Digital Ethics
Rhetoric and Responsibility in Online Aggression

Edited by
Jessica Reyman and Erika M. Sparby
2020

1 Introduction

Toward an Ethic of Responsibility in Digital Aggression

Jessica Reyman and Erika M. Sparby

Digital aggression in today's digital world is widespread, lacking easily traceable sources and causes and without clear paths for accountability and resolution. In July 2017, the Pew Research Center reported that four in ten U.S. adults (41%) have experienced online harassment and many more have witnessed it. From name calling and public shaming to physical threats and stalking, much of this activity targets women, transpeople, and members of racial or ethnic minority groups. Research shows that one in four Black people has been the target of harassment, and women are twice as likely as men to be harassed online (Duggan, 2017). In addition, religious views and political views and affiliations have also been the basis for aggression. In 2014, GamerGate brought the gendered tensions and gatekeeping practices present in some gaming communities to the attention of the mainstream public. The 2017 white nationalist demonstration in Charlottesville, VA (which was organized largely through Discord, a voice and text chat app), and the online political and public responses following it gave renewed visibility to hate groups. Automation and algorithmic technologies present their own ethical quandaries surrounding digital data. The initial release of Apple Watch was unable to work properly on heavily tattooed or darkly pigmented skin (Profis, 2014), and Fitbit data has been used to catch cheating spouses (Pilon, 2015) and reveal pregnancy (Jackson, 2016). Twitter bots have been credited with influencing the 2016 U.S. presidential election. Facebook's Cambridge Analytica data privacy scandal revealed how the extent of the personal information and data collected through Facebook can be mismanaged and subject to abuse (Granville, 2018).

Within the current context, questions arise about response, responsibility, and accountability that the field of digital rhetoric is uniquely poised to address. The chapters in this collection, as a whole, build on what Porter (1998) has called "rhetorical ethics," which do not constitute a moral code or a set of laws but rather a "set of implicit understandings between writer and audience about their relationship" (p. 68). While Porter's work appeared before the rise of social media, automation, widespread digital data collection, and other contemporary web contexts, we have more recently seen how these implicit agreements

extend beyond writer and reader (who often occupy both roles) to also include the individuals, communities, and institutions that build and manage technological spaces for discourse and engagement. Furthermore, as Brown (2015) argues, digital platforms, networks, and technologies themselves carry ethical programs with rhetorical implications.

Following the Association of Internet Researchers' "Ethical Decision-Making and Internet Research" (Markham & Buchanan, 2012), this collection considers digital ethics as deeply embedded within rhetorical contexts. Approaches from our authors take into account Markham and Buchanan's (2012) questions:

- Does the research definition of context match the way owners, users, or members might define it? …
- Are there distinctions between local contextual norms for how a venue is conceptualized and jurisdictional frameworks (e.g. Terms of Service, other regulations)? …
- What are the ethical expectations users attach to the venue in which they are interacting? (p. 8)

In doing so, the chapters in this collection seek to understand ethics within dynamic digital ecosystems and ecologies. The authors demonstrate the value of casuistic approaches to studying digital ethics, considering complex ethical issues by analyzing the tensions among regulation-driven and context-driven considerations.

In a time when more people are on Facebook than live in the country of China, when people wear social media in their watches and carry it in their pockets, and when the U.S. president uses Twitter as a political platform, attention to digital ethics is more important than ever. How users participate online and through digital media and how researchers and scholars theorize about ethical participation have the potential to shape the norms, laws, and practices that will determine the future of the social web and digital data. This edited collection provides a discussion of what principles, based on research and theory in rhetoric and composition, *should* guide our thinking about responsibility, accountability, and ethics. Through examinations of unethical practices in digital spaces and through digital technologies, this collection contributes to the field's research and theorizing about ethical participation by providing contextualized, case-based analyses of varying forms of digital aggression and exploring the tensions inherent in minimizing harm in a digital age.

Ethics and Digital Aggression in Rhetoric and Writing Studies

Ethical issues have captured the attention of rhetoric and composition scholars from early iterations of the web to present day. *Foregrounding*

Ethical Awareness in Composition and English Studies (Fontaine & Hunter, 1998) offered a collection of essays on ethical issues relevant to teaching and administrating composition classes and programs at a time when the field was reportedly taking an "ethical turn." In a review essay titled "The Ethical Turn in English Studies" (1999), Harrington asserted a rise in interest in ethics at the time. While she observed that "current discussions vary considerably in approach as well as well as how ethics relates to what we do as teachers and scholars," she located its roots in the connections developing at the time between administrative practices and pedagogical choices to the social, cultural, and political.

At that same time, the web was emerging as a force for communication, civic discourse, public activity, and education, leading rhetoric and composition researchers to define and theorize the concept of "rhetorical ethics" for online rhetoric (Porter, 1998) and later a "digital ethic" (DeVoss & Porter, 2006) for new online environments. New environments for reading and writing presented new ethical issues for consideration by rhetoric and composition scholars. Privacy online was identified early as an ethical issue of rhetorical import by Gurak in *Persuasion and Privacy in Cyberspace* (1999), and has since been examined by Beck et al. (2016). Digital copyright and authorship have been widely examined as contentious legal issues rife with ethical implications (Herrington, 2010; Logie, 2006; Reyman, 2010; Rife, Slattery, & DeVoss, 2011; Westbrook, 2009). Johnson-Eilola (2010) has written about the implications of "spimes" and user data for writers and readers in "Among Texts," where he concludes with "an incomplete list of concerns" that includes issues of privacy and ethics. A 2011 special issue of *Computers and Composition Online* titled "Ethics in Digital Age: Ethics and Digital Media in the Writing Classroom" (Coley, 2011) presents discussion of "ethical literacy" for a digital age, which includes discussions about ethical issues presented by both student composing activities such as photo manipulation and remix and instructor activities such as adopting course management systems. These particular ethical issues—of privacy, of copyright, of remix and attribution—have attracted the interest of scholars interrogating ethics within digital ecosystems and ecologies for composing, communication, participation, and engagement.

As far back as Selfe and Selfe's "Politics of the Interface" (1994), digital rhetoric has been concerned with the ways technologies and interfaces reinscribe dominant ideologies and power structures by their very design, and equally important are explorations of how humans can perpetuate these dynamics through discourse and participation in online spaces. More recently, digital ethics scholarship within rhetoric and composition studies has examined these concerns within the context of human–machine collaborations. Brown's *Ethical Programs* (2015) offers a compelling and extended analysis of the ethics of networked

software, showing how software promotes particular arguments and advances an ethical agenda, thus contributing to the complexities of locating responsibility and accountability within and across digital ecologies. Automated systems and algorithms have been addressed as an ethical quandary for rhetoricians studying rhetorical agency (Kennedy, 2016; Miller, 2007; Reyman, 2010).

Other recent work addresses digital aggression through examinations of case studies, analysis of examples, and observations of user activity in online communities. Warnick and Heineman's *Rhetoric Online* (2012) examines cyberterrorism to contribute to an understanding of why and how digital aggression can be so persuasive, revealing the global and political implications for damaging digital discourses (Warnick & Heineman, 2012). Several scholars have observed gendered aggression in online spaces (Cloud, 2009; Jane, 2014; Milner, 2013; Phillips, 2015), including Poland (2016), who provides a first-person exploration of cybersexism and harassment in *Haters: Harassment, Abuse, and Violence Online*. She follows Phillips's (2015) nuanced treatment of the topic, examining different types of trolls and aggressors and acknowledging the intersectionality of harassment and abuse. Clinnin and Manthey (2019), Gruwell (2017), and Sparby (2017) connect what we know (and can learn) about digital aggression to practical applications for teaching writing. They urge writing teachers to engage students in becoming more civically engaged in the digital spaces they occupy.

The field of rhetoric and composition is only just beginning to grapple with and untangle the implications of the widespread abuse and harassment distributed across the social web, the ethical dilemmas presented to us in the writing and rhetoric classroom or through our research, and the complex questions surrounding automation, wearable technologies, and issues of privacy and surveillance. Poland (2016) offers a call to action: "Those of us with the power to do research, educate others, enforce consequences, and build safer spaces have a responsibility to do so" (p. 252). The authors in this collection aim to meet this challenge, offering new frameworks for digital ethics from a uniquely rhetorical perspective.

Toward an Ethic of Responsibility in Digital Aggression: From Not Feeding the Trolls to an Ecology of Response

Typical advice and calls for civility amidst aggression, hate speech, and harassment are frequently too-optimistic, misguided, and ineffectual. Phillips and Milner (2018) critique sentiments such as, "[i]f only people would lower their voices, stop posting rude memes, and quit with the name-calling, we could start having meaningful conversations. We could unite around our shared experiences. We could come together as a nation." Poland (2016) brings attention to the problematic nature of

a similar adage for responding to digital harassment: "don't feed the trolls." The idea behind these approaches is that because those who spew hate and harass others online feed on attention, if users simply ignore the comments and behaviors, aggressors will get bored, cease their behavior, and go elsewhere. However well-intentioned, this urging toward civility is inadequate because it flattens contexts, puts an emphasis on intentionality over effect, and can silence the targets of aggression, including already marginalized voices.

Put simply, existing approaches to address digital aggression fail today's digital media users. In contemporary digital contexts, the number of users has risen so dramatically and the boundaries between digital communities are more fluid and diverse than ever before. In *The Internet of Garbage*, Jeong (2018) notes the ways in which such fluidity can contribute to digital aggression:

> When looking through the lens of online harassment, the internet is simply too small. When one platform links to another platform in these cases, it creates a pipeline of hate with very little friction. Even if the targeted platform maintains certain norms, the oncoming invaders ignore them, operating only under the norms of their originating platform. A simple Google search can connect together all the disparate aspects of a person's digital life, allowing bad actors to attack each and every part even without knowing them particularly well to begin with.
>
> (p. 74)

As users move quickly and easily between platforms and spaces, often gaining membership in multiple communities with varying and sometimes conflicting goals, purposes, and values, it's no longer easy to maintain an exclusive online space and drive out those with dissenting opinions or who don't fit within narrow identity categories for participants. Digital spaces are, therefore, rife with conflict, sometimes productive but at other times manifesting as aggression, harassment, and abuse. Such conflict is not necessarily a bad thing, as Milner (2013) points out, because disengagement of marginalized voices can threaten the promise of the Internet as public sphere, and "voice should be evident over exclusion, even if that voice is not monolithic in content and tone."

In these spaces, the advice to "be kind" or "don't feed the trolls" can mean "don't engage" or "don't point out abuse and its effects." In this way, calls for civility problematically place blame for aggression on the target, not the aggressor. Such advice suggests to targets that because speaking within digital spaces leads to harassment, they should no longer speak or even inhabit those spaces at all. Often when someone responds to hostile digital discourse, they are met with "what do you expect when you feed the trolls?" or another such victim-blaming sentiment. In such a

way, calls for civility can lead to silencing some of the most marginalized voices online. As Poland (2016) explains:

> Online threats of violence seem to have a very simple purpose: they are intended to act as a reminder to women that men are dominant, that women can be attacked and overpowered if men choose to attack, and that women are to be silent and obedient. Many threats contain ultimatums: if a woman doesn't stop engaging in activities that the men issuing threats find undesirable, she will be punished with physical violence or even death. The intent of threats is to establish offline patterns of violence against women in online spaces. (p. 53)

Calls for civility can downplay the seriousness of the behaviors that some people experience online. "Troll" sounds playful, harmless even, and belies the serious nature of the range of behaviors that are often grouped under it. But actions like violent rape and/or death threats, stalking, and doxxing—things that invalidate identities and make users feel unsafe beyond digital spaces—are often grouped under this term. "Don't feed the trolls" can suggest that aggressive behaviors shouldn't be taken seriously, and that the effects of abuse aren't real and important to the people who experience them.

Platform providers often approach managing and responding to digital aggression through after-the-fact action and enforcement of punishments rather than through the cultivation of guiding values and preventative measures. Social media platforms, such as Twitter, offer reporting mechanisms for those who experience harassment and technical tools users can employ to block harassers and aggressive content. These mechanisms, however, are limited in their aims and outcomes, and do not do enough to protect the most vulnerable groups (Jhaver et al., 2018). Roberts (2016) has examined the effects that commercial content moderation has on the workers that perform the labor involved in the curation and decision-making processes, what she calls the "dirty work of social media." Immersion in aggressive content—racist, homophobic, misogynistic, and disturbing—can leave workers vulnerable to the potentially devastating effects of daily exposure. Understanding how digital aggression is managed, often through inadequate technical tools or through potentially harmful labor practices that are largely hidden from the general population of Internet users, reveals the limitations of existing approaches. While acknowledging that "content moderation is hard," Gillespie (2018) asserts that social media platforms must embrace their role as "custodians of the Internet":

> This would be a very different understanding of the role of 'custodian'—not where platforms quietly clean up our mess, but where they take up guardianship of the unresolved tensions of public

> discourse, hand back with care the agency for addressing those tensions to users, and responsibly support that process with the necessary tools, data, and insights.
>
> (p. 211)

Such an approach requires a new ethic of responsibility to guide the development of platforms, the management of digital communities, moderation practices, and responses to digital aggression. An ethic of responsibility calls for more engagement rather than less, for value in designing for protection against digital harassment rather than after-the-fact cleanup, for accountability and tactical response rather than civility within digital contexts. From platform designers, developers, and managers, to digital community leaders, to everyday users, to content moderators, to policymakers and legal experts, diverse actors must become more aware of their own positionality within particular spaces and moments; the consequences of their decisions, words, and actions; and the embodied experiences of users with which they engage across diverse networks of digital communities. Value systems and ethical principles must be considered from the point of design of platforms, sustained through the careful development and management of communities, and supported through appropriate corrective actions.

We posit that productive digital communities be proactive and act in ways that recognize the multidimensional nature of aggression. Figure 1.1 displays a comprehensive view of such an approach.

1 *Platform designers and developers*. First, we find some responsibility with platform providers, technology developers, and media

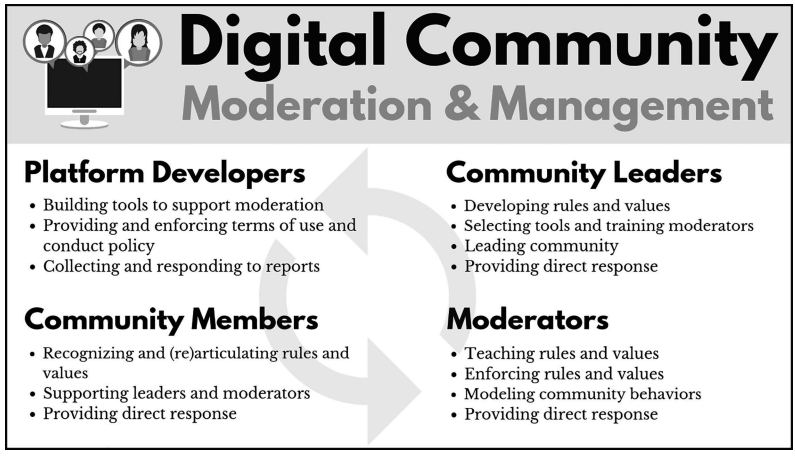

Figure 1.1 Digital community moderation and management (Developed by E. M. Sparby, 2019).

companies to design and offer more powerful tools for moderation and management of aggression, as well as harassment policies and terms of use that can be drawn on to support community members. There must be transparency in actions being taken, tools employed, and terms of use enforced. Some design decisions could benefit from including more diversity in the hiring of developers, managers, and entrepreneurs, as lack of diversity can lead to design decisions with unintended consequences for minority groups and vulnerable populations (Noble, 2018).

2 *Community leaders*. Second, community leaders and content creators themselves must clearly establish and articulate the values and norms of their communities. Leaders can employ technical tools and moderation options made available to them through the platform as well as communicative practices that support values of inclusivity and productive discourse. Community leaders are uniquely positioned to establish these values through the generation of codes of ethics, community mission statements, and through direct responses to transgressions.

3 *Moderators*. Third, human moderators who participate in digital communities can consistently enforce the rules and values of those communities. They can accompany corrections of transgressions and enforcement of rules with reminders to users of the connections between rules and the values established by community leaders. Their actions and comments should aim to support a culture based on a shared value of inclusivity rather than more limited rule-following.

4 *Community members*. Fourth, members and participants who are not official moderators can also help to reinforce values, norms, and rules, and teach others. By distributing the activity of moderation among many participants, community leaders and moderators do not suffer the sole burden. Rather than remaining silent or "not feeding the trolls," participants can respond to aggression with clear articulation of shared values.

This model acknowledges that the multidimensional nature of digital communication requires an ecological response to digital aggression. Much of this work must come from platform designers and developers and social media managers and entrepreneurs, who must build tools that (1) protect vulnerable groups and (2) allow for collective and collaborative management of digital aggression by users and communities. For instance, Citron and Wittes (2017) ask us to consider the impact of features such as shared block lists and Gillespie (2018) proposes *collective* flagging systems where users can identify content as "racist" or "sexual" or "violent," which would then lead to actionable data such as content warnings. Offering design and moderation options would allow users to collectively employ tools and tactics to establish (and reestablish, when