CHAPTER 16

Social Infrastructure: Designing for Online Civility

Ramona Pringle

In our hyper-connected world, for many the term microaggressions conjures images of digital infractions: the racist rant of an angry Twitter troll, the toxic rhetoric left behind by a user in the comments section of a newspaper, or the sexist banter in an online forum that is unwelcoming to anyone new or different. When we speak of microaggressions, "brief and commonplace daily verbal, behavioral, or environmental indignities, whether intentional or unintentional, that communicate hostile, derogatory, or negative racial slights and insults toward people of colour" (Sue et al., 2007, p. 273), it is hard not to immediately think of a Twitter timeline, full of hostile remarks and hateful rhetoric from angry "eggs" out to cause controversy and incite uproar.

It would appear that nowhere are microaggressions more commonplace, or more inherent, than across the Internet, where platform design seems to foster, if not fuel, toxic behavior. From Twitter, to online games such as *League of Legends*, to the comments section of Canada's public broadcaster, the *CBC*, few corners of the online world are free from the mounting toxicity that has become all too common in these digital spaces. The list of news outlets that have gotten rid of commenting

R. Pringle (✉)
RTA School of Media, Ryerson University, Toronto, Canada

C. L. Cho et al. (eds.), *Exploring the Toxicity of Lateral Violence and Microaggressions*, https://doi.org/10.1007/978-3-319-74760-6_16

features altogether includes the *Toronto Star, NPR, Reuters, Popular Science, The Telegraph,* and *Recode*. Canada's broadcaster closed down commenting on all articles related to the Indigenous population, because they were consistently filled with bigotry and hateful rhetoric instead of the intended helpful dialogue (Pringle, 2017). While some of these outlets claim to have closed down commenting in order to better focus their limited resources on social media channels, others, like Popular Science admit that "trolls and spam-bots" have overwhelmed their ability to provide intellectual debate, saying "even a fractious minority wields enough power to skew a reader's perception of a story" (Labarre, 2013).

Colloquial words of wisdom such as "don't read the comments" and "don't feed the trolls"—troll being slang for someone who seeks out discord by posting inflammatory remarks online—have become widely adopted strategies for managing online incivility; unfortunately, they do little to remedy the issue. As a result, online public spaces—the digital equivalents of the town square where ideas are shared and ideologies are discussed—have become hijacked by the toxic minority, whose loud and angry presence often overwhelms any attempt at civil dialogue or debate. While communal online environments have the potential to be valuable collaborative spaces, the opportunity to learn from each other is negated when civil discussion is prohibited by a dominant aggressive culture. In fact, in the context of news platforms, studies show that comments can actually taint how content is perceived; once the comments section associated with a piece of media has been overtaken by venomous or discordant posts, those who might have previously been interested in a meaningful discussion will stop engaging. Maria Konnikova calls this "the nasty effect," whereby the nastier the comments, the more polarized readers become about the contents of the article in question (Konnikova, 2013). Users who are exposed to polite comments do not change their view of the contents of the article, but those who read nastier comments tend to have a more negative take on the topic at hand.

The challenge of how to make the commenting that takes place within online communities less aggressive spaces spans beyond news outlets, into other Internet domains such as gaming and social media, and left unresolved, can have negative implications for businesses, as users opt to disengage rather than face unnecessary hostility. Just as *CBC* closed down commenting on articles about Canada's Indigenous population, other platforms are plagued by the vile treatment of women, people of color, and minorities. The harassment of female players is notorious in

online games, and social network Twitter is plagued by its reputation for being a hotbed of verbal abuse. While strategies including user registration, the prohibition of anonymity, and pre- and post-moderation, whereby posts are reviewed by a moderator before being made visible to the public or reviewed by a moderator shortly after being submitted, are already being implemented by different organizations (Ksiazek, 2015), fixing the toxicity of online culture is a time-, money- and labor-intensive undertaking, all of which can be prohibitive factors for organizations seeking to increase the civility of their communities.

Across sectors and disciplines, aggression and toxicity have become commonplace online, and need to be addressed. After all, these spaces, though digital, are where we spend a great deal of our time, attention, and energy, and despite being pixels and data, the impact of harmful online comments can be very real. As Marshall McLuhan stated, "we shape our tools, and then our tools shape us" (McLuhan, 1994, p. xxi). In this sense, it can be argued that offline microaggressions can be fueled by the rampant toxicity online (Johnson, 1997)—especially when there are little to no repercussions for this kind of behavior.

Given that the design of these platforms seems to foster rampant toxicity, there is a strong argument to be made that it is by design, too, that these issues can be combatted. This chapter will examine how design solutions and social infrastructure can be developed for the online world, with examples from several digital platforms, including gaming environments, news portals, online forums, and other collaborative spaces, in order to provide a framework by which we can start to mitigate microaggressions online.

The Social and Commercial Value of Online Commenting and Social Spaces

On the one hand, it is understandable that so many outlets and organizations have made the decision to remove commenting, the communal conversation around a given topic, and close down comment sections, the destinations where these digital exchanges take place, because it can feel like an unwieldy issue. But for all the trolling that occurs online, it is vital to not forget how powerful the Internet can be as a tool that connects us to information, people, and ideas. This is the central premise of Clay Shirky's *Cognitive Surplus* (Shirky, 2010), in which he discusses the immense potential for creative and intellectual output that can result from the collaborative efforts of people's online hobbies and pastimes.

But the problem of poisonous online rhetoric cannot be solved by avoidance, or by simply closing down forums. Where some organizations, such as those in journalism and media, might feel as though they have the option to close commenting sections, as those are only tangentially related to the original content, for other organizations such as social networks and online games, the communal space is inherent to the platform itself. Moreover, removing the opportunity for users to comment and engage communally negates the opportunity to gain from the greatest affordance of the Internet: the ability to collaborate and learn from each other. According to Johanna Blakely, the director of the Norman Lear Centre, the worst outcome of closing comment sections is that we lose out on the potential of these interactive platforms to learn about each other, collaborate, and grow. Blakely is one of several domain experts from various disciplines interviewed as part of a multidisciplinary effort to identify design solutions to online toxicity. As a researcher who studies the impact of entertainment and media on society, she says,

> just having that archive itself is one of the most valuable things on earth that exist right now. It's not that it suddenly tells us everything about ourselves, but it's this opportunity that we've never had before to at least start interpreting this information about our attention and how we allocate it and our desires and how we record them. (J. Blakely, personal communication, January 3, 2017)

In many cases, the opportunity to share thoughts and perspectives on an issue—be they anything from politics, to pop culture, to hobbies—is part of the motivation to engage in the first place. Removing the ability to comment affects the experience itself: It may take away the motivation to engage with a topic more deeply and to share it with a wider audience (Konnikova, 2013).

The benefit of finding ways to remedy the current toxicity found in online commenting spaces is not just experiential. While there is evidence that the communal element of online interactions is a driving factor in audience engagement, there are financial repercussions to the decision to remove commenting, as well. Just as newspaper readers have transitioned to digital platforms in seek of up-to-date information, television viewers are cutting cords and moving to online sources of content such as Netflix and Amazon (Strangelove, 2015). As such, it would behoove media corporations to foster their online communities. If commenting sections

are closed down and users cannot engage on the proprietary platforms where the media on which they are commenting is housed, the result is not that those users will cease to comment, rather, they will comment elsewhere. The net result is that that "elsewhere" will grow, and benefit financially from the activity of those users, while the outlet generating content will suffer based on lower engagement. As media consumption habits become increasingly digital-first, there is a strong business case to be made, to consider the design of commenting features and user engagement, alongside the design of Web and mobile platforms, and the creation of the content itself.

Just as there is a business argument for making the commenting sections of traditional media outlets less hostile, the same is true of digital environments that are innately interactive, such as social networks and online games, says game designer Jeffrey Lin whose credits include the massive multiplayer online game *League of Legends*. Lin was interviewed for this project because of his complementary expertise in game design and human interaction; as a designer with a PhD in cognitive neuroscience, he studies the way players engage with each other and develops applied design solutions based on those findings. According to Lin, the more negative behaviors an individual is exposed to when playing a game, the more likely he or she is to quit and never come back. And so, he concludes, this is an essential—and valuable—problem for companies to solve (J. Lin, personal communication, February 21, 2017).

Design Solutions: Digital Social Infrastructure

By design, social media can foster toxic behavior. Platforms favor short, quippy remarks, shock value is key to virality (Olsen & Gaude, 2015), and the speed by which timelines scroll past a user's field of vision fuels a sense of ephemerality, whereby users are more likely to comment quickly without necessarily considering the impact of their words. Just as the design of these platforms seems to encourage uncivil discourse, it can be argued that it is through design strategies that this growing toxicity can be combatted.

In *The Design of Everyday Things* (Norman, 2013), the seminal text on the design of everything around us, author Don Norman explains, that "design presents a fascinating interplay of technology and psychology" and that good designers "must understand both" (Norman, 2013, p. 7). He says,

> All artificial things are designed. Whether it is the layout of furniture in a room, the paths through a garden or forest, or the intricacies of an electronic device, some person or group of people has to decide upon the layout, operation and mechanisms. Not all designed things involve physical structures. Services, lectures, rules and procedures and the organizational structures of businesses and governments do not have physical mechanisms, but their rules of operation have to be designed, sometimes informally, sometimes precisely recorded and specified. (Norman, 2013, p. 4)

It is understandable that the abuse that is encountered online can make individuals feel defeated and demoralized about the potential for positive change. Based on rampant trolling and flame wars—hostile, aggressive online exchanges—it is not uncommon for the Internet to be understood as the digital equivalent of a giant flood light, bringing into stark relief the worst of human nature. From this perspective, viewing the Internet as a mirror of humanity, it can seem as though there is a sort of inevitability to the toxicity that overflows online. But addressed from Norman's design-centric point of view, the Internet's tendency toward toxicity is neither innate nor unsolvable. Rather, it is an issue of bad design.

To date, the Internet has widely been seen as a digital frontier, a 'Wild West' where anything goes (Schneiderman, 2014), with little enforced regulation. On the one hand, this open sensibility has created an unprecedented arena for the democratization of ideas and ideals, but that is in jeopardy when toxicity strangles the air out of communal online spaces. In this seemingly lawless, ruleless, and often repercussion-less free-for-all, users feel empowered to say, or do, anything without a thought to social tolls. But even offline, when there are no rules, anarchy prevails. This is the premise of the classic work of literature, *Lord of the Flies* (Golding, 1954), in which lawless disorder surfaces, after a group of young boys are stranded on a deserted island without adult supervision, having survived a plane crash. Without rules to abide by, the group's social infrastructure quickly deteriorates into a cruel and dangerous free-for-all. A fable about the fragility of peaceful coexistence from before the invention of the modern Internet, the tale is a precursor to what is common today across much of the online world, wherein harassment emerges when there are no rules, or no enforced social contract.

Offline, society has developed codes of conduct that the majority of the public lives by; agreeing to these design systems or sets of rules helps people to coexist. These design systems, or rules, comprise a "social infrastructure," designed to help people keep themselves and each other

free from harm. Laws exist as guidelines, and there are repercussions when laws are disobeyed. Take for example the system by which traffic flows. Traffic control is social infrastructure, a system designed to make our transportation interactions easier and mitigate the potential for damage. While there are repercussions when drivers break established rules, signals such as stop signs and changing colored lights have been designed to help drivers steer clear of wrongdoing in the first place. The traffic system has all of the markers that define what Norman would consider good design: Rules are clearly defined and communicated, and users are aware of what is possible, as well as the repercussions for deviating from the intended behavior. He explains, "good design requires stepping back from competitive pressures and ensuring that the entire product be consistent, coherent, and understandable" (Norman, 2013, p. 263). Perhaps the toxicity we see online is not an "internet issue", but rather a design issue, and online communities would benefit from a system similar to the traffic system, with widely understood and accepted codes of conduct and repercussions for bad behavior. After all, while anonymity is often blamed as the culprit for bad online behavior, in fact, it seems that it is the lack of consequences or repercussions that fosters a hostile or aggressive nature (Birk et al., 2016). As Norman explains, "designers need to focus their attention on the cases where things go wrong, not just when things work as planned" (Norman, 2013, p. 9). If the early vision for the Internet was a hope that it would grow into a communal network wherein individuals could share ideas and co-create solutions, then the evolution of spaces such as open-source communities can be seen to be best-case scenarios. The rampant toxicity that fills forums and social media feeds is the worst-case scenario of an initial utopian vision and as such is ripe for redesign.

Dialogue vs Debate

In considering the design of a successful online social infrastructure, it is important to keep in mind that aspiring to online civility does not necessitate that everyone be in agreement all of the time, or that comments should be banal and homogenous. Nor does it in any way equate to censorship, or the limiting of an individual's freedom of speech. Rather, the ideal is a system that is designed to encourage dialogue. At its best, the Internet is a tool that democratizes. This is evident in diverse cases, ranging from the Arab Spring, in which digital tools were used to voice

the concerns and experiences of the populous, to the anti-establishment celebrity status of early YouTubers, who garnered massive audiences and success despite breaking from traditional entertainment models and gatekeepers (Shirky, 2008).

As Daniel Yankelovitch comments in *I'm Right and You're an Idiot* (Hoggan & Litwin, 2016), "Democracy requires space for compromise, and compromise is best won through acknowledging the legitimate concerns of the other. We need to bridge opposing positions, not accentuate differences" (Hoggan & Litwin, 2016, p. 7). Author James Hoggan goes on to add, "When we use dialogue rather than debate we gain completely different insights into the ways people see the world" (Hoggan & Litwin, 2016, p. 9). Though neither Yankelovitch nor Hoggan is referencing the Internet expressly, their comments are relevant to this discussion of online civility. After all, the inherent strength of the Internet is its ability to connect users with diverse points of view. The challenge is simply that despite its potential, all too often, open spaces for online discussion are hijacked by disparaging abusers, as opposed to those wishing to acknowledge different perspectives.

A person can be argumentative and still be civil. As long as an argument is made without insulting or offensive language, it can maintain its civility, even if it might not be considered "polite" or "nice" (Ksiazek, 2015). Unfortunately, oftentimes in the worst of online confrontations, there is no attempt at dialogue, let alone civility. As psychologist Daniel Kahneman states, "we can be blind to the obvious, and we are also blind to our blindness" (Kahneman, 2011, p. 48). Or, as Hoggan so aptly named his book, the toxicity of online commenting is due to the phenomenon of "I'm right and you're an idiot" (Hoggan & Litwin, 2016). When a user attacks another's views or posts, with comments that target them based on factors such as gender, sexual orientation, or even political leanings, the intention is to rile that person, rather than to educate or inform.

For Steve Ladurantaye, who at the time of writing was the manager of digital news for *CBC*, overseeing not only the content being posted to the national broadcaster's Web site, but also the strategy for online community and commenting, the hostility that prevents civil online discourse is not a new occurrence. Rather, the industries that are now struggling to remedy rampant online toxicity have fostered a me-versus-you or us-versus-them sensibility for a long time now, in order to provoke responses. In the digital age, the term for this kind of fabricated provocation is "click bait," content

that is designed and presented to attract attention, even if it that attention is negative. From his experience working in multiple newsrooms, Ladurantaye explains, "Stories are set up to be provocative and they are deliberately framed as somebody versus somebody… It's sort of the way journalism has worked for the last 100 years." But, he adds, if journalists and media makers can develop a model where they provide context and offer solutions, they can promote conversation. "I think once you start [providing solutions instead of provoking responses] you've taken away the natural inclination to oppose. There's not your side and my side, rather it's 'this is a problem and this is how it might be fixed'" (S. Ladurantaye, personal communication, February 15, 2017).

While the Internet seems to inherently foster incivility, it is not the first platform to encourage debate. While on the one hand, the Internet's open, networked nature makes it uniquely well equipped to help diverse users work together to solve problems, there are lessons that can be learned from other platforms that have managed to facilitate debate while avoiding the pitfalls of harassment and abuse. Charles Shanks is the senior producer of *CBC* radio's national call-in program *Cross Country Checkup*. In his role, Shanks has been designing debates for a long time. For over forty years, the radio program has been taking calls from diverse listeners all across the country on current affairs issues; the strategy is to highlight the places where opposing views might actually coincide or overlap.

> We try to frame it more towards the middle where people are a little more ambivalent, more willing to move and listen to each other's opinions. I think we've worked hard at that over the years and people know that this is not the place you go to see banging heads, this is the place you go to actually talk… Acknowledging similarities, instead of focusing on differences can lay the basis for dialogue versus debate. (C. Shanks, personal communication, March 3, 2017)

And while this strategy has been largely successful for the decades-old radio call-in show, it would appear to be equally beneficial online.

While Ladurantaye and Shanks were interviewed for their perspectives from the trenches of the newsroom, and the potential design solutions that can be gleaned from their experiences interacting with audiences in radio and digital platforms, Sean Stewart was included in this research for his understanding of game mechanics and user engagement, specifically

as it pertains to collaboration among users. Stewart is credited as being one of the founders of Alternate Reality Games (ARGs), a breed of collaborative games that take place across a range of platforms spanning the Internet and the offline world. Just as there is wisdom to be gained in terms of making the online environment more civil from the experience of keeping a call-in program like *Cross Country Checkup* on the air for over four decades; likewise, there are design lessons from gaming that can be useful in the redesign of comment sections of news articles and journalistic media. According to Stewart, the key to success in ARGs lays in bringing players together, as opposed to pitting them against each other for the sake of competition. (S. Stewart, personal communication, January 4, 2017). With *The Beast*, for instance, an ARG created to accompany the Steven Speilberg film, *A.I. Artificial Intelligence*, players take responsibility for themselves, from the start, to host their own conversations as a means of pooling knowledge and solving puzzles. "Players communicate with one another, share their knowledge, offer storyline interpretations and gather info necessary to solve the game" (Kim, Allen, & Lee, 2008). In this context, commenting was established not as a means of expressing a polarizing opinion, or attributing value to the content in question, but rather as a means of collaboratively engaging with the content to extend the experience.

"One of the things that was interesting about *The Beast*, which is different from the comments section on *Sports Illustrated* or the *New York Times*, is there was no conversation that we hosted. There was only a conversation driven by the players themselves," says Stewart. "So they took responsibility for it from the beginning. There was no authority against whom to rebel. It was communal" (S. Stewart, personal communication, January 4, 2017). This is an explanation, also, for why niche online communities experience less hostility than platforms with broader scopes. In niche communities or fan sites, the users agree on shared values and interests when they opt in. As a result, says Blakely, these communities are full of constructive dialogue. "Of course there are fights and battles and tiffs. But, generally the interaction is incredibly positive, because people are constantly learning from a community that they did not have immediate geographical access to" (J. Blakely, personal communication, January 3, 2017).

The explanation for this, according to Stewart, is that on a niche site like *Ravelry.com*, a popular knitting community, despite different backgrounds or even levels of prowess, there is a shared assumption that

everyone is there to learn, and by default, everyone there is imperfect. This, as a foray into the community, prevents the me-versus-you premise that can so quickly yield toxicity in public forums. "If you are on a site like *Ravelry*, everyone drops a stitch, everyone makes a horrible lumpy thing, everyone admires the work that is hard because they know it's hard and everyone shares their stories of failure," says Stewart. "Anytime there is a community of doers it is also a community of failures, because that's the price of admission" (S. Stewart, personal communication, January 4, 2017). He points out that, with opinion and punditry, the forms of communication that are dominant in many commenting platforms that serve broader audiences, there is no failure. Rather, each time a user speaks, or posts, or tweets, it is coming from a place of authority or certainty.

The takeaway for the brave designers tackling the issue of online civility is to focus on commonalities. For the incivility that emerges in news-based commenting sections, this could be as simple as trying to solve the problems being addressed through an approach such as solutions journalism, which focuses on how people are addressing challenges and gives readers resources to be able to help in a given cause (Curry & Hammonds, 2014), instead of overextending and trying to be everything to everyone. With this approach, users can be actively engaged in a meaningful and purposeful way, without being inflammatory or argumentative, by raising awareness, or contributing funds, for example, to the issue being addressed, through solutions provided by the journalist.

Additionally, there is a benefit to a design that is both top-down, wherein the social infrastructure is designed and enforced by the company, and bottom-up, whereby users create their own rules and community standards. Reiterating Stewart's findings that healthy communities tend to include an element of self-moderation, Lin notes that while many companies choose one approach or the other, either controlling the community or taking a hands-off approach, in fact, a healthy balance is ideal; a design-centered approach (top-down) can solve half the problems, and a community-centered approach (bottom up) can solve the other half. In other words, in addition to whatever mechanism the designers create, the more that the community can take control of the space for themselves, to establish shared goals and values, the more that community will self-enforce civil discourse. While Lin and Stewart cite examples from gaming, the presence of hostility in these environments is no less challenging to contend with than what is found in the comment

sections of news and media organizations, and their findings provide practical design solutions that can be applied in other contexts to help foster civil online engagement.

Anonymity, Consequences, and Repercussions

While the anonymity that is prevalent online, and unique to online discourse, is often cited as the culprit for bad behavior (Cho, Kim, & Acquisti, 2012), there is reason to believe that anonymity alone is not to blame for the rampant toxicity that is expressed online. That said, studies note that by humanizing the Web, and developing strategies whereby posters see other commenters as more than just anonymous generators of text on a screen, the level of civility is increased. For Ladurantaye, initiating human interaction during Facebook live streaming of news programming, whereby a moderator responded to and interacted with the community in real time, made a big difference to the tone of the subsequent audience conversations. The benefit of reminding users that their online peers are also real human beings on the other side of the computer screen is substantial. The risk of forgetting that the profiles people engage with online are also real human beings is a trap that even seasoned professionals can fall into, without a face looking them back in the eye. "Even I have a really hard time thinking of people in the comments section as people," says Ladurantaye. Many outlets have found that the level of civility increases when the author of a post or article engages in the comment section (Stroud, 2014). It should be noted, however, that there is a human toll for wading into an already toxic forum, especially as a self-identified female or minority, wherein the bashing often has little to do with the substance of the original content, and more to do with preconceptions and bias.

While several organizations have tried implementing real name policies, whereby users are required to create online profiles linked to their offline identities, to combat what they consider to be the negative effects of online anonymity, many have yielded better results through the implementation of a code of conduct, with consequences and repercussions for those who step out of line (Lin, 2015). Sometimes, the two are correlated. For example, Ladurantaye explains that the advantage of using Facebook as a platform for commenting is that most people can be held accountable because they use their real names. "You can report toxic behavior and the user can have their account lost.

That level of accountability is important." But as Lin points out, that model is imperfect, as the repercussions are not directly related to the user's goal, which in the case of Ladurantaye's *CBC* audience would be to read and comment on news articles. "Even if I say something super racist, I still get access to the news site, I don't get any repercussions, I don't get punished at all and in fact I kind of enjoy everybody giving me more attention for me being the person that I am on that site" (J. Lin, personal communication, February 21, 2017).

Working as a designer on *League of Legends*, Lin found that by implementing meaningful consequences for bad behavior, negativity was greatly reduced. In broad strokes, "if the community finds that you've behaved inappropriately, you can be temporarily banned from the game," a punishment powerful enough to impact the decisions, behavior, and language of individual players. After a year of research, Lin and his team realized that significant punishment for bad user behavior had never been integrated into the game's design, so users were free to behave badly without consequence, "We had to approach it from a consequences perspective first because the culture had gone to a point where it was out of control" (Lin, 2017). Implementing a system, or social infrastructure, with penalties for negative behaviors was the best way to get what Lin calls "a meaningful and necessary reset." Designers of the game implemented *The Tribunal*, wherein community members can collectively vote on whether a flagged infraction does in fact break the agreed-upon code of conduct, and then they administer a punishment accordingly, often booting players from the game and preventing them from being able to play for a length of time deemed proportionate to their offense. This method was found to successfully mitigate toxic behavior, with a 50% reduction in recidivism after a player is punished for an infringement (Blackburn & Kwak, 2014).

Similarly, the online forum *Reddit*, despite its reputation for xenophobia, has also managed to successfully implement a design strategy that centers on user-led moderation, and repercussions for bad behavior. With "shadow-banning," a user is blocked, but unaware of it; as Lin explains it, "they can keep posting, and they think they are posting so other people can see, but nobody else can. What they learn is if they keep posting this toxic stuff, nobody actually gives them any feedback so they just stop" (Lin, 2017). If the incentive for posting to a platform such as *Reddit* is to be seen and have your comments read, this punishment will incentivize a change of tone or approach, so that the user can stay in the conversation. As Stewart notes, "We are, even the very trolliest among

us, social creatures. And if your comments get conclusively down-voted you feel less" (Stewart, 2017).

For this kind of punishment to be most effective, repercussions should be immediate, in order to draw a connection between cause and effect, so that the offending poster is aware of the relationship between their toxic behavior and the resulting punishment. Several companies—including Riot Games, the makers of *League of Legends*, and Google, which has launched a tool called Jigsaw—have now implemented strategies involving machine learning, to pick up negative keywords, with an immediate consequence of 30-second loss of chat or similar repercussions. Lin (2017) explains, "The closer the feedback loop to the actual time of the incident, the much better the results are," adding that the real-time repercussions are far more effective in changing user behavior than punishment after the fact. Granted, systems using artificial intelligence that look for keywords are far from perfect solutions; this approach still has a tendency to identify false positives and punish commenters for their use of flagged words, even when they are not being used in a harmful context. Nonetheless, the premise of delivering repercussions with enough immediacy that users are made aware of their infringement is a lesson that has been shown to yield positive results.

Incentives and Rewards

Inevitably, users will not always be in agreement with each other. In fact, it is the diversity of opinions that is meant to be protected, even in cases where individuals are as polarized as they could be—for example, supporters of opposing political parties or ideologies—systems can be redesigned to incentivize good behavior and foster online civility. While Wikipedia, like Reddit, is not immune to sexism and flame wars, according to research from the Harvard Business School, individuals who edit political articles on the platform seem to grow less biased. Users who have a particular political bent tend to edit pages with opposing political positions; a right-wing contributor is likely to edit a left-wing page and encounter different views and vice versa. Because of the collaborative nature of the site, which relies on user-generated content and moderation, no article is ever "complete," and any change to the content of an article can be edited, or deleted, at any time. In a study of 70,000 articles (Greenstein & Zhu, 2014), the researchers found that

contributors who started out with extreme political stances developed more neutral language over time, breaking out of their filter bubbles, the echo chamber of like-minded opinions that often manifests in online communities and social networks, due to the necessity to post edits in such a way that they would not be removed by someone with opposing views, thus making all articles more balanced. In other words, inherent in the design of Wikipedia is a reward for presenting content as objectively as possible, as the content that is deemed acceptable by the widest array of users is the content that is most likely to remain visible and not be deleted or edited.

Blakely (2017) notes that some design strategies rely on incentive as much as punishment

> Generally, the incentive on Wikipedia for editors is to edit something and have it stay up… They know that if they just go on an opposing political site and rail, it will be deleted immediately. But if they can find a way to put it in just the right terms that it will slide past the censors, who supposedly hate them, it's a victory for them. And, it's a victory for discourse because suddenly we have an encyclopedia entry that reflects everyone's point of view. Having feedback when they do something right will shape people's behavior.

Stewart experienced similar patterns of behavior in *The Beast*, noting that good outcomes tend to lead to more good behavior. He explains it as a type of cognitive dissonance whereby the mentality of the player is, "I'm working with these people therefore I must like these people" (Stewart, 2017). Stewart points out that the community that played *The Beast* and subsequent ARGs came from movie review sites where they were always engaging in hostile arguments and flame wars. But the ARG designers found that as long as the community was kept busy with challenges and tasks, and felt as though their involvement was necessary, the quality of engagement was really positive. He notes, "It even surprised the players themselves!" (Stewart, 2017).

As an extension of this kind of reward-based engagement, Lin suggests designing a platform where the more a user contributes valuable discussion and content the more privileges he or she can unlock, such as the ability to help moderate the conversation as a super-user. But, the platform must be designed in such a way from the start,

before behavior patterns become ingrained. "The behavior you're seeing is the behavior you've designed for," says Shirky (2010, p. 196), who explains that behaviors follow opportunity: Even after a designer decides why users will want to participate in their new service, he or she has to give them an opportunity to do so in a way that they can understand and care about.

Conclusion

While for many the term microaggressions is evocative of the toxicity that has become commonplace in online commenting sections across the Internet, perhaps it is not too late to fix this culture of digital incivility.

Just as contemporary society has implemented systems of social infrastructure to help people coexist in their offline lives, so too can design help foster civility online. Through strategies including systems of consequences for breaking established and widely understood codes of conduct, to incentives for pursuing meaningful dialogue in a constructive way, several organizations have started to see positive results in their communities, often when they thought that perhaps the problem had already passed the tipping point.

What is also understood is that the Internet is an innately interactive space; online, no conversation is one way, and no content is static. As such, for new systems to be successful, designers need to consider how the infrastructure can be implemented so as to be both a top-down and bottom-up design, wherein the organization and the community members all have a voice, and a stake, in the success of the community.

References

Birk, M. V., Buttlar, B., Bowey, J. T., Poeller, S., Thomson, S. C., Baumann, N., & Mandryk, R. L. (2016, May). The effects of social exclusion on play experience and hostile cognitions in digital games. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 3007–3019). San Jose, CA, USA: ACM.

Blackburn, J., & Kwak, H. (2014). STFU NOOB! Predicting crowdsourced decisions on toxic behavior in online games. CoRR, abs/1404.5. Retrieved from http://arxiv.org/abs/1404.5905.

Cho, D., Kim, S., & Acquisti, A. (2012, January). Empirical analysis of online anonymity and user behaviors: The impact of real name policy. In *System*

Science (HICSS), 2012 45th Hawaii International Conference (pp. 3041–3050). Maui, HI, USA: IEEE.

Curry, A. L., & Hammonds, K. H. (2014). The power of solutions journalism. *Solutions Journalism Network*.

Golding, W. (1954). *Lord of the Flies*. New York: Perigee.

Greenstein, S., & Zhu, F. (2014). *Do Experts Or Collective Intelligence Write with More Bias? Evidence from Encyclopedia Britannica and Wikipedia*. Cambridge, MA, USA: Harvard Business School.

Hoggan, J., & Litwin, G. (2016). *I'm Right and You're an Idiot: The Toxic State of Public Discourse and How to Clean It Up*. Vancouver, BC: New Society Publishers.

Johnson, D. G. (1997). Ethics online. *Communications of the ACM, 40*(1), 60–65.

Kahneman, D. (2011). *Thinking, Fast and Slow*. London: Macmillan.

Kim, J. Y., Allen, J. P., & Lee, E. (2008). Alternate reality gaming. *Communications of the ACM, 51*(2), 36–42.

Ksiazek, T. B. (2015). Civil interactivity: How news organizations' commenting policies explain civility and hostility in user comments. *Journal of Broadcasting & Electronic Media, 59*(4), 556–573.

Konnikova, M. (2013). *The Psychology of Online Comments*. Retrieved from http://www.newyorker.com/tech/elements/the-psychology-of-online-comments.

Labarre, S. (2013). *Why We're Shutting Off Our Comments*. Retrieved from http://www.popsci.com/science/article/2013-09/why-were-shutting-our-comments.

Lin, J. (2015). *Doing Something About the 'Impossible Problem' of Abuse in Online Games*. Retrieved from https://www.recode.net/2015/7/7/11564110/doing-something-about-the-impossible-problem-of-abuse-in-online-games.

McLuhan, M. (1994). *Understanding Media: The Extensions of Man*. Cambridge, MA: MIT Press.

Norman, D. (2013). *The Design of Everyday Things: Revised and Expanded Edition*. New York: Basic Books.

Olsen, C. B., & Gaude, C. (2015). *Show Me What You Share and I'll Tell You Who You Are*. Retrieved from: http://lup.lub.lu.se/student-papers/record/5463392.

Pringle, R. (2017). *Online Hate Might Just Be an Issue of Bad Design*. Retrieved from http://www.cbc.ca/news/opinion/online-toxicity-1.4001767.

Schneiderman, E. (2014). *Taming the Digital Wild West*. Retrieved from https://www.nytimes.com/2014/04/23/opinion/taming-the-digital-wild-west.html.

Shirky, C. (2008). *Here Comes Everybody: The Power of Organizing Without Organizations*. New York: Penguin Books.

Shirky, C. (2010). *Cognitive Surplus: How Technology Makes Consumers into Collaborators*. New York: Penguin Books.

Stroud, N. J. (2014). *Journalist Involvement in Comment Sections*. Report prepared for the Engaging News Project. [online] https://engagingnewsproject.org/enp_prod/wp-content/uploads/2014/04/ENP_Comments_Report.pdf.

Strangelove, M. (2015). *Post-TV: Piracy, Cord-Cutting, and the Future of Television*. Toronto: University of Toronto Press.

Sue, D. W., Capodilupo, C. M., Torino, G. C., Bucceri, J. M., Holder, A., Nadal, K. L., & Esquilin, M. (2007). Racial microaggressions in everyday life: Implications for clinical practice. *American Psychologist, 62*(4), 271.