2 ‘Don’t Feed the Trolls’

Social Media and the Limits of Free Speech

Tom Clucas

Abstract
This chapter examines ‘comment culture’ using examples from social media platforms and the tabloid press. It explores the shift from a model of free speech as a collective responsibility to one that presents it as an absolute individual right. The chapter examines the consequences of this shift by analyzing a series of vitriolic exchanges on its chosen platforms. Three main trends emerge: first, social media have developed a unique comment culture focussed on combat, disinhibition and the contest for popularity. Second, online platforms are governed by their own distinctive ‘thread logic’, which disdains rational argument in favour of passionate display. Third, the conventional rules of conversation and argumentation have been disapplied, with consequences for communities both on- and offline.

Keywords: online vitriol, social media, free speech, comment culture, public sphere

This chapter examines ‘comment culture’ using examples drawn from the comment sections on YouTube, Twitter, and Facebook, as well as in the British tabloid press. These comment sections aim to make the websites more interactive by enabling individuals to post comments on other users’ content (e.g. YouTube videos, tweets, Facebook posts, or news articles). The comments are generally gathered beneath the original post in a thread that often also allows users to respond to one another’s comments. The chapter argues that these comment sections develop their own particular cultures – sets of generally established rules and practices about the content, tone, and format of comments – which are tacitly agreed between the community of users on the platform. However, there is also a tendency for certain

Polak, Sara, and Daniel Trottier (eds), *Violence and Trolling on Social Media*. Amsterdam, Amsterdam University Press 2020
DOI: 10.5117/9789462989481_CH02

individuals to flout these conventions by posting deliberately provocative, aggressive, or hate-filled material. As a result, the comment sections are often particularly rife with examples of online vitriol, as individuals take advantage of anonymous posting options and/or disguised user names to post vitriolic comments that would not be tolerated in face-to-face conversation.

The chapter begins by briefly surveying modern debates about free speech, from John Milton's *Areopagitica* to Timothy Garton Ash's *Free Speech*. In the process, it posits a recent shift from a model of free speech as a collective responsibility to one that presents it as an absolute individual right. This strong model of free speech, coupled with users' ability to post anonymously online, helps to shape the cultures that evolve in the comment sections of various online platforms. Although each platform has its own unique community of users and set of rules governing the comments section, three broad trends can be seen to emerge. First, social media have developed a unique 'comment culture', in which largely self-regulating communities of users reach a consensus about what forms of expression are or are not available. Recent research by Jude P. Mikal and others suggests that:

> As individuals learn the generally accepted rules of conduct associated with the site, their online interactions will shift to reflect their social – rather than individual – identities. The result will be a common voice emanating from the website: a generally cohesive tone, characterized by overall consistent responses, and overt behaviour correction.[1]

Partly, this comment culture is established by the host of the site, through moderation and rules governing the site's usage. However, the culture is also the product of communities of users responding to one another's posts with praise or blame. Often, it is the site's users who decide whether or not to report a particular comment to the moderators. As a result, the community of users has a large say over the extent to which hate speech and online vitriol are tolerated.

The second trend is that online platforms are governed by their own distinctive 'thread logic', in which traditional models of debate are supplemented by appeals to humour and the popular phenomenon of 'trolling'. In his study of this phenomenon, Whitney Phillips observes that:

> Engaging in racism or sexism or homophobia, disrupting a forum with stupid questions, or generally being annoying does not automatically make

1 Mikal et al., 'Common Voice', p. 506.

> one a subcultural troll. Trolling in the subcultural sense is something a self-identifying troll sets out to do, as an expression of his or her online identity.[2]

Real trolling aims to disrupt accepted practices of online debate and conversation for the sake of 'lulz', which Phillips defines as 'a particular kind of unsympathetic, ambiguous laughter'.[3] As discussed below, the phenomenon of trolling is problematic because internet users often conflate it with hate speech. When forms of online vitriol (including those that relate to racism, sexism, or homophobia) are mislabelled as 'trolling', there is a tendency for users to downplay or ignore them.

The third trend is that the conventional rules of conversation and human interaction are often modified online, with some users being more willing to resort to insult and abuse when they have the ability to appear anonymously. This is reinforced by the fact that online users are often hesitant to 'feed the trolls' by calling out examples of hate speech or online vitriol: such displays of aggression are often mislabelled as trolling and allowed to remain on the site without being censured or removed. The chapter concludes by considering the implications that these recent developments might have for offline modes of communication.

The effects of online vitriol were powerfully demonstrated during the 2017 General Election campaign in the UK. During this campaign, the Shadow Home Secretary Diane Abbott Member of Parliament (MP) was subjected to extensive scrutiny by the mass media and the electorate. Abbott became the first black woman elected as a UK MP when she won her seat in 1987. However, in an article published in *The Guardian* on 14 February 2017, she revealed that her role had become harder in recent years due to the daily racist and sexist abuse she receives online. Abbott wrote that 'I have received rape threats, death threats, and am referred to routinely as a ***** and/or ******, and am sent horrible images on Twitter.' The death threats include an English Defence League-affiliated account with the tag 'burn Diane Abbott'.[4] It is likely that some of these comments constitute hate speech and can be reported under the UK hate speech laws in Part 3 of the Public Order Act 1986. However, the sheer scale of hate speech on the internet makes it impractical to place the burden solely on victims to report each incident. On a cultural level, it is important to understand the

2 Phillips, *This Is Why*, p. 24.
3 Ibid.
4 Abbott, 'I Fought Racism'.

possible causes of this epidemic of online vitriol. This chapter proposes two interrelated factors: first, a strong model of free speech as an absolute, individual right and second, a rapid rebalancing of the equilibrium between the private and public spheres encapsulated by YouTube's catchphrase: 'Broadcast Yourself'.

The right to hate

Recently, there has been a shift from a model of free speech as a collective responsibility to one that presents it as an absolute individual right. This marks a significant departure from the model of free speech that prevailed from the early modern period into the twentieth century. In an often-quoted formulation from his prose polemic *Areopagitica; A Speech* [...] *for the Liberty of Unlicenc'd Printing* (1644), the poet John Milton exclaimed: 'Give me the liberty to know, to utter, and to argue freely according to conscience, above all liberties.'[5] What is remarkable in this quotation is its emphasis on 'conscience'. Milton stresses that the right to speak freely comes with the attendant burden of accountability for what is said. Similarly, John Locke argued in 'A Letter Concerning Toleration' (1690) that:

> [N]o private person has the right to attack or diminish another person's civil goods in any way because he professes a religion or ritual differing from his own; all of that person's human rights as well as his civil rights are to be scrupulously observed.[6]

Locke wrote his 'Letter' centuries before the European Convention on Human Rights came into force in 1953 and before the US enacted the Civil Rights Act in 1964. His conceptions of 'human rights' and 'civil rights' inevitably differ from those in the present, yet they remain expansive in encompassing 'life, liberty, physical integrity, and freedom from pain, as well as external possessions, such as land, money, the necessities of everyday life, and so on'.[7] What unites these two early texts is their interpretation of free speech as a collective responsibility as well as an individual right. In order for free speech to function, Milton and Locke contend that it must be moderated by individuals respecting one another's mutual rights.

5 Milton, *Areopagitica*.
6 Locke, 'A Letter Concerning Toleration', pp. 12-13.
7 Ibid., p. 7.

In principle, this model of free speech as a collective responsibility remains inscribed in European law. Article 10 of the European Convention on Human Rights, which governs 'freedom of expression', provides that 'Everyone has the right to freedom of expression', but that the

> exercise of these freedoms, since it carries with it duties and responsibilities, may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society.[8]

With the advent of the internet and the growth of the tabloid press, however, the popular conception of free speech appears to have departed from this model. Tracing the western history of free speech since the French Revolution, Elizabeth Powers argues that the debates foreshadowed in the eighteenth century have intensified since the fall of the Berlin Wall in 1989.[9] One reason for this, she argues, is that the 'western liberal order would seem to have become a victim of its own success, so long without competition that it has forgotten the source of the freedoms it enjoys'.[10] As a result, arguments for free speech in western societies have tended to be expressed in stronger terms in recent years, giving rise to the strong model of free speech on which this chapter focuses.

In his 2016 book *Free Speech: Ten Principles for a Connected World*, the journalist and academic Timothy Garton Ash argues that 'free speech has never meant unlimited speech – everyone spouting whatever comes into his or her head, global logorrhea'.[11] Despite this, the ten principles that Ash proceeds to outline adopt a relatively laissez-faire approach to free speech. The second principle states that 'we neither make threats of violence nor accept violent intimidation', but under the fifth heading Ash argues that 'mature democracies should move beyond hate speech laws'. Rather than simply abandoning laws, Ash implies the need to replace them with more robust social mechanisms for dealing with hate speech. However, the examples of online comment sections show that communities of users are often peculiarly hesitant or reticent to deal with hate speech, due to the belief that calling it out will only escalate the conflict and exacerbate the problem. As a result, it is unclear how the

8 European Convention on Human Rights, Art 10.

9 Powers, *Freedom of Speech*, p. 10.

10 Ibid.

11 Ash, *Free Speech,* p. 4.

racist, sexist, and body-shaming abuse described by Diane Abbott could be regulated in Ash's system, particularly where it does not constitute a direct 'threat of violence'. Its vitriolic force appears to nestle safely in the realms of opinion. In a study of racist hate speech, Caroline West notes that '[s]ome liberals have argued that the regulation of hate speech should be resisted as a matter of principle because our commitment to free speech must be absolute'.[12] Yet, as West observes, this model of unregulated free speech is not sustainable, because the 'visceral hostility' of racist hate speech 'forecloses' the possibility of further discussion: 'In the immediate aftermath of a verbal attack, it is rare that victims are able to produce words at all, let alone to gather themselves together to offer a clear-headed and balanced response.'[13] As a result, the strong model of free speech struggles to deal with the problem that one person's unlimited free speech can effectively silence another's.

Despite this, some netizens have defended their 'right to free speech' in absolutist terms. When Katie Hopkins was fired by the LBC radio station for tweeting that 'We need a final solution' after the Manchester terror attack on 22 May 2017, the libertarian commentator Brendan O'Neill defended her in a piece entitled 'The Mob Claims Another Scalp'. Despite suggesting that Hopkins's tweet knowingly echoed one of the most harrowing slogans of the Holocaust, O'Neill argues that:

> [I]t's one thing to be offended by something (you can be offended by whatever you like) – it's quite another to mobilise your feelings of offence to the end of getting someone sacked, and by extension warning everyone else in public life that if they say anything like this, if they venture too far from the realm of Acceptable Thought, then they too will face fury, punishment, and potentially the loss of their livelihood.[14]

This argument implies that those who are 'offended' by what O'Neill himself suggests was a call for genocide are exercising the privilege of oversensitivity, while Hopkins's freedom to issue this call stands as an indefatigable right. One might argue that this comment embodies an extreme position in the free speech debate. However, the absolutist approach to free speech has rapidly entered the mainstream.

12 West, 'Words That Silence?', p. 246.
13 Ibid., pp. 235-236.
14 O'Neill, 'Katie Hopkins' Sacking'.

'Broadcast yourself': The expanding public sphere

In addition to the strong model of free speech outlined above, online platforms have been influential in dissolving the traditional boundary between the public and private spheres. With the rise of social media, the distinction between what one would say and do in private and in public has become more porous. The benefits of this development are expressed in the mission statements of various prominent social media platforms, for example YouTube ('Our mission is to give everyone a voice and show them the world'),[15] Twitter ('Our mission: Give everyone the power to create and share ideas and information instantly, without barriers'),[16] and Facebook ('Facebook's mission is to give people the power to build community and bring the world closer together').[17] Social media platforms like these have proved important in democratizing the access to information and the right to self-expression around the world. However, the expansion of the public sphere into the realms of formerly private opinion has also given rise to an increased potential for the clash of beliefs and the uncensored expression of prejudices online. This section examines how the expanded capacity for instantaneous communication inevitably increases the potential for conflict, as well as how prominent online platforms have attempted to the regulate the resulting phenomenon of online vitriol.

In arguing for the transformative power of social expression, social media platforms rely on a model of free speech as a civilizing force which emerged during the Enlightenment. Since the eighteenth century, it has often been suggested that speech should be allowed to function as a free market. In this vein, the eighteenth-century philosopher the Earl of Shaftesbury argued that:

> All politeness is owing to liberty. We polish one another and rub off our corners and rough sides by a sort of amicable collision. To restrain this is inevitably to bring a rust upon men's understandings. It is a destroying of civility, good breeding and even charity itself, under the pretence of maintaining it.[18]

Shaftesbury's metaphor of 'amicable collision' implies that social values can only be developed by individuals participating in the conversation of

15 YouTube, 'About'.
16 Twitter, 'About'.
17 Facebook, 'Investor Relations'.
18 Cooper, *Characteristics of Men,* p. 31.

culture. In other words, there is a need for individuals to air and discuss their private opinions in public, so that the members of a community can arrive at a consensus through mutual critique and discussion. This model of amicable collision continues to be upheld in a modified form by social media platforms, for example Twitter, which reminds its users that 'as a policy, we do not mediate content or intervene in disputes between users.'[19] Along similar lines to Shaftesbury, the current Twitter policy treats conflict as a potentially productive process in which users regulate one another's posts and arrive at a consensus about what can or cannot be expressed.

However, the hugely increased capability which these platforms offer their users – the opportunity to 'create and share ideas and information instantly, without barriers'[20] – has also led to an increased capacity for conflict. Hate speech like that aimed at Diane Abbott, including rape threats, death threats, and the Twitter hashtag 'burn Diane Abbott',[21] shows that the deregulation of speech on the internet does not necessarily lead to 'politeness' and 'amicable collision'. In this respect, it is important to consider the fundamental shift that has occurred with the rise of the internet in the equilibrium between the private and public spheres. In his seminal text *The Structural Transformation of the Public Sphere*, Jürgen Habermas argued that the 'model of the bourgeois public sphere presupposed strict separation of the public from the private realm'.[22] Though Habermas recognized that the boundary between these spheres was porous, the rise of social media has almost erased this boundary by encouraging individuals to share their private thoughts and emotions in the public sphere with virtually no mediation. Users of these platforms are invited to project every aspect of their mental experience into the public debate. This approach is illustrated by YouTube's slogan – 'Broadcast Yourself' – which imagines the possibility of superimposing every individual's stream of consciousness in the public sphere. One consequence of this is that public abuse is more likely to occur in this context than in Shaftesbury's model of 'civility', which was developed in a culture where social groups were much smaller and even access to print media was limited.

Like any major cultural change, the rebalancing of the private and public spheres carries both advantages and disadvantages, which will take years if not decades to understand. Writing just before the rise of Twitter, Facebook,

19 Twitter, 'About Offensive Content'.

20 Twitter, 'About'.

21 Abbott, 'I Fought Racism and Misogyny to Become an MP'.

22 Habermas, *The Structural Transformation*, pp. 175-176.

and YouTube, Luke Goode contended that the internet could contribute to greater reflexivity in the public sphere:

> [I]n the Habermasian model, the public sphere and its reflexive context must be mutually reinforcing: the public sphere takes on the role of a kind of exemplary space for the considered, deliberative and, as far as possible, egalitarian weighing of competing claims, an ethic that can at least rub off on – though by no means colonise – the more unruly and visceral micro-practices and discourses of everyday life.[23]

To some extent, this prediction has come true, with social media platforms allowing their users to criticize celebrities and those in the public eye for displaying prejudice. This occurred, for example, when the musician Ten Walls posted homophobic comments on his Facebook page in 2015 and was rapidly criticized by fans and dropped by sponsors.[24] Subsequently, a number of other artists, including UK grime artist Stormzy, have apologized for homophobic posts and tweets made in the past.[25] While this may hold true for public figures, the proliferation of online content has also made it possible for private individuals, such as the owner of the English Defence League-affiliated hate account aimed at Diane Abbott, to project what Goode calls 'the more unruly and visceral micro-practices and discourses of everyday life' into the public sphere without censorship. While social media platforms can make the public sphere more reflexive of democratic ideals, they also risk transforming areas of the public sphere into an unregulated space where unjustified prejudice and legitimate, reasoned opinion become interchangeable. To this extent, the traditional model of the public sphere as a space in which communities come together to negotiate that shared cultural practices and public opinions becomes less tenable online.

Unruly spaces: The problems of enforcement

The phenomenon of online vitriol is arguably exacerbated by the sheer scale of online content, which necessitates the relatively non-interventionist and reactive approach to moderation adopted by many online platforms. On its support page, YouTube prohibits 'content that promotes violence or hatred

23 Goode, *Jürgen Habermas,* p. 120.
24 Channel 4, 'Ten Walls Dropped'.
25 BBC Newsbeat, 'Stormzy Apologises for Homophobic Tweets'.

against individuals or groups based on certain attributes, such as: race or ethnic origin, religion, disability, gender, age, veteran status, and sexual orientation/gender identity'.[26] However, the site also instructs its users to 'keep in mind that not everything that's mean or insulting is hate speech', adding: 'If you're upset by content that a specific person is posting, you may wish to consider blocking the user.'[27] Arguably, it is not feasible for platforms like YouTube, Facebook, and Twitter to moderate comments in real-time. However, the reliance on individual blocking and reporting mechanisms once again places the burden on the victims of abuse and risks denigrating their reactions as hypersensitive. Not only does this mean that many instances of online vitriol go unreported, but it also presents the victims of online abuse with the task of deciding whether the comments they have received are severe enough to qualify as 'abuse', 'harassment', or 'bullying'.

Even when the identification of hate speech is unequivocal, there remains a problem of enforcement. On a video of a subway performer covering Fleetwood Mac's song 'Landslide', one person commented: 'As soon as she started singing I got chills, damn'.[28] Another user, responded: 'Write a ***** book about it. You ******.' At this point, the conversation escalated until a user with a pseudonymous and deliberately offensive name made a comment about slavery. When a fellow commenter expressed outrage at this post, they were reprimanded by a third user, who commented: 'Why people try and argue with trolls on this I'll never understand. They just want everyone to be as mad and ignorant as they are! Save your energy bud ;)'.[29]

This exchange exemplifies the problem referred to above, where forms online vitriol and actual hate speech are mislabelled as 'trolling', with the result that users agree to ignore rather than denounce them. The sentiment embodied in the phrase 'don't feed the trolls' is now widespread on social media platforms. The problem with this approach is that even if the other users do not sanction such racist and dehumanizing language, they are pressured (as in the example above) into overlooking it. Meanwhile, the perpetrators feel empowered to post sentiments online which they often would not expect to be able to express in person. Furthermore, even if such comments are removed and the users' accounts are suspended, the individual concerned can simply create another pseudonymous account and continue posting.

26 YouTube Help, 'Hate Speech'.

27 Ibid.

28 https://www.youtube.com/watch?v=x--yddOolRQ. Accessed 23 June 2017.

29 https://www.youtube.com/watch?v=x--yddOolRQ. Accessed 11 March 2018.

Arguably, there are limits to the steps that platforms like Facebook, YouTube, and Twitter can take to tackle the epidemic of online vitriol. Following the recent media scrutiny of cyberbullying, these platforms have taken steps to make their stances on online abuse and harassment more robust.[30] Nonetheless, there is still a dearth of concrete solutions to tackle the scale of the problem. In a recent study of online misogyny, Emma A. Jane observes that 'cyber-harassment such as rape threats and sexualized vitriol [...] have become part of the everyday experience for many women online'.[31] Not only this, but 'the discourse involved is more rhetorically noxious and is occurring in far broader communities than earlier iterations of gender-based harassment'.[32] The current case-by-case approach to regulation cannot help but leave many instances of online vitriol unreported. In addition, it can appear punitive when some individuals are policed more strictly than others. An example of this came in 2017, when Rose McGowan was suspended from Twitter for remarks she made in the wake of the Harvey Weinstein abuse scandal. Believing that McGowan had been unfairly targeted in a draconian instance of enforcement, many women boycotted the platform for a day using the hashtag '#WomenBoycottTwitter'.[33] Though potentially effective, this form of protest also involved women removing themselves from the social platform and therefore renouncing this channel of self-expression. Given the limitations of regulation and enforcement when dealing with online vitriol, there is a strong case for analyzing the conditions which enable it to prevail in the first place.

As discussed in the previous section, social media platforms encourage the expansion of the public sphere into a public hypersphere, in which individuals are invited to share their immediate, emotional responses to every event. This, coupled with the sites' non-interventionist approach to free speech, has enabled the development of a unique comment culture focussed on combat, disinhibition, and the contest for popularity. In the process, the conventional rules of conversation, argumentation, and mutual respect have been disapplied. Between them, these factors make the public hypersphere a revolutionary but vitriolic space, which has far-reaching consequences for language, logic, and the constitution of societies. In the absence of active enforcement by users, online platforms can be governed by the 'unruly and visceral micro-practices' that Goode hoped would be

30 Twitter, 'Clarifying the Twitter Rules'.
31 Jane, 'Online Misogyny and Feminist Digilantism'.
32 Ibid.
33 Griffin, 'Boycott Twitter'.

transformed in a more reflexive public sphere.[34] Unless users are willing to enforce the social rules of debate and politeness as they would in an offline space, online platforms develop their own 'thread logic', governed not only by attempts to shock and troll other users, but also in some cases by very real attempts to threaten and intimidate.

Limitless free speech?

The alternative to ignoring online vitriol and hate speech is for users to call it out by denouncing the content and/or blaming those who post such comments. However, this raises the problem that those who condemn online vitriol risk being accused of being vitriolic themselves, or of attempting to stifle free speech. In this case, perpetrators of hate speech can quickly be transformed to be presented as victims. On 17 April 2015, Katie Hopkins published an article in *The Sun* in which she referred to migrants as 'cockroaches' and proposed using gunships to prevent them from reaching British shores. At the time, there was widespread condemnation of this article on Twitter and other social media platforms, but the press regulator Ipso found on 1 May 2015 that Hopkins's comments were not discriminatory because they did not refer to a specific individual. This finding arguably reveals the shortfalls of relying on national defamation and libel laws to shoulder the burden of regulating free speech. Clearly, hate speech can be targeted at groups as well as individuals, and it appears significant that the British press regulator in 2015 was not equipped to deal with that fact. Equally significant was the willingness of numerous individuals to defend Hopkins's comments in the name of free speech.

In a blog published by *The Spectator* on 20 April 2015, Brendan O'Neill argued that 'she's wrong, but Katie Hopkins has a right to call migrants 'cockroaches''.[35] The problem of the strong model of free speech, based on the fallacy that one person's freedom of expression cannot harm another's, has already been considered. However, O'Neill raised a second point when he described the 'Twitterstorm' which followed Hopkins's remarks:

> She's a fascist, they said. She's a Nazi. She's indistinguishable from the authors of the Rwandan genocide. Her comments would have made Hitler blush, said an *Independent* journalist. Congratulations! You win the war

34 Goode, *Jürgen Habermas*, p.120.
35 O'Neill, 'She's Wrong'.

> of hyperbole, the thesaurus-bombing competition to see who can hate Hopkins the most.[36]

What is significant is O'Neill's recognition that hate speech is often countered with hate on the internet. This raises the question of how to distinguish hate from denunciation: whereas denunciation involves a reasoned rejection of hateful comments, hate involves a more aggressive and/or abusive response to the poster as an individual. Increasingly, those interacting in the public sphere respond to online vitriol in its own vituperative terms, with scorn and vitriol. As O'Neill observes, this leads to a 'war of hyperbole', as the internecine internet hosts a perpetual escalation of anger and resentment.

The contemporary manifestation of anger as a form of critique has a long heritage. No one has done more to understand this heritage than the philosopher Hannah Arendt. In her essay *On Revolution*, Arendt traced the progress of political violence from Rome to her own time, arguing that:

> since the days of the French Revolution, it has been the boundlessness of their sentiments that made revolutionaries so curiously insensitive to reality in general and to the reality of persons in particular, whom they felt no compunctions in sacrificing to their 'principles,' or to the course of history, or to the cause of revolution as such.[37]

Though the reaction to Katie Hopkins's comments on Twitter did not occur in a revolutionary context, it is significant that Hopkins's critics continue to pride themselves on the 'boundlessness of their sentiments'. Due to their vehemence, writers like O'Neill are ultimately able to portray Hopkins as a victim of the debacle, since so many terms of abuse have been levelled at her for exercizing what he and others perceive as her absolute right to free speech. In the process, some of the accusers have become like their target in adopting the same language and logic. In the context of the May 1968 events in France, Arendt argued in her essay *On Violence* that 'loss of power becomes a temptation to substitute violence for power'.[38] Arguably, one reason that online posters increasingly adopt anger and vitriol as a means of critique is that they no longer possess the power to set the terms of the debate. As the principles of rationality and compassion lose their hold over the public hypersphere, vitriol begins to seem like the only means

36 Ibid.
37 Arendt, *On Revolution*, p. 85.
38 Arendt, *On Violence*, p. 54.

of expression. Arendt argues that: 'Where violence rules absolutely [...] not only the laws [...] but everything and everybody must fall silent.'[39] In adopting anger as a mode of critique, and joining the 'war of hyperbole', those who wish to defend against online vitriol and denounce hate speech risk abandoning the principles of reason and compassion that they seek to uphold. No doubt, there are highly articulate and compassionate forms of anger, but it is important for opponents of social injustice to retain these, rather than surrendering to an expressive but incoherent vitriol.

Conclusion

This brief survey of 'comment culture' on YouTube, Twitter, and Facebook, as well as in the British tabloid press, has examined how the phenomenon of online vitriol has developed and, to some extent, been regulated on social media. In particular, it suggests that the development of a strong model of free speech, coupled with the expansion of the public sphere into the realm of private and opinion and belief, helps to shape the cultures that evolve in the comment sections of various online platforms. Although each platform has its own unique culture, three broad trends can be seen to emerge. First, social media have developed a unique 'comment culture', in which largely self-regulating communities of users reach a consensus about what forms of expression are or are not available. Second, online platforms are governed by their own distinctive 'thread logic', in which traditional models of debate and conversation are undermined by the phenomenon of trolling and by the confusion of this with more problematic forms of online vitriol and hate speech. Third, the conventional rules of conversation and human interaction are often modified online, with some users being more willing to resort to insult and abuse when they have the ability to appear anonymously and shelter behind other users' unwillingness to 'feed the trolls'.

Ultimately, this chapter considers how the phenomenon of online vitriol begins to influence offline behaviour, as the displays of anger and aggression commonly tolerated and accepted online begin to permeate public life. While it remains difficult for the providers of social media platforms to police every post, the onus of moderation falls on the users of these platforms, who are able to negate and report extreme instances of abuse and prejudice online. In the absence of a consensus and a concerted democratic effort to uphold the principles of debate and mutual respect, the comments sections on even

39 Arendt, *On Revolution*, p. 9.

the most highly-respected online platforms risk becoming unruly spaces, where the public sphere – rather than becoming a democratic market of opinion – is colonized by 'the more unruly and visceral micro-practices and discourses of everyday life'.[40] The problem for users and operators of these sites is developing a response to online vitriol and hate speech which finds a middle ground between simply ignoring these posts and responding to them in similarly vitriolic terms. A more robust culture of denunciation needs to develop, in which users respond to vitriol with a reasoned rejection of its content rather than with abuse and outrage. Without this, it does not seem possible to realize Ash's ideal that 'mature democracies should move beyond hate speech laws'.

Works cited

Abbott, Diane. 'I Fought Racism and Misogyny to Become an MP. The Fight Is Getting Harder'. *The Guardian,* 14 February 2017, https://www.theguardian.com/commentisfree/2017/feb/14/racism-misogyny-politics-online-abuse-minorities. Accessed 23 June 2017.

Arendt, Hannah. *On Revolution*. London: Faber & Faber, 2016.

—, *On Violence*. New York: Harcourt Brace, 1970.

Ash, Timothy Garton. *Free Speech: Ten Principles for a Connected World.* New Haven: Yale University Press, 2016.

BBC Newsbeat. 'Stormzy Apologises for Homophobic Tweets'. *BBC*, 22 November 2017, http://www.bbc.co.uk/newsbeat/article/42078366/stormzy-apologises-for-homophobic-tweets. Accessed 11 March 2018.

Channel 4. 'Ten Walls Dropped from Festivals after Anti-gay Rant'. *Channel 4*, 8 June 2015, https://www.channel4.com/news/festivals-drop-ten-walls-following-vicious-homophobic-rant. Accessed 11 March 2018.

Cooper, Anthony Ashley, Third Earl of Shaftesbury. *Characteristics of Men, Manners, Opinions, Times*, edited by Lawrence E. Klein. Cambridge (UK): Cambridge University Press, 1999.

European Convention on Human Rights, Art. 10. https://www.echr.coe.int/Documents/Convention_ENG.pdf. Accessed 11 March 2018.

Facebook, 'Investor Relations', https://investor.fb.com/resources/default.aspx. Accessed 11 March 2018.

Goode, Luke. *Jürgen Habermas: Democracy and the Public Sphere.* London: Pluto Press, 2005.

40 Goode, *Jürgen Habermas*, p. 120.

Griffin, Andrew. 'Boycott Twitter: Women Across the World Abandon Site after it Bans Rose McGowan'. *The Independent*, 13 October 2017, https://www.independent.co.uk/life-style/gadgets-and-tech/news/boycott-twitter-latest-explained-rose-mcgowan-harvey-weinstein-ben-affleck-ban-a7998341.html. Accessed 11 March 2018.

Habermas, Jürgen. *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*, transl. by Thomas Burger. Cambridge (MA): MIT Press, 1991.

Jane, Emma A. 'Online Misogyny and Feminist Digilantism'. *Continuum: Journal of Media & Cultural Studies*, vol. 30, no. 3, 2016, pp. 284-297.

Khommani, Nadia, ''Day of Rage' Protestors Demand May's Resignation'. *The Guardian*, 21 June 2017, https://www.theguardian.com/uk-news/2017/jun/21/day-of-rage-protest-march-westminster-grenfell-tower. Accessed 23 June 2017.

Locke, John. 'A Letter Concerning Toleration', *Locke on Toleration*, edited by Richard Vernon. Cambridge (UK): Cambridge University Press, 2010.

Mikal, Jude P., et al. 'Common Voice: Analysis of Behavior Modification and Content Convergence in a Popular Online Community'. *Computers in Human Behavior*, vol. 35, no. C, 2014, pp. 506-515.

Milton, John. *Areopagitica*. London, 1644, http://www.dartmouth.edu/~milton/reading_room/areopagitica/text.html. Accessed 23 June 2017.

O'Neill, Brendan. ''Katie Hopkins' Sacking: The Mob Claims Another Scalp'. *Spiked*, 26 May 2017, https://www.spiked-online.com/2017/05/26/katie-hopkins-sacking-the-mob-claims-another-scalp/. Accessed 23 June 2017.

—, 'She's Wrong, But Katie Hopkins Has a Right to Call Migrants 'Cockroaches.'' *The Spectator*, 20 April 2015, https://www.spectator.co.uk/article/she-s-wrong-but-katie-hopkins-has-a-right-to-call-migrants-cockroaches-. Accessed 23 June 2017.

Phillips, Whitney. *This Is Why We Can't Have Nice Things: Mapping the Relationship Between Online Trolling and Mainstream Culture*. Cambridge (MA): MIT Press, 2015.

Powers, Elizabeth. *Freedom of Speech: The History of an Idea*. Lewisburg (PA): Bucknell University Press, 2011.

Twitter, 'About', https://about.twitter.com/en_us/company.html. Accessed 11 March 2018.

—. 'About Offensive Content', https://help.twitter.com/en/safety-and-security/offensive-tweets-and-content. Accessed 11 March 2018.

—. 'Clarifying the Twitter Rules', 3 November 2017, https://blog.twitter.com/official/en_us/topics/company/2017/Clarifying_The_Twitter_Rules.html. Accessed 11 March 2018.

West, Caroline. 'Words That Silence? Freedom of Expression and Racist Hate Speech' in *Speech and Harm: Controversies over Free Speech*, edited by Ishani Maitra and Mary Kate McGowan. Oxford: Oxford University Press, 2012, pp. 222-247.

YouTube. 'About', https://www.youtube.com/about/. Accessed 11 March 2018.
YouTube Help. 'Hate Speech', https://support.google.com/youtube/answer/2801939. Accessed 23 June 2017.

About the author

Tom Clucas completed his B.A., M.St. and D.Phil. in English Literature at the University of Oxford, where he also taught at several of the Colleges. Subsequently, he worked as a researcher at the International Graduate Centre for the Study of Culture and Deputy Professor of English and American Literature and Culture at Justus Liebig University, Giessen. He has published articles on Romantic Poetry in journals including *Notes and Queries*, *The Wordsworth Circle*, *The Cowper and Newton Journal*, *Textual Cultures*, *The Review of English Studies*, *European Romantic Review and The Coleridge Bulletin*. He now lives and works in London.