method of moderating comment threads is proposed.

While moderation policies are constantly under revision and undergo improvements, comment threads have become so popular that they act as the intermediary between readers and journalists. Despite the amount of attention abusive comments attract, news organizations are not legally responsible for these comments; in fact, they can only respond to complaints from readers after the comment has been reported or flagged as abuse by other readers (Canter, 2012, p. 611). Because journalists hold no legal authority to control commenters' behaviors or intervene in comments that they believe to be controversial, many news publications find themselves eliminating comment threads (Canter, 2012). The unclear rules of moderation prove to be a double edge sword for readers as consumers and journalists as producers (Braun, 2011).

Journalists as Producers and Readers as Consumers

Journalists seek information because they want to increase readership and financially optimize their business models. News organizations' financial motives can be outlined as what James T. Hamilton, Professor of Journalism at Stanford University, calls the "subscription model" (Hamilton, 2011, p. 280). Under this model, news organizations do not meet the producer information demand due to the high volume of profanity, heated debates, and irrelevant advertisements in their readers' comments. News organizations' unmet information demand stems from the lack of moderation in comment threads.

On the other hand, readers as consumers seek information as an experienced good that they can judge only after they have read it. Consumers comment on online news because of two primary reasons: to express personal opinions on a subject matter of interest and to interact with other readers (Canter, 2012). The lack of moderation gives readers a chance to freely express thoughts, which includes abusive comments and spam. These unfiltered comments and ads create a negative experience to consumers who genuinely want to engage in relevant discussions pertaining to the article (Canter, 2012, pp. 607-608).

Challenges and Conflicts

Lack of moderation in comment threads dilute journalists' expert role and authority over readers (Braun et al., 386). Journalists have a duty to inform readers, but humorous or abusive comments hinder them from fulfilling this duty. Journalists view comments from readers as "amateur" (Heinonen, 2011, p. 39) and feel as though they are "chaperones" (Braun et al, 2011, p. 384) of online discussion rather than their role as messengers of news. Moreover, some readers challenge facts or direct ad hominem attacks toward journalists. Journalists find these attacks the most salient because personal comments damage their image and reputation (Canter, 2012). A higher number of comments on a journalist's article correlates to a journalist's lower trust of their audience's ability to get facts right or act ethically (Meyer et al., 2013). Such distrust of their audience influences journalists' views of moderation. Journalists wish to maintain enough editorial control in a "one-way communication model" (Canter, 2012, p. 605) to exclude readers. Therefore, a lack of moderation in comment threads discourages journalists from engaging with their readers.

Readers and journalists have varying views on how comment threads should be monitored. Journalists believe non-moderation is a better choice for managing comment threads because they do not have any legal responsibility and therefore remove themselves from the conversation (Canter, 2012). Many news organizations turn to a non-moderation approach because journalists would have to handle readers' comments in addition to their articles. Ed O'Keefe, senior producer at ABC News, notes the difficulties that journalists face when monitoring readers' unpredictable behavior and the desire of news organizations to facilitate community discussions, which become a part of the organization's content. He advises that news organizations "open the gate" to encourage active communication and take risks of readers' "inappropriate, offensive, [and] threatening remarks" (Braun, 2011, p. 388), because the relationship between journalists and readers is worth more than the benefits in flagging users and deleting profanity. Although journalists favor non-moderation, it poses risks because it often leads to abusive comments, hate speech, and defamation of news organizations. Therefore, the primary concerns of journalists, which include avoiding legal liability and damaging their organization's brand, can be addressed if news organizations balance the "economic, processional, and ideological aspirations" (Braun et al., 2011, p. 384) and implement moderation methods that uphold traditional journalistic principles.

Two-Way Interaction and Community

News organizations need a new method of moderation that satisfies both readers and journalists. In a joint survey by the University of Texas Engaging New Project (Morrison, 2017), approximately 75% of commenters said they wanted a reporter to participate and engage in their discussions. Furthermore, the likelihood of uncivil comments decreased by 15% when reporters participated in comment threads. Clearly, there is a two-way benefit for readers who want reporters' engagement and reporters who want less abuse from trolls and ideas from comments for their next articles. Mónica Guzmán, a co-founder of "The Evergrey," observes that journalists must have a "mutual beneficial relationship" with their readers. Both journalists and readers would benefit equally from collaborative moderation. Readers and journalists must engage in an active two-way interaction, not a hierarchical communication model; in fact, Chung (2008) advocates that interactivity is an essential quality of online journalism that provides readers with increased choice options. More importantly, interactivity of comments allows them to participate in the production of information. Comment threads would increase readership and commercial value for journalists and give readers a chance to engage with others real-time. Journalists must also realize that their articles are valued not only for their information, but also the sense of community that they provide.

Community is an important factor of comment threads. Readers feel a sense of community when they read comment threads because they can often relate to the thoughts and experiences of other commenters. The desire to belong could explain why readers decide to read news articles in the first place. Comments foster feelings of "membership, identity, belonging, and attachment" (Blanchard, 2007). Readers express opinions to those outside their echo chambers and feel much less entitled because they are anonymous online. Community influences readers to maintain a stronger and longer online presence (Meyer et al., 2013) as they build virtual relationships through their online personas. However, the freedom to engage in online communities comes at a price; some readers may feel offended or outnumbered. Therefore, to prevent bullying and to promote a safer community for reader engagement, journalists must establish rules and moderate comment threads (Meyer et al., 2013). Several news outlets have identified the vital role that a community plays in online comment threads and have experimented with moderation techniques. The most successful projects thus far include 'The Coral Project,' Google's 'Jigsaw,' and 'Civil.'

Current Moderation Technologies

'The Coral Project' is a collaborative effort by Mozilla, *The Washington Post,* and *The New York Times* to increase readership and create a troll-free environment. Greg Barber, the director of digital news projects of the *Washington Post,* decided to partake in the project because he believes that commenters are the most loyal readers who pay the bills (Mendelez, 2016). The project offers three main products: ‘Ask,’ ‘Talk,’ and ‘Guides’ (Coral). ‘Ask’ is a form builder and an embedded polling tool that gathers content data that readers post and produces galleries of responses for news organizations to share on their website. ‘Talk’ offers data about community and identifies troubled resources, such as tracking commenters with the highest percentage of flagged comments and using filters to find the most liked commenter. Together with its third product ‘Guides,’ which recommends ways to improve online communities, The Coral Project may just appear to be a moderation tool on the surface. But this digital platform, targeted to publishers, serves to enliven community building in a safe and sustainable environment (Coral). It differentiates from other moderation projects because it is open source; any newsroom or publisher can use or modify The Coral Project’s three products.

Another noteworthy effort to moderate comment threads is ‘Civil.’ Civil offers ‘Civil Comments,’ ‘Civil Live,’ ‘Civil Reviews,’ and ‘Civil Audience.’ The mission of these products is to provide “real audiences with real engagement’ (Civil). Civil aims to help readers be the best versions of themselves while interacting with others online by cross-checking comments and allowing readers to rate civilities of other readers using an algorithm. Civil differs from The Coral Project because of its “behavioral approach” (Civil) directed to commenters to think twice about their posts and change or rephrase words if necessary. And unlike The Coral Project, Civil is an embedded tool in the comment threads that targets readers, not publishers. David Hulen (2016) of “Alaska Dispatch News” mentions that the cross-checking technique that Civil uses, makes commenters act as their own moderators, which in turn, makes discussions more humanistic. Hulen believes Civil’s blended system of technology and peer review creates a better experience for all readers that encourages civil debates and conversations without spams.

‘Jigsaw’ is the third effort that improves moderation in comment threads. An incubator of Alphabet, Google’s parent company, Jigsaw seeks to “expand comments to more articles” and “increase the speed at which comments are reviewed” (Press, 2016). Partnering with *New York Times,* Jigsaw improves the workload for fourteen moderators who sift through as many comments as 11,000 on a given day (Press, 2016). It alleviates their intensive labor with predictive models that group comments of similar content. Heavily grounded in machine learning, Jigsaw uses a technology called ‘TensorFlow’ to generate separate algorithms for each news organization. Although it is open source like The Coral Project, Jigsaw differs technically in its integration of machine learning to aid moderators make decisions faster and read more comments. Modeled on tremendous input data of comments of varying content and tones, machine learning in Jigsaw automatically discerns irregularities in comments. Jigsaw’s technology advances assisted moderation by moving beyond using filters to find unacceptable comments.

A New Moderation Technology

'The Coral Project', 'Civil', and 'Jigsaw' are current moderation techniques that benefit news organizations and help journalists take some pressure off manually dealing with unwanted comments, fighting spam, and identifying trolls. Should news organizations use these tools, journalists in turn can better facilitate discussions and commenters can have more opportunities to openly engage with those outside their echo chamber in a safer environment. However, despite the numerous benefits of these emerging moderation technologies, they fail to draw clear distinctions between which comments are abusive and which practice free speech (Canter, 2012, p. 615). Thus, a truly successful method of moderating comments must avoid legal liability, protect readers' freedom of speech (Braun, 2011), and foster democratic ideals for readers.

News organizations should implement a live function in comment threads so that journalists can have a real-time interaction with readers. A chat box or a Twitter-type format of automatically updated comments will curate discussions and create a community for readers and journalists. A joint survey by University of Texas reveals that 80 percent of journalists "never" or "rarely" respond to comments (Morrison, 2017), indicating readers' passive experience with articles. Live comment threads will transform this one-way relationship into an interactive two-way communication. Incorporating 'live' functions in articles has been successful in attracting readers. In a study by Thurman and Walters (2012), live blogs on *Guardian* received most attention and the highest number of site visits. In addition, readers are twice more likely to participate in live blogs than other forms of online news. Live blogs are an expanding journalistic phenomenon and have gained popularity globally, especially for their coverage on breaking news (Thurman et al., 2012).

The same effective concepts of live blogs can be applied to comment threads. Although moderating live comments requires more effort by journalists, machine learning could lessen this pressure. Algorithms gathered from open access sources, like The Coral Project and Jigsaw, can automatically identify inappropriate comments and decrease journalists' efforts to identify problematic comments. Further, the joint survey reveals that 50% of readers wanted newsrooms to highlight quality comments (Morrison, 2017). Journalists can spotlight "quality commenters" (Heinonen, 2011, p. 42) who are worthy of recognition during live comment threads. Journalists can also shape future article ideas from readers' enriching and triggering stories (Morrison, 2017). Live comment threads with tools utilizing machine learning will be the most effective moderation strategy that gives journalists, who are concerned about amateur comments, authority over their readers. More significantly, live comments eliminate journalists' legal liability of readers' post-publication concerns, as they are not "legally responsible for content of contributions the moment they appear" (Singer, 2011).

Live comment threads will foster a sense of community more so than current comment threads because they require readers and journalists to actively participate in real-time discussions. The more readers live comment threads attract, the more variety of opinions and less bias in the articles. Instead of a single journalist's viewpoint on an issue, live comments threads will allow readers to directly engage and weigh in on the issue as well. Moreover, news organizations will experience positive externalities from live comment threads because collaborative comment threads lead to a marketplace of ideas, lessen polarization among readers and create a safer, online environment. Another positive spillover is the influence that live comment threads can have on the readers who do not directly participate in live comment threads. These readers can learn additional information on an issue, read opinions that they agree or disagree with, monitor the discussion as a third party, highlight important comments that journalists miss out on, and even contribute by identifying abusive comments that machine learning fails to detect.

Conclusion

Lack of moderation in comment threads leads to an unsafe online environment that forces news organizations to remove these threads and take away readers' freedom of speech. However, comment threads foster a community among readers and inspire journalists with new ideas. While 'The Coral Project,' 'Civil' and 'Jigsaw' seek to improve reader engagement and create a troll-free environment, news organizations have yet to find a perfect moderation technology that balances the readers' and journalists' responsibilities. A combination of machine learning and live function in comment threads is the ideal moderation technology that serves as an interactive two-way communication between readers and journalists. This technology would also adhere to the subscription incentive of news organizations, as high demands for live comment threads will increase their revenues.

Journalists must alter their negative views on moderation before making improvements on existing moderation techniques. But because technology reflects the needs and demands of society, the social incentives of journalists must change before news organizations implement new technologies. Journalists are likely to view moderation more positively should news organizations raise their professional returns, such as status. This will motivate journalists to have a positive view on moderating comment threads, collaboratively create a safe troll-free space for discussions and fulfill their duty to inform the public (Hamilton, 2004).

References

Braun, Joshua and Gillespie, Tarleton. (2011). Hosting The Public Discourse, Hosting The Public. *Journalism Practice, 5*(4), 383-398. doi:10.1080/17512786.2011.557560

Canter, Lily. (2012). The Misconception Of Online Comment Threads. *Journalism Practice, 7*(5), 604-619. doi:10.1080/17512786.2012.740172

Chung, Deborah. S. (2008), Interactive Features of Online Newspapers: Identifying Patterns and Predicting Use of Engaged Readers. Journal of Computer-Mediated Communication, 13: 658–679. doi:10.1111/j.1083-6101.2008.00414.x (Chung)

Civil. (n.d.). Why Civil? Retrieved February 26, 2017, from https://www.getcivil.com/why-civil/

Coral. (n.d.). The Coral Project. Retrieved February 26, 2017, from http://coralproject.net/about.html

Hamilton, James T. (2004). *All the News That's Fit to Sell*. Princeton, NJ: Princeton UP. Print.

Hamilton, James T. (2011) essay in, Will the Last Reporter Please Turn Out the Lights: The Collapse of Journalism and What Can be Done to Fix It," edited by Robert Mc Chesney and Victor Pickard (New York: New Press), pp. 277-288.

Heinonen, Ari. (2011). The Journalist's Relationship with Users, in Participatory Journalism: Guarding Open Gates at Online Newspapers (eds J. B. Singer, A. Hermida, D. Domingo, A. Heinonen, S. Paulussen, T. Quandt, Z. Reich and M. Vujnovic), Wiley-Blackwell, Oxford, UK. doi: 10.1002/9781444340747.ch3

Hulen, David. (2016). A new approach to online comments at Alaska Dispatch News. Retrieved February 26, 2017, from https://www.adn.com/commentary/article/were-going-more-civil-approach-story-comments-adncom/2016/03/23/

Melendez, Steven. (2016). Internet Comments Are Awful. Could They Be Awesome? Retrieved February 26, 2017, from https://www.fastcompany.com/3065844/internet-comments-are-awful-could-they-be-awesome

Meyer, Hans. K., & Carey, Michael. Clay. (2013). In Moderation. *Journalism Practice, 8*(2), 213-228. doi:10.1080/17512786.2013.859838

Morrison, Sara. (2017). The Future of Comments. Retrieved February 26, 2017, from http://niemanreports.org/articles/the-future-of-comments/

Press. (2016). The Times is Partnering with Jigsaw to Expand Comment Capabilities. Retrieved February 26, 2017, from http://www.nytco.com/the-times-is-partnering-with-jigsaw-to-expand-comment-capabilities/

Singer, Jane. (2011). ''Taking Responsibility: Legal and Ethical Issues in Participatory Journalism.'' In Participatory Journalism: Guarding

Gates at Online Newspapers, edited by Jane Singer, 119 38. Chichester, UK: Wiley-Blackwell.

Thurman, Neil. & Newman, Nic. (2014). The Future of Breaking News Online? A study of live blogs through surveys of their consumption, and of readers' attitudes and participation. Journalism Studies, doi: 10.1080/1461670X.2014.882080

Wang, Shan. (2017). Commenters say they want journalists and experts to join them in the comments. Retrieved February 26, 2017, from http://www.niemanlab.org/2017/01/commenters-say-they-want-journalists-and-experts-to-join-them-in-the-comments/