Colin Muller cam2299@columbia.edu

Alba, Davey. “Defining ‘Hate Speech’ Online is an Imperfect Art.” *Wired*. 22 August 2017.

https://www.wired.com/story/defining-hate-speech-online-is-imperfect-art-as-much-as-science/

After the white supremacist rally in Charlottesville YouTube took down a video of US soldiers destroying Nazi swastikas. Around the same time, Facebook allowed for the now infamous Daily Stormer article attacking Heather Heyer to be shared 65,000 times before they started deleting links to the post since it violated community standards on hate speech. These two incidents highlight the difficulty of properly identifying hate speech and doing so on time, and they also show the flaws that exist in current algorithms meant to identify hate speech.

Part of the problem is the sheer volume of content to be sorted through. There are around 400 hours of content uploaded to YouTube each minute and Facebook has over 2 billion users. Therefore one of the jobs of the algorithms is “to determine whether content flagged for review should be given higher priority for a human reviewer,” but as of now there is always a human who makes the final determination to remove something from platforms. Computer Scientist Bart Selman says that it will take at least a decade or so for artificial intelligence to properly understand the context around words and pictures that would allow them to make filtering decisions without human approval.The current issue is that current AI for filtering has to be revised every time there is a mistake such as the case of the Nazi video on YouTube. Another issue is on the human end, content reviewers for social media sites are trained very quickly and are even told to scim videos rather than watching them beginning to end which reduces accuracy.

****

Thompson, Nicholas. “Instagram’s Kevin Systrom Wants to Clean Up the Internet.” *Wired*. August 14 2017.

https://www.wired.com/2017/08/instagram-kevin-systrom-wants-to-clean-up-the-internet/

In 2016, Instagram’s CEO Kevin Systrom got fed up with seeing hateful and offensive comments flooding his and other people’s comment sections and he decided he wanted to make instagram a kinder place so he put his engineers to work. Instagram built a filter that “would automatically delete specific words and emoji from users’ feeds.” The first test case was on Taylor Swift’s page where her photos were flooded with comments which included snake emojis because of some celebrity beef she had. Instagram added a “hide inappropriate comments” button which blocks a list of words that the company selected “ including racial slurs and words like whore” but also allows for custom keywords or even emojis. Additionally, they launched tools that would send users using suicidal language in posts a message that says “If you’re going through something difficult, we’d like to help.” and could redirect a user to a page with support including a suicide hotline’s number.

The decision makers at Instagram are well aware of their role as private owners of a quasi-public space, and therefore take the role of the moderating algorithms seriously as it determines the sort of environment users will experience.

Colin Muller cam2299@columbia.edu

Instagram is now moving beyond simple fixes like automated deletion of specific words or emojis. Now they are using Facebook's DeepText machine learning software which does "word embedding." They are training DeepText with millions of comments, teaching it to distinguish which do not meet community guidelines even in cases that require contextual understanding since there are no overtly offensive keywords. The AI compares the content of the post with factors such as the author and commenter's relationship as well as the quality of the commenter's posting history. The DeepText product launched in late June, and has gone relatively unnoticed. It has had trouble with words with different cultural meanings and recognizing song lyrics. Now instagram is going a step further and trying to elevate high-quality comments on feeds in order to create a "mimicry effect" where people say nice things because they see others doing the same.

****

Levin, Sam. "Sexual harassment and the sharing economy: the dark side of working for strangers." The Guardian. 23 August 2017. https://www.theguardian.com/business/2017/aug/23/sexual-harassment-sharing-economy-uber-doordash-airbnb-twitter

Gig economy workers face both physical and online harassment and threats and feel there is little recourse. One woman reported receiving pornographic videos from a client on DoorDash, and after reporting it to the company the order was cancelled but the harasser was able to continue sending her inappropriate messages.

Being your own boss in gig economy means that the company is less liable for your safety and that there is no requirement for the company to provide legal or economic assistance in combatting it. Some people, like uber drivers are afraid of reporting harassment to the company out of fear of retaliation or having their accounts cancelled and losing their jobs.

"In the same way that female engineers and startup founders struggle to report harassment for fear of retaliation or lost funding, gig economy workers are in precarious positions when they are victimized, since they aren't classified as employees."

"While highly paid tech workers have complained about dysfunctional HR operations, gig economy drivers lament the fact they can struggle to even get a human on the phone when they are facing dangerous situations on the job. Often, drivers receive automated replies to their complaints."

****

Colin Muller cam2299@columbia.edu

West, Lindy. *Save Free Speech From Trolls,* NYT. 1 July 2017

https://www.nytimes.com/2017/07/01/opinion/sunday/save-free-speech-from-trolls.html?action=click&pgtype=Homepage&clickSource=story-heading&module=opinion-c-col-left-region&region=opinion-c-col-left-region&WT.nav=opinion-c-col-left-region

West pushes back against critics that conflate political correctness, cultural criticism, and feminist discourse with anti-free speech action. This is dangerous as it ignores the fact that the first amendment protects speech against government censorship and has nothing to do with opposition by other citizens to offensive or hateful speech. It is even more troubling, because people use their "defense" of free-speech to justify online harassment including death threats and threats of rape to feminist media critics.
****

Allanajune, Alia. *WhatsApp, Crowds and Power in India,* NYT. 21 June 2017

https://www.nytimes.com/2017/06/21/opinion/whatsapp-crowds-and-power-in-india.html

Widely circulated messages on whatsapp conveying sensationalized and often entirely false stories have fueled communalist violence in India. Most of these fake viral stories originate from right wing groups who attach an unrelated video with gossip of violence perpetrated by a member of a minority group. The accusations in the messages range from child-trafficking, defaming hindu icons, or killing & eating a cow. The messages are widely circulated throughout India (India comprises 200 million of whatsapp's 1 billion users) and this contributes to the mob-mentality fueled nationalist and sectarian violence.
****

Angwin, Julia and Hannes Grassegger. *Facebook's Secret Censorship Rules Protect White Men from Hate Speech But Not Black Children,* ProPublica, Julia Angwin, and Hannes Grassegger
https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms

The article reviews some of facebook's leaked slides for their content reviewers, and explains why statuses calling for the murder of all "radicalized muslims" is allowed to remain on facebook whereas calling "all white people racist" resulted in the suspension of a facebook account. This opens up to a broader discussion of how facebook decides what speech to censor on the platform.

Facebook will delete "curses, slurs, calls for violence and several other types of attacks only when they are directed at "protected categories"—based on race, sex, gender identity, religious affiliation, national origin, ethnicity, sexual orientation and serious disability/disease" however it does not protect as consistently against attacks on 'subsets' of these protected categories. So black children, female drivers, or radicalized muslims are considered a subset and are fair game for online attack whereas the category white men is a protected group. Facebook has described their approach as color-blind and equally protective of all groups, but this has been criticized for avoiding substantive equal protection and only

establishing formal equality while allowing more vulnerable subsets of protected groups to be attacked. Organizations have called on facebook to be more proactive in protecting more vulnerable groups.

How facebook differentiates between hate speech and legitimate political expression formally in their guidelines is already an issue to critics. What is more frustrating to some is having posts that clearly do not violate any of the sites policies. This happens frequently to journalists, activists, and professors who often make posts criticizing racism and police brutality. Some activists report monthly post deletions for posting things such as “white people are racist” or “White folks. When racism happens in public — YOUR SILENCE IS VIOLENCE.” This is attributed to the high density of posts that facebook reviewers have to go through, and having to make rapid decisions leads to inaccuracies that make already marginalized voices feel unjustly served by the social media platform. The pressure to rapidly remove reported hate speech on facebook is even more intense in Europe where EU regulation demands that illegal content be removed within 24 hours of being reported.

For many critics, it seems like accidentally deleted posts are only restored if they gain media attention or they are posts made by celebrities.The feeling that celebrities get special treatment is amplified by the fact that there was internal debate within facebook as to whether or not they should take down Donald Trump’s post about banning muslim immigrants as it violated the company’s policy against “calls for exclusion.” Ultimately Zuckerberg called for exemptions for Trump’s posts.

****

Tokunaga, Robert S. "Following you home from school: A critical review and synthesis of research on cyberbullying victimization." *Computers in human behavior* 26.3 (2010): 277-287.

In this article Tokunaga reviews and conducts a meta-synthesis of 25 peer-reviewed academic articles that quantitatively evaluate “the relationship between cyberbullying victimization and age, gender, negative outcomes, or coping strategies, and/or incidence rates” (279).

Table 1
Conceptual definitions of cyberbullying used in research.

Study	Conceptual definition of cyberbullying
Besley (2009)	The use of information and communication technologies to support deliberate, repeated, and hostile behavior by an individual or group, that is intended to harm others
Finkelhor et al. (2000)	*Online harassment*: Threats or other offensive behavior (not sexual solicitation) sent online to the youth or posted online about the youth for others to see (p. x)
Juvoven and Gross (2008)	The use of the Internet or other digital communication devices to insult or threaten someone (p. 497)
Li (2008)	Bullying via electronic communication tools such as e-mail, cell phone, personal digital assistant (PDA), instant messaging, or the World Wide Web (p. 224)
Patchin and Hinduja (2006)	Willful and repeated harm inflicted through the medium of electronic text (p. 152)
Slonje and Smith (2007)	Aggression that occurs through modern technological devices and specifically mobile phones or the Internet (p. 147)
Smith et al. (2008)	An aggressive, intentional act carried out by a group or individual, using electronic forms of contact, repeatedly or over time against a victim who cannot easily defend him or herself (p. 376)
Willard (2007)	Sending or posting harmful or cruel texts or images using the Internet or other digital communication devices (p. 1)
Ybarra and Mitchell (2004)	*Internet harassment*: An overt, intentional act of aggression towards another person online

Tokunaga first looks at the definition of cyberbullying and finds discrepancies across the literature. The main point of disagreement is whether or not definitions include deliberateness and repetition of the hostile behavior. He points out that the discrepancy in definitions of cyberbullying throughout the literature make cross comparison or consolidating studies, so he proposes the following definition:

> "Cyberbullying is any behavior performed through electronic or digital media by individuals or groups that repeatedly communicates hostile or aggressive messages intended to inflict harm or discomfort on others. Additionally, the following addendum may be included with the definition.... In cyberbullying experiences, the identity of the bully may or may not be known. Cyberbullying can occur through electronically mediated communication at school; however, cyberbullying behaviors commonly occur outside of school as well" (278).

The definition of cyberbullying is derived from definitions of traditional bullying which generally includes the 3 conditions of 1. Repeated behavior, 2. Psychological torment, and 3. Carried out with intent. While most cyberbullying literature draws definitions and methodological approaches from studies on traditional bullying, Tokunaga points out the following differences between traditional and cyberbullying: Cyberbullying is seen more as an "opportunistic offense" since individuals who don't engage in traditional bullying will do so online since of the low threat of being caught. However, about half of cyberbullying cases involved perpetrators and victims who know each other and done so in a non-anonymous manner. Another distinction between traditional and cyberbullying is the lack of supervision online. While schools have clear enforcement agents, there is no set individual who regulates online behavior. Finally, cyberbullying allows for greater access to the victims even when not in the same location. Despite these differences, there is a significant number of bullies engage in both traditional and online bullying.

Across the literature, the prevalence of cyberbullying victimization is placed between 20-40% of youths reporting being victimized. In studies that cast the net more broadly, by asking about "mean things" being said about them online, as high as 70% of respondents reported experience. Tokunaga points out that in some cases the statistics are lower because the instances were not repeated. He also notes that a limitation of most studies is that there is no information collected on the duration of the victimization or the time elapsed between each encounter.

Measuring the role of age is also important for tailoring types of intervention to the age being intervened based on their distinct behaviors, but there is no consensus across the literature of what age has the highest rate of cyberbullying. Similarly, most studies had inconclusive findings or found that no particular gender is targeted in victimization, though a handful of studies found that females are disproportionately represented among online victims. This diverges from traditional bullying literature which finds that males are more involved both as bullies and victims. Females may be targeted more online than in traditional bullying since "males tend to bully others and be bullied through physical threats and aggression" whereas females tend to be implicated in bullying "involving psychological torment" (280), and the electronic medium lends itself more to this than face-to-face interaction. Despite the inconclusiveness of demographic variables, studies generally show that it is most frequent across gender in junior high school, so Tokunaga proposes that junior high employees should be trained to detect and remediate cyber bullying, but prevention programs should be implemented prior to 7th grade (280).

The effects on the victims of cyberbullying are similar to those of traditional bullying including drop in academic performance, strained family relationship, and development of psychosocial problems and affective disorders. However, it is not clear if cyberbullying is the cause of academic and psychosocial problems or if these characteristics are an antecedent of cyberbullying. In most cases there is some method of dealing with the cyberbullying experience (only around 25% of victims do nothing). These include technological coping (stricter privacy settings or changing online identities) which is frequently used, but its effectiveness is not known. Another strategy is the victim confronting the bully. It

is very infrequent that victims tell an adult (1-9% report telling their parents) due to fear of embarrassment or having their internet privileges suspended, instead friends are usually turned to for support.

Tokunaga argues that one of the greatest limitations of the literature on cyberbullying is the lack of theory building on the causes and effects of cyberbullying. Most studies assess who is the victim and perpetrator and what the effects of the bullying are, but fail to construct theoretical models that could better outline routes for intervention. He does note exceptions to this indifference to theory pointing to Damian Maher who takes a socio-cultural discourse approach to the learning process underpinning cyberbullying: "The socio-cultural discourse framework explains cyberbullying behaviors as a product of the minimal social cues, or anonymity, available on the online media through which the bullying occurs. Internet-supported technologies such as chat rooms, e-mails, and instant messengers offer fewer social cues than traditional interpersonal interactions, which renders divergent learning practices and behaviors." He also points to Social cognitive theory which could help explain how victims of cyberbullying become bullies themselves.

Overall, the greatest limitation of the study of cyberbullying for Tokunaga is the lack of an agreed upon definition, and his proposed definition seeks to address this issue.

Useful articles cited:
Maher, D. (2008). Cyberbullying: An ethnographic case study of one Australian upper primary school class. Youth Studies Australia, 27, 50–57.

****

Barlett, Christopher P., Douglas A. Gentile, and Chelsea Chew. "Predicting cyberbullying from anonymity." *Psychology of Popular Media Culture* 5.2 (2016): 171-180.

In this article Barlett, Gentile, and Chew are testing the validity of the Barlett and Gentile Model developed in their 2012 study *Attacking others online: The formation of cyberbullying in late adolescence* which argues that cyber bullying attitudes mediate the relationship between anonymity and cyberbullying behavior since users who realize their anonymity online will dissociate their online actions from their "real" self which will contribute to positive cyberbullying attitude and increased cyberbullying. The authors use Tokunaga's definition of cyberbullying :"any behavior performed through electronic or digital media by individuals or groups that repeatedly communicates hostile or aggressive messages intended to inflict harm or discomfort on others" (Tokunaga, 2010, p. 278). They also draw on Suler's theory of the online disinhibition effect which argues that online users will tend to detach their actions online from actions of their perceived 'real self,' this online disinhibition is partially a result of dissociative anonymity.

In reviewing the literature on cyberbullying, the authors find that in addition to anonymity, other predictors of cyberbullying are frequency of internet use, level of computer/internet ability, and risky internet use. They also point out that males in their late adolescence and early adulthood and children 12-15 years old are the most likely to engage in cyberbullying.

Colin Muller cam2299@columbia.edu

The authors hypothesize that perceived anonymity gives potential cyber bullies a sense of impunity online as well as empowerment since they can attack individuals they would not be able to attack offline either because of differences in physical strength or because they do not actually know or live near that person offline. Perceived anonymity leads cyberbullies to distance themselves from their own actions and gain a more positive perception of cyberbullying thus encouraging further cyberbullying.

The study's findings

The longitudinal study comprised of four waves of surveys that assessed 146 participants attitudes towards cyberbullying, their perceived anonymity, and their own cyberbullying behavior over time. In phases spread over 2 months, respondents rated their level of agreement with a series of statements related to cyberbullying perception and practice and perceived anonymity "to test whether aggressor-perceived anonymity in the cyber-world is related to subsequent cyberbullying, and if positive cyberbullying attitudes mediated these relations" (172).

The findings demonstrated that perceived anonymity was indeed related to cyberbullying behavior as those who felt more anonymous online were more likely to cyberbully. Furthermore, the findings suggest that "anonymity supports positive feelings regarding cyberbullying behavior, leading to the manifestation of those attitudes in behaviors" (177).

Potential interventions

The authors argue that their findings confirm the "online disinhibition effect" whereby online users compartmentalize their "online self" and "real life" self and the normal cognitive processes that guide their "real life" behavior are suspended when they are online. The authors propose intervening against cyberbullying by informing "Internet users that they are not anonymous and show them evidence of IP address tracking and how History folders operate, then perhaps cyberbullying will decrease" (178). However this form of intervention would not be effective in cases where users are actually able to be anonymous online.

Another finding is that "cyberbullying attitudes mediate the relation between anonymity and later cyberbullying behavior" (178) which means that individuals internalize a positive attitude towards cyberbullying as they learn how anonymous their cyberbullying behavior is. So intervention could also take the form of downplaying positive perception of cyberbullying on online platforms, but the authors fail to propose concrete implementations of this intervention.

Limitations

A limitation that the authors point out is that there are other variables that can mediate the relationship between a user's perceived anonymity and cyberbullying behavior such as normative aggressive beliefs and empathy. Another limitation is that the study doesn't clearly define anonymity and distinguish it from other categories such as pseudonymity or perceived anonymity. For example one of the statements that participants were asked to rate on level of agreement to gauge anonymity was : "Sending mean emails or text messages is easy to do because I am not face-to-face with the other person." but not being face to face does not constitute anonymity if your email address of cell number reveals your identity. Other limitations in the methodology is that a majority (78%) of participants were female and the gap between the 4 surveys was not long enough to gauge longer term change over time.

Overall, while the theory and the methodology is compelling, I feel like the author's loose use of the term anonymity conflates dissociative anonymity with more general feelings of inhibition and impunity users feel online even when they are cyberbullying on platforms that are tied to their identity.

Colin Muller cam2299@columbia.edu

Useful Articles that they cite:

Barlett, C. P., & Gentile, D. A. (2012). Attacking others online: The formation of cyberbullying in late adolescence. Psychology of Popular Media Culture, 1, 123–135. doi:10.1037/a0028113

Suler, J. (2004). The online disinhibition effect. CyberPsychology and Behavior, 7, 321–326. doi:10 .1089/1094931041291295

Tokunaga, R. S. (2010). Following you home from school: A critical review and synthesis of research on cyberbullying victimization. Computers in Human Behavior, 26, 277–287. doi:10.1016/j.chb .2009.11.014

Wright, M. F. (2013). The relationship between young adults' beliefs about anonymity and subsequent cyber aggression. Cyberpsychology, Behavior, and Social Networking, 16, 858–862.

Santana, D. (2014). Virtuous or vitriolic: The effect of anonymity on civility in online newspaper reader comment boards. Retrieved from http:// www.tandfonline.com/doi/abs/10.1080/17512786 .2013.813194#.U1qNBZVZplYonApril,25,2014

Wright, M. F. (2013). The relationship between young adults' beliefs about anonymity and subsequent cyber aggression. Cyberpsychology, Behavior, and Social Networking, 16, 858–862

Slonje, R., Smith, P. K., & Frisen, A. (2012). Processes of cyberbullying and feelings of remorse by bullies: A pilot study. European Journal of Developmental Psychology, 9, 244–259. doi:10.1080/ 17405629.2011.643670

****

Colin Muller cam2299@columbia.edu

Levy, Nathaniel, Sandra Cortesi, Urs Gasser Edward Crowley, Meredith Beaton, June Casey, and Caroline Nolan. "Bullying in a networked era: A literature review." (2012). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2146877

In September 2012 The Berkman Klein Center published *Bullying in a Networked Era: A Literature Review.* The paper provides an overview of much of the literature on cyberbullying, but the second section "What can be done about bullying" focuses almost entirely on school and parental intervention and attributes little role to the online platforms.

The review points to Olweus's definition of bullying as "an aggressive act with three hallmark characteristics: a) it is intentional; b) it involves a power imbalance between an aggressor (individual or group) and a victim; c) it is repetitive in nature and occurs over time" (8).

This defintion of bullying encompases a variety of forms of aggression which include:

- Physical contact, words, or faces or obscene gestures may be means of bullying (Olweus, 1994).
- "Proactive" aggression is usually unprovoked, instrumental, and goal-directed – for instance, a bully may want to gain power, property, or a certain affiliation or relationship status (Price & Dodge, 1989; Espelage & Swearer, 2003; Griffin & Gross, 2004).
- "Reactive" aggression can be a defensive or angry response to a threatening, angering, or frustrating event (Price & Dodge, 1989; Espelage & Swearer, 2003; Griffin & Gross, 2004).
- "Indirect" or "relational" aggression uses rumors, gossip, secrets, and social exclusion as means of harming (often humiliating) the victim. Stated simply, relational bullying occurs through relationships (Crick & Grotpeter, 1995; Espelage & Swearer, 2003; Griffin & Gross, 2004; Low, Frey, & Brockman, 2010; Mishna, Cook, Saini, Wu, & MacFadden, 2010).
- "Bias-based" bullying (also referred to as aggression or harassment) refers to bullying that co-occurs with discriminatory prejudice such as racism, sexism, and homophobic teasing. The term also reflects the understanding that bullying and such forms of discrimination often converge (Russell, Sinclair, Poteat, & Koenig, 2012).

The main differences between traditional forms of bullying and cyberbullying is that a cyberbully's identity is less likely to be known, and cyberbullying is more likely to take place outside of school.

In a lot of the ethnographic data collected it was found that "teens do not identify with the bullying or cyberbullying rhetoric" but rather use terms like 'drama' which can blur the line between serious and non serious conflict as well as blurring the distinction between bully/victim/bystander (10). For example, most studies find low rates of experiencing cyberbullying among respondents (<25%), however when respondents are asked about instances of cruelty or meanness on social networking sites, as high as 88% report experiences (16).

Comparing online and Offline Bullying

The authors of the review assert that the degree of anonymity of cyberbullies is often overstated, but anonymity is something that can contribute to the imbalance of power that is central to definitions of bullying. The power of anonymity can sometimes be substituted for other forms of power that the bully lacks--such as strength, intelligence, popularity etc.

While central to most definitions of bullying, in some cases the concept of **repetition** of harmful acts is less applicable in online bullying. This is because a post can be made once but stay online and

continue to harm the victim repetitively. This is arguably the case with the offline bullying method of writing a nasty message on the bathroom wall, but in the cases of notes in bathrooms, researchers have not viewed this as meeting bullying's definitional component of repetition. Some researchers have tied the question of repetition to the size of the audience or bystanders, since bystanders online can play a role in the impact of the harassment (ex: is it a public post or private message?). While most studies fail to distinguish between different online activities, online bullying occurs more often using social messaging than social networking sites, suggesting that more cyberbullying happens privately between the bully and victim than in front of virtual bystanders.

One significant difference between online and offline bullying is that the distinction between bully and victim online is not as rigid since it is common for a victim to retaliate against their bully online which would then qualify as "reactive" bullying. However, there are some forms of cyberbullying (anonymous messages or creating hostile websites) that are not conducive to online retaliation. In many surveys around ⅓ of respondents reported being involved both as bullies and victims in cyberbullying. Relational aggression is the type of bullying that is most common among bully-victims. One survey found that 50% of cyberbullies surveyed were also victims of online bullying (21), and researchers have argued that being the victim and perpetrator of cyberbullying predict one another.

Like offline bullying, poor social problem-solving skills, poor academic performance, and being less well adjusted are associated with cyberbullying. These characteristics are found in bullies, victims, and bully-victims which makes it hard to make a one-way causal argument. Offline bullying reaches its peak during middle school whereas online bullying peaks in high school.

Norms around bullying

The concept of "reciprocal socialization" is useful for explaining where bullying behaviors might come from. The theory holds that bullies tend to associate with peer groups that show a lot of aggressiveness between themselves. I think a good example of this is online communities like 4chan where within the community there is a culture of hostility and aggressiveness towards each other which socializes them for their attacks on outsiders. What reinforces the theory of reciprocal socialization is findings that children who tend to bully are part of peer groups who identified "popular and aggressive peers as cool" whereas non-bullies identified popular and nonaggressive peers as cool. (19). It was also found that students are more likely to perpetrate bullying when they perceive bullying as occurring frequently and being socially acceptable in their school (30).

LGBTQ youth and youth with disabilities are more likely to be victimized by bullying and also tend to be more psychologically affected by the bullying. Contrary to earlier theories of bullying, girls don't engage more in "indirect" or "relational" bullying than boys. Overall boys are more involved in offline bullying as both victim and perpetrator, but online the gendered difference is unclear (17). Surprisingly, the research shows that bystanders to bullying offline will reinforce the bully more often, whereas in cases of cyberbullying, bystanders are more likely to either ignore the instance of cyberbullying or report or confront the bully.

Addressing Bullying

Research shows that rigid disciplinary rules such as "zero tolerance" do not effectively reduce bullying in schools, but some degree of discipline coupled with social-emotional learning seems to be the most effective approach to school policy. The school's social climate and students' feeling connected and supported by their school's community is also important for preventing bullying.

While cyberbullying may raise concerns about teens' online activity in general, nearly all studies show that students also have positive experiences online including interactions that make them feel good about themselves or being introduced to new perspectives on social and political issues. Therefore teachers and parents should try to understand students' social environment to foster a sense of 'digital citizenship' rather than entirely banning online activity. This is related to one of the reasons that most victims of bullying do not report it to their parents, because they are afraid that reporting it will result in their loss of access to the web.

The review mentions the following two articles as proposing a framework for differentiating between different sorts of online peer victimization and harassment:

Finkelhor, D., Turner, H.A., & Hamby, S. (2012). Let's prevent peer victimization, not just bullying. Child Abuse & Neglect, 36(4), 271-274.

Hong, J. S., & Espelage, D. L. (2012). A review of research on bullying and peer victimization in school: An ecological system analysis. Aggression and Violent Behavior, 17(4), 311-322.
****

Franks, Mary A. *Sexual Harassment 2.0,* 71 Md. L. Rev. 655 (2012)Available at: http://digitalcommons.law.umaryland.edu/mlr/vol71/iss3/3

Franks reviews "classical" sexual harassment doctrine which is limited to single-setting scenarios. This means that protection against sexual harassment is largely confined to places like schools, workplace, and prisons. Under current sexual harassment law, there are limited protected spaces along two axes: spatial and relational. The spatial axes means that harassment is only protected if it occurs in a protected setting (work, school, prison, home) and the effects of the harassment are felt in the same setting. The relational axes means that harassment is only protected from a limited set of individuals who are under the control of the agents who exert control over the protected settings (such as employer or school administrator). For example, this means that in the workplace, employees are protected against harassment from employees as well as other third parties over which the employer has control such as customers in a restaurant. The scope of the setting has been expanding in recent court rulings as it is found that "Conduct that takes place outside of the workplace has a tendency to permeate the workplace" (p. 667 citing Blakey v. Continental Airlines, 751 A.2d 538, 549 (NJ 2000)). Despite expansion of the understanding of setting, relational axis is still very limited to people directly related to the protected setting and fails to incorporate cases of harassment where it occurs outside of the protected setting by unrelated individuals but affects the victims in their protected setting. An example of this is a woman being harassed on the subway by a stranger on her way to work.

Franks argues that the major limitation of sexual harassment law as it stands is that it only covers cases where either "the action and the effects of the harassment occur in the same protected setting" (p. 671) or where there is "off-site harassing action but with targets and harassers strongly connected to the protected setting" (p. 672). What is excluded from the current framework are cases where 1. "The harassing activity takes place in an unprotected setting but produces effects in the protected settings" or 2.

cases where "both the harassing activity and its harmful effects occur in unprotected settings" (p. 672). The limitation of the current sexual harassment legal framework is particularly relevant for online sexual harassment and Franks proposes a multiple-setting conception of sexual harassment which will better cover harassment that "occurs completely or partially outside of traditionally protected settings" (p. 655).

The relevant law that prevents extending discrimination law to the internet is Section 230 of the Communication Decency Act (CDA) which says that "no provider or user of an interactive computer service shall be treated as the published or speaker of any information provided by another information content provider." This has immunized website hosts from being liable for their users illegal activity related to sexual harassment. Frank argues that CDA should be amended to treat website hosts and other Internet entities like employers or school administrators since they are the intermediaries who can exert control over those settings. This could be done simply by including explicit language in section 230 of the CDA about compliance with federal discrimination law, since the "immunity" that web hosts currently have is already limited by other federal laws such as copyright violations.

****

Huang, Guanxiong, and Kang Li. "The effect of anonymity on conformity to group norms in online contexts: a meta-analysis." *International Journal of Communication* 10 (2016): 18.

This article starts by rejecting the antiquated view of mass psychologists such as LeBon which holds that a mass of people behave irrationally, negatively and dangerously as they feel less social obligation in their anonymity. The essay instead seeks to test the validity of the Social Identity model of Deindividuation Effects (SIDE) model which rejects LeBon's "deindividuation" and instead supports "depersonalization." "Depersonalization refers to the process through which individuals perceive that their certain group identity is more salient than other identities in a particular context, termed as 'the emergence of group in the self' (399).

The article emphasizes that conformity caused by anonymity is especially strong when there is an in-group/out-group dynamic and the correlation is also dependent on the type of anonymity (visual, physical or personal)[1], with visual anonymity having the strongest correlation with group conformity. The study used a meta-analytic approach on 13 articles where they searched for academic studies relating to the effects of anonymity on group behavior and coded different attributes in each article such as type of anonymity, effect size, and sample size. They had mixed results, but found a positive correlation between anonymity and outgroup identification and group conformity (with visual anonymity being the clearest causal variable—visible anonymity is not seeing a partner's picture on a screen while using text based communication). A limitation of this study that the authors pointed out was that meta-analysis privileges quantitative studies with significant findings; so studies with findings that have insignificant findings or that do not support a theory might not be published or might be published using non-quantitative methodology. I thought the study was rather interesting because it serves to further disprove crowd theories which in my opinion are rather anti-democratic and disdainful towards the average person. The

[1] Visual anonymity: not being able to see others' image on a screen
Physical anonymity: not being in the same room as others'
Personal information anonymity: not knowing others' personal information

article also did a good job of not ascribing much normative value to the notion of conformity, noting that it can be a force for greater respect and cohesion in a group, not just a violent mob-mentality.