

issue 23: General Issue 2014



FCJ-170 Challenging Hate Speech With Facebook Flarf: The Role of User Practices in Regulating Hate Speech on Facebook.

Benjamin Abraham
University of Western Sydney

Abstract:

This article makes a case study of ‘flarfing’ (a creative Facebook user practice with roots in found-text poetry) in order to contribute to an understanding of the potentials and limitations facing users of online social networking sites who wish to address the issue of online hate speech. The practice of ‘flarfing’ involves users posting ‘blue text’ hyperlinked Facebook page names into status updates and comment threads. Facebook flarf sends a visible, though often non-literal, message to offenders and onlookers about what kinds of speech the responding activist(s) find (un)acceptable in online discussion, belonging to a category of agonistic online activism that repurposes the tools of internet trolling for activist ends. I argue this practice represents users attempting to ‘take responsibility’ for the culture of online spaces they inhabit, promoting intolerance to hate speech online. Careful consideration of the limits of flarf’s efficacy within Facebook’s specific regulatory environment shows the extent to which this practice and similar responses to online hate speech are constrained by the platforms on which they exist.

Introduction

A recent spate of high profile cases of online abuse has raised awareness of the amount, volume and regularity of abuse and hate speech that women and minorities routinely attract online. These range from the responses garnered by Anita Sarkeesian's (2012; 2014) video series 'Tropes vs. Women in Video Games,' which included photoshopped images of Sarkeesian made to appear bruised and brutalised, (Lewis, 2012) to the abuse directed at the activist Caroline Criado-Perez following her campaign to have a woman represented on a UK banknote (Guardian Staff, 2014), to the countless instances of more pernicious 'Everyday Sexism' documented by the activist group of the same name.

As a result of these and a host of similar recent events, it has become increasingly apparent that online abuse and instances of hate speech directed at women, people of colour, transgender individuals and other minorities is on the rise in many online spaces. This awareness is so prevalent that Robinson Meyer (2014) writing for *The Atlantic* in late October 2014 described the GamerGate movement (an online hate mob responding to the increasing visibility of women in the games industry and in gaming culture, and a kind of culmination of these larger trends) as 'an existential crisis for Twitter.' This is particularly the case given that Twitter has become an important public space for users in not just the videogame industry but also similar 'cultural communities that have strong online presences, including tech, science fiction, fantasy, comics, and film' (Meyer, 2014). The mental and emotional burdens borne by the women and minorities that use these platforms are clearly disproportionate, as they expose themselves to almost limitless online abuse. While many users clearly feel that this situation is unacceptable and can no longer be ignored, the question remains, what can practicably be done about the problem of harassment and hate speech on social networking sites like Facebook and Twitter?

In part, this problem is constituted by the approach taken by many social media sites, as they deliberately position their services in a very particular relationship with their users in order to limit the responsibilities they have towards them. Tarleton Gillespie (2010) has productively focused on the terminology of the 'platform' itself, frequently deployed by sites like Twitter and Facebook, arguing that it does important discursive work positioning these services and managing expectations across users, potential legislative regulators, and advertisers – eliding and diffusing the tensions that might otherwise arise from their competing interests. As Gillespie (2010: 11) notes:

... online content providers who do not produce their own information have long sought to enjoy limited liability for that information, especially as the liabilities in question have expanded from sordid activities like child pornography and insider trading to the much more widespread activities of music and movie piracy.

Missing from this list of indemnities that platform holders regularly seek for themselves is the problem of hate speech generated on and transmitted through social media platforms. Many platforms simply do not want the burden (whether economic or legal) entailed by responsibility for dealing with hate speech, and their arguments against the imposition of further regulation often rest upon claims as to the ‘impossible’ scale of the problem. In a similar fashion, intervention by state authorities often faces a host of issues preventing or hampering effective regulation of online hate speech. These include cultural and legislative reluctance, as well as technical or practical difficulties facing enforcement and regulation of the internet, which many governments and state agencies are not well placed to handle. As a result of these issues, users of online services are often left to take responsibility for hate speech themselves. With this paper I explore one case of users acting creatively within Facebook’s technical and regulatory environment to take small-scale actions against hate speech.

I begin by briefly considering the literature addressing questions of hate speech and the varying degrees of comfort different national traditions have with greater or lesser state intervention in speech. This provides important context for the later discussion of the practice of flarfing. Examining the literature around online hate speech reveals a complex negotiation between states, internet intermediaries, and the agency afforded to users themselves, with this paper taking a particular focus on the regulatory environment presented by Facebook. The remainder of this paper discusses a little-known creative activist strategy employed on Facebook which redeploys the site’s tagging algorithm to ‘flarf’ – posting absurdist, nonsense, or subtly reflexive messages to Facebook threads and status updates, often those containing bigotry, sexism or hate speech. I examine this practice in the context of the ongoing debate around internet hate, situating the case of Facebook flarf as a creative, discursive activist practice deployed by individuals and small communities as a user-led response to harmful speech online. I argue that it demonstrates the narrow utility of individual citizens’ creative small-scale interventions in online discourse, acknowledging that it is importantly limited in its ability to address the scope of the problem sufficiently, existing as it does within the limitations of the regulatory spaces both Facebook itself and larger international contexts define.

This paper adds to the evidence for what Banks (2010: 234) calls the need for ‘a broad coalition of citizens, government and businesses’ in addressing hate speech online, and provides a clearer picture of what kinds of creative responses may be and have been available to individual users in the interim. Users can be instrumental in *taking responsibility* for the promotion of online cultures intolerant of bigotry and hate speech on social media platforms, as Banks (2010: 238) suggests, but it is important to also acknowledge the limitations and opportunities created by the context of these platforms own regulatory responses.

The research undertaken for this paper was performed by observing Facebook user activities online over a period of years, though the majority of examples are drawn from the period which saw the most flarf activity, during the year 2012. This approach necessitated by design a reliance on personal contacts drawn from the author’s network of activist acquaintances, as well as some observation of public posts made by other users unknown to the author. The relatively small scope of the study reflects the transient and contingent nature of these kinds of online user practices, existing within and frequently in response to the changing and contested regulatory contexts of the platforms and services themselves. The precise number of users involved in the practice is incredibly difficult to determine, given the nature of the Facebook platform and the fairly organic manner in which the practice grew from a larger community context. It is safe to say, however, that the practice was not widespread, and was very much a product of a particular online community at a certain time and place.

Despite the limited extent of its application, Facebook flarf remains important for our understanding of the development of these sites and the cultures that emerge alongside them, as users test the extent of their personal agency and their ability to ‘take responsibility’ for the problem of hate speech online. The examples included in this paper were chosen for their ability to demonstrate the key features of Facebook flarfing, from a very limited set of materials collected at the time, or just after, these incidents occurred. The ‘real time’ nature of much of the Facebook platform and the algorithmic selection of material it chooses to present in the news feed make the discovery and documentation of examples that illuminate these practices particularly difficult. All examples are drawn from what were at the time of writing publicly visible ‘posts’ and ‘pages’ on Facebook.

Responses to Internet Hate

When examining how the issue of harmful speech online has been addressed, an obvious division appears, based quite clearly on the differing historical and political traditions of the United States and Europe. The former's history and culture of resolutely defending free speech under the protections of the first amendment causes some friction with the latter's greater comfort with state-based intervention in the prohibition of hate speech. This comfort can be largely attributed to Europe's recent history and experiences with hate speech in the lead-up to the Second World War. It reflects an awareness of its role in enabling the demonisation of minority groups, ranging from Jews and Roma to homosexuals and the disabled, acts that prefaced the Holocaust.

Awareness of the unique problems presented by internet hate speech dates back at least to the early BBS (Bulletin Board System) era. Chip Berlet (2000) has described the early history of US based hate sites and the range of responses, from individual hackers to US Government efforts and the work of prominent civil liberties groups who have extended their concern for freedom of speech protections in the public sphere to also encompass speech on the internet. Expanding on this history, Frydman and Rorive (2002) have examined the involvement of "intermediaries" such as ISPs in preventing or removing hate speech online. While acknowledging that European legal frameworks are likely to give 'public authorities and human rights activists... better tools to limit the influence of racist, Nazis, anti-Semitic and other kind of hate speeches on the Internet,' they caution that it could be a "slippery slope" to new regimes of censorship (Frydman and Rorive, 2002: 55). Intermediaries also include websites and web services like Facebook and Twitter, and I return to discuss these sites and their typical reluctance to intervene in a moment.

The concern that too much state intervention in these services may be a "slippery slope" to regimes of censorship appears most strongly and is repeatedly emphasised in American scholarship on the issue, such as Barnett's (2007) *Untangling the Web of Hate*, which examines the US Constitution's first amendment protections as they apply to hate speech online. Through a content analysis of the material hosted on hate sites such as Ku Klux Klan and Neo-Nazi websites, Barnett applies the US Supreme Court's jurisprudential tests of what constitutes *unprotected speech* according to the first amendment and finds that the vast majority of material hosted on these US based hate sites would be considered protected expressions. Similarly, Foxman and Wolf's (2013) *Viral Hate*, produced with the support of the Anti-Defamation League (a political lobby group founded primarily to counter anti-Semitism), articulates this distinctly American position on hate speech with greater nuance than some first amendment advocates who are reluctant to view any restrictions

on speech as acceptable. In Foxman and Wolf's (2013: 60) view, hate speech 'laws are the least effective way to deal with the problem.' Instead, they argue:

... the best antidote to hate speech is counter speech – exposing hate speech for its deceitful and false content, setting the record straight, and promoting the values of respect and diversity. (Foxman and Wolf, 2013: 129)

Though a noble goal, it is clearly unable to account for the unequal burden that it places upon individuals most harmed by hate speech. Women and minorities are, in effect, caught between a lack of state regulatory intervention in online hate speech, and the reticence of internet services such as Facebook and Twitter which, as Gillespie (2010: 12) remarks, deploy the rhetoric of the platform in order to position themselves as a simple 'facilitator that does not pick favourites.' Gillespie (2010: 11) unpacks this opting-out of intervention, explaining that:

... in the effort to limit their liability not only from...legal charges [arising from users infringing copyright] but also more broadly the cultural charges of being puerile, frivolous, debased, etc., intermediaries like YouTube need to position themselves as just hosting – empowering all by choosing none.

The implication of this state of affairs is that those harassed are now left to address and take responsibility for the conditions of their own harassment via 'counter speech' – if we take Foxman and Wolf (2013) at their word. Challenging their fairly simplistic conception of efficacious 'counter speech' is a body of work from feminists and other theorists who criticise the conception of the neutral liberal state and the 'marketplace of ideas' assumptions upon which these types of claim rest.

Representative of the more European perspective, Abigail Levin (2010: 1) argues that the idea that the state should remain a neutral facilitator of a 'marketplace of ideas' which assumes hate speech is defeated by truthful, efficacious counter speech (Foxman and Wolf, 2013) is incompatible with another commitment of the liberal state – that of the equality of citizens, which is often sacrificed in service of non-intervention in the expression of ideas. Most crucially for its theoretical legitimacy, Levin (2010) argues the hands-off neutral state does not lead *naturally* to better (or more truthful) ideas winning out via competitive market forces, since among other reasons:

... our systemically racist, sexist, and homophobic society has had the effect that certain dominant racist, sexist, and homophobic views have become so deeply held as not to be amenable to rational discussion, with the effect that minorities' and women's voices are not heard fairly in the marketplace. (Levin, 2010: 1)

This sentiment echoes a body of literature pointing towards the amount of work required to produce free markets themselves, for instance Karl Polanyi's *The Great Transformation* (2001 [1944]) and Nikolas Rose's *Powers of Freedom* (2004 [1999]: 65), along with a host of others who have criticised the 'neutral marketplace of ideas' on a variety of grounds (see: Brazeal, 2011; Goldman and Cox, 1996; Sparrow and Goodin, 2001). Levin's (2010: 4) conclusion, drawn with regard to the current skewed marketplace, is that 'the state as a neutral facilitator of private ideas is untenable and must be dropped' and an interventionist liberal state conceptualised. Such ideas apply just as much to social media sites' reluctance to intervene in their users' generation of content.

Addressing concerns that immediately follow any proposed greater state interventionism and the automatic cries of 'censorship' that ensue when the prospect of state intervention into speech and truth claims are raised, both Judith Butler (1994) and Frederick Schauer (1994) have offered important critiques of uncritical understandings of censorship as simply 'preference frustration' without consideration of the impact of state power and discourses on the formation of these preferences themselves. In other words, the neutral marketplace of ideas is not, and cannot ever be, perfectly and entirely neutral.

Banks (2010: 234) construes enforcement as the main difficulty for contemporary European and other nations' interventionist approaches to hate speech:

[The] rise in hate speech online is compounded by difficulties in policing such activities which sees the Internet remain largely unregulated. Criminal justice agencies are unlikely to proactively dedicate time and money to investigate offences that are not a significant public priority. Consequently, the police will rarely respond to online hate speech unless a specific crime is reported.

This reluctance to intervene is repeatedly encountered when individuals seek intervention by state authorities such as US police departments. For example, in early 2014 a high-profile piece discussing online hate speech by journalist Amanda Hess (2014) detailed her own experiences with online threats to her person, the mental and emotional cost of

the permissiveness of online hate speech directed at women in the United States, and US law enforcement's prohibitive jurisdictional limitations and frequent reluctance to investigate the majority of these incidents. This seemed to set the pattern for the year, as in October of 2014, feminist media critic Anita Sarkeesian made headlines when Utah State University received a highly credible and detailed anonymous threat to carry out a mass shooting if Sarkeesian refused to cancel a speaking event (Wingfield, 2014). Because of Utah's concealed carry law, the police in that state would not (or could not) prevent attendees from carrying weapons into the meeting, and the event was subsequently cancelled. Examples like these underscore the difficulties facing effective state regulation, given the practical reality that authorities in various nations (none less than the United States, where many internet sites and social networking sites are based) cannot be relied upon to do the work of policing and preventing hate speech, particularly online.

Given the issues state authorities and legislators face, one might be tempted to hope that internet intermediaries such as social networking sites might take the initiative to combat the issue of online hate speech themselves. However, barring one significant victory which I discuss later in this paper, many sites (Facebook included) have been reluctant to take a more proactive role in preventing or responding to hate speech on their services. The reasoning for this is partly due to a lack of legislative compulsion – as we have already seen, the United States is reluctant to legislate against hate speech – and partly because sites like Facebook must negotiate the competing interests of users, advertisers, and government legislatures (Gillespie, 2010: 7). In order for these sites to function, and remain profitable, they:

... must present themselves strategically to each of these audiences, carve out a role and a set of expectations that is acceptable to each and also serves their own financial interests, while resolving or at least eliding the contradictions between them. (Gillespie, 2010: 7)

While Banks (2010: 234) rightly believes that 'a broad coalition of government, business and citizenry is likely to be most effective in reducing the harm caused by hate speech' this ideal scenario does not seem likely at present. It may be unjust to expect individual users (particularly those most harmed by hate speech) to take responsibility for online hate speech, but from a practical perspective Banks argues there is still a role to be played by users of social media sites. He concludes that:

... individual responses to online hate may only have a limited impact on access to online material, but the degree of responsibility of individual users can

both promote a culture of intolerance towards online hate and contribute to efforts to 'reclaim' the web. (Banks, 2010: 238)

While Banks' argument is persuasive, besides 'alerting relevant authorities to incidents of cyberhate which may warrant law enforcement intervention' (Banks, 2010: 238), he mostly leaves unelaborated what precisely is entailed by his call for individual users to take a degree of responsibility for online culture in spaces like Facebook and Twitter. Furthermore, his claims to the minimal effectiveness of these individual responses may remain unconvincing for those who hold to US approaches that eschew state regulation in favour of individual counter speech.

In light of this context, I now turn to the creative activist practice of Facebook flarf as a method of challenging individual incidents of hate speech on the site. I interpret this practice as users taking a degree of responsibility for online hate speech and attempting to do something about it. This case study demonstrates the utility (and limits) of these individual responses, and perhaps more crucially, shows one way in which users at a particular moment attempted to 'reclaim the web,' negotiating and acting within the regulatory regime that Facebook presented in 2012 when the practice was most active. I argue that Facebook flarf belongs to an emerging trend of discursive activist strategy that takes an agonistic approach to online discursive norms, repurposing some of the tools and tactics more traditionally associated with online trolling, but which are more simply reflective of current internet culture, described as a culture of ubiquitous memes (Phillips, 2013; Leaver, 2013).

What is Facebook flarf?

Facebook flarfing consists of tagging Facebook pages and apps in text fields such as status updates and comment threads, building up strings of phrases into an often-absurd or ironic comment, message or poem. Introduced by the site sometime in late 2009, a tag is made by typing the '@' symbol followed by one or more characters, resulting in the appearance of a drop down text box with options reflecting choices suggested by the tagging algorithm, chosen from the pool of total Facebook pages. Selecting one of these options inserts the name of the page or app into the text field which then appears in blue as a hyperlink, visually distinguishing the tag (or 'flarf') from 'ordinary' text comments which appear in black. Tags may be anything from a single letter or word up to whole sentences or even paragraphs of text, and this otherwise innocuous technical feature, most commonly used for tagging individual users, briefly blossomed into a rich, if relatively niche, variety of poetic and activist practices.

The exhausting breadth of pages that one can tag in order to flarf transforms it into a creative, playful practice of digging through the extensive source material, filled with cultural detritus, typos, and memes that make up the site. The distinctive content of this giant archive of text is primarily Facebook page titles, populated by the hundreds of millions of Facebook users over the many years the site has existed, giving Facebook flarf much of its distinctive character. As a result, flarf poems and comments take on a characteristically ‘Facebook’ feel, involving cultural tropes the site has facilitated, such as briefly popular meme pages that conform to certain repeated tropes or structures – for example, the myriad ‘*The awkward moment when...*’ pages, a meme that was quite popular on the site around 2011.

In attempting to describe the effect and aesthetic of Facebook flarf, Caleb Hildebrand (2012) suggested that it repurposes the objects of preference (pages that people have ‘liked’) to subversive ends, and that it ‘asks naïve users of [Facebook] to consider the ways in which their sincerely expressed sentiments may be twisted and blasphemed.’ Earnest expressions of ‘likes’ and interests, captured in the form of Facebook pages, become hijacked by playfully irreverent flarfers and the noise of these nonsense tags – by standing out in their distinct, hyperlink-blue text – adds to an aesthetic of confusion or disarray that is often in stark contrast to the otherwise clean corporate experience of Facebook. This effect can also be viewed as a process of ‘personalisation’, or user-customisation of the Facebook environment – an attempt to ‘take control’ or have some say in the tone or feel of an online space. In the next section I return to this theme in order to connect this impulse with a sense of taking responsibility for online spaces and the cultures that emerge within them.

Facebook flarf has roots in an earlier poetic form based around tailoring Google searches to achieve exquisitely bizarre results. The poet Gary Sullivan, who was involved with these early efforts (emerging as early as 2001), has described the practice as emerging in response to the sombre, patriotic mood that descended upon the US post-9/11 (Sullivan, 2009). Calling it an early symptom of the ironic phase of US culture, early examples of poetic flarf were often filled with ethnic and racial slurs, using repeated Google searches with obscure and unrelated strings of words (e.g. “peace” + “kittens” or “pizza” + “kitty”) as a way of finding offensive, horrible or absurdist text to share amongst a community of internet-savvy poets (Sullivan, 2009). In his history of the practice, he gives a three-pronged definition of flarf as a particular noisy/messy/irreverent aesthetic (similar to the aesthetic described above), as a verb (as in bringing out something innately ‘flarfy’ about a text) and, importantly, as a community practice (Sullivan, 2009).

Facebook flarf is often performed as part of a small group or community, such as the Alt.

Lit community, which was responsible for adapting the practice to Facebook. A diffuse and difficult to define online social movement centred around self-publishing, Alt.Lit writers and poets like Steve Roggenbuck played a large role in popularising the use of Facebook flarf (Roggenbuck, 2011a; Roggenbuck, 2011b). Hildebrand (2012) presents screen captures of Roggenbuck interacting with his fans and fellow community members, as they take turns posting short absurdist messages in combinations of flarf tags riffing off of each other's postings. Taking the lead in a post, Roggenbuck's comment "Hashtag Whale Sex" (formed by tagging Facebook pages 'hashtag' as well as 'whale' and 'sex') elicits response like "Hashtag Whale Hunger Games" (a combination of tags for 'hashtag' 'whale' and 'hunger games'), as well as "Hashtag Haveing Fun" [sic] (pages 'hashtag' and 'haveing fun' [sic]). Flarf seemed in many early instances to beget flarf from others in a contagious and spontaneous elicitation of group activity, with flarf threads often degenerating into absurdity while facilitating mutual engagement with community members.



Figure 1

As is clearly on display in Figure 1, left on a publicly shared photo posted to Roggenbuck's Facebook wall in mid 2012, much of the flarf is absurdist, vaguely sexual or confrontational, echoing earlier pre-Facebook found-text flarf. Many references in the thread to 'yolo' (You Only Live Once, an acronym peaking in popularity and cultural awareness at the time) give the flarf a distinctly pop culture and timely feel. While the pages tagged ("What you lookin' at? You all a bunch of fuckin' assholes.") might be read as antagonistic or confrontational if expressed outside this context, being on a prominent figure's Facebook wall contributes to this being more of a playful public performance than a provocative activity.

The practice gels with what Whitney Phillips has observed as the changing relationship between trolling as a practice and the cultural signifiers such as memes (which are often used in flarf posts) that would once have identified these activities more clearly as 'trolling.' Phillips (2013) notes that, 'what used to provide unequivocal proof that trolling was afoot no longer (necessarily) denotes anything, other than a basic familiarity with memes.' Likewise, when evaluating a public Facebook page organised to protest the Australian broadcast coverage of the 2012 Olympic games which also frequently deployed memes and tropes associated with trolling and troll culture, Tama Leaver (2013: 226) has noted that 'the iconography of trolling, if not the wholesale practice itself, has entered mainstream culture, moving away from the subcultural fringes.' In flarf postings on Facebook we often find something perhaps even more reflexive: a subculture (Alt.Lit poets) repurposing the ubiquity of memes and redeploying its artifacts (Facebook pages) for an altogether different subcultural practice.

Facebook flarf in general and distinctly activist deployments of flarf in particular, seemed to peak in 2012, and instances of the practice have become fewer and farther between. There was a particularly noticeable decline in its use for activist ends around the same time, possibly attributable to both technical changes to the Facebook platform (consideration of which are beyond the scope of this article), and more importantly to changes in Facebook's regulatory environment, particularly around enforcement of its hate speech policies, which I turn to in the final discussion. But for a while Facebook flarf seemed on the verge of becoming a more widely known and accepted practice with potential to alter the content and tone of online discussions in the circles of those who deployed it. In the next section I describe some of these more active and confrontational uses of flarf, within the limited scope for individual agency allowed within Facebook's regulatory context.

The Utility of Facebook Flarf

On the 18th of January 2013 Melbourne based Facebook user Kristina Arnott left a public comment on the wall of McDonald's Australia's public Facebook page. Her complaint refers to a McDonald's advertisement that was screening on free to air television at the time, and it quickly attracted likes and comments from other Facebook users:

Is it really necessary to include stupid young men beeping at a woman in your advertising? And that woman smiling shyly as if she is flattered and even enjoys such interactions? I don't know a single woman who likes being beeped at or yelled at or leered at from cars, in fact most of my female friends find it annoying or even enough to make them feel a little uncomfortable in certain circumstances. Do you really need to encourage such behaviour by further normalising it and making it seem like a positive experience for all involved?? (Arnott, 2013)

The post, which was publicly visible and thus able to appear in the newsfeeds of those whose friends liked or commented on the post, quickly gained a significant amount of 'likes' and attention. Many however disagreed with Arnott's assessment, and dismissed her concerns, with one (male) commenter telling her to 'settle down' and another (also male) telling her to 'find something better to complain about.' The thread was quickly derailed from initial discussions of the advertisement by commenters, who made sexist jokes and steered the discussion into irrelevant territory. It initially received only a small amount of engagement from individuals attempting to argue against or reason with these detractors.

Several acquaintances of mine noticed the thread and began intervening, first with earnest comments and attempts to engage argumentatively, with what Foxman and Wolf (2013: 129) would call 'counter speech' that revealed 'its deceitful and false content.' When it quickly became clear no good-faith discussion was to be had with these detractors, my acquaintances began posting noisy, agonistic and nonsense flarf. I joined in myself and, along with a small group of acquaintances, began leaving flarf comments which referenced or played with the original nature of the complaint, such as 'I Wish I Were Diving In a hotted up stolen taxi Beepin at random pedestrians & waving just to mess with their heads Hard Cunts' [sic]. Other comments included very simple or short non-sequiturs, such as 'börp' [sic] and longer phrases made out of several stitched together page names like:

I was only 19 when i stopped talking to you, i found out how depressing u made my life "Sandwich Jokes!!! DeR DaH Im FuCKInG sTuPiD lAla" it is so hard being so fucking funny all the time NOT [sic]

Even with a small number of commenters, it was possible to partially drown out the offensive jokes and comments under a tide of blue-text nonsense, which often played with or obliquely commented on the sexist comments that were being left. The inclusion of a reference to “Sandwich Jokes!!!” was clearly referring to a comment left earlier in the thread expressing the classic sexist trope ‘make me a sandwich,’ which the flarf here was criticising and calling out for not being funny. From a practical perspective, the flood of flarf comments was making the derogatory comments harder to see, both visually (due to the blue text ‘noise’ that stood out and surrounded them) and statistically (with more nonsense comments quickly being left than harmful ones). A feature of Facebook comment threads is that once a thread reaches a certain number of comments earlier ones become hidden and a user needs to select “view previous comments” in order to expand the thread and see comments left earlier. This leads to a dynamic in which it becomes something of a contest over who can have the last word, with newcomers to the thread less likely to see the offensive comments, having to scroll back up to see them. As trifling an achievement as this may seem, moving comments further up a thread chronologically by adding to the end of a thread is precisely the kind of limited strategy afforded to individuals in their responses to hate speech on Facebook. For one, it does nothing to prevent *new* instances of hate speech from being posted, merely hiding older ones. It also doesn’t take flarf in particular to do this kind of flooding of threads, however the nature of flarf as somewhat viral, often eliciting further flarf responses from others, does contribute to its efficacy in this respect. Flarf also provides users with a pattern or template to employ when they wish to flood or drown out hate speech in threads like this, perhaps performing a similar function to the ‘SAGE’ feature of the 4chan imageboard – when a user wishes to avoid ‘bumping’ a thread on 4chan, either to express disagreement or dislike, and avoiding drawing further attention to the thread, users can place ‘SAGE’ in the ‘options’ field and the post will not be bumped to the top of the imageboard in question upon posting.

These socio-technical practices carry real weight, often acting as ‘boundary-policing social practice’ (Manivannan, 2013) and can be considered a somewhat more efficacious aspect of Facebook flarf than simply the practical aspect of flooding deleterious comments to make them harder to notice. Furthermore, I would position the kind of action undertaken by these flarf commenters as what Frances Shaw (2012: 42) calls ‘discursive activism’, which she defines as:

... speech or texts that seek to challenge opposing discourses by exposing power relations within these discourses, denaturalizing what appears natural (Fine, 1992: 221) and demonstrating the flawed assumptions and situatedness of mainstream social discourse.

The way flarf performs this challenging of norms is twofold: firstly by demonstrating both competence or mastery of the Facebook platform itself and secondly by being a form of meta-textual play that can be read as a form of personalisation and imposition of a ‘flarfy’ tone, style or aesthetic upon an online space. I would argue that flarf is similar to the creative use of emoticons and other typographical elements unique to digital textual communication, a phenomenon Brenda Danet (2001) has called ‘Cyberplay.’ These kinds of personalisations demonstrate a high level of experience with navigating the social and technical spaces of Facebook – flarf-tagging being a relatively rarefied practice. Significantly for this argument, Kelly, et al. (2008: 2381) agree with this interpretation, having found that, in certain online discussion situations, greater employment of these kinds of textual ‘personalisations’ (such as emoticons, abbreviations like ‘lol,’ and to which I would suggest we could add flarf) often result in the users employing them being perceived as more experienced, or more intelligent than users who do not personalise their text.

Engaging in flarf is also implicitly a form of meta-textual discussion, drawing attention to the medium of communication itself. Facebook flarf performs a reflexivity that makes it very difficult to forget one is speaking or writing *on Facebook* whenever it appears. This is both because of the contrasting blue hyperlink text colour that makes it stand out from ‘ordinary’ text, and because it frequently, often as much by design as by incidence, makes reference to or invokes the cultural tropes of Facebook itself. Engaging in flarf can send a meta-textual message to those leaving hate speech comments to remind them that they are on the shared-public space of Facebook, and without having to say as much, that the flarfist(s) are actively refusing or resisting entering into ‘good faith’ discussion with this material, thereby relegating it to ‘beneath discussion.’ Metaphorically, this could be compared to a situation where, in a face-to-face conversation, one or more individuals turn away from or present their backs to a person in order to communicate social displeasure or a wish to exclude someone. In the thread on the McDonald’s Facebook wall, some good faith discussion with the dissenting parties who rejected Arnott’s argument was attempted, but was quickly redirected into flarf-based mockery when this good faith dialogue was not reciprocated. Not much else can be done when those espousing hate speech cannot even be engaged with sincerely. Foxman and Wolf’s (2013) claims about the efficacy of counter speech seem to miss their mark when facing such recalcitrance as regularly occurs in online spaces, a fact no doubt encouraging users who wish to take responsibility for harmful speech online to explore alternative methods of engagement, like flarf.

A similar dynamic of refusal-to-engage occurred on a thread on the Facebook page for fictional character and “lambassador” Sam Kekovich, created for a series of advertisements promoting lamb eating on Australia Day. One exchange in this thread consisted of subtly poking-fun at the reactionary nature of some of the sites fans and their use of the phrase

“un-Australian” – a dogwhistle term with often racist connotations. Most of the preceding comments were subtle criticisms of the page made by playing at ignorance, however once other sincere commenters were drawn to engage with them, flarf made an appearance, communicating something beyond just the content of the flarf, and again indicating some kind of social displeasure or opprobrium via meta-textual play. At one point in the thread, a fellow user (unknown to me) directed a comment at my acquaintance, to which he replied with a pithy plain-text ‘let me help clear it up a little for you’ followed only by a single, lengthy flarf tag constructed out of one extraordinarily-long page title:

i dont like drama , childish people . uma hella cool person till uu disrespect me . i hate shit talkers . if uu referrin tah me in a status go on ahead n tag me in it...if uu got balls enuff . i dont care about much anymore thats just how thee cookie crum [sic] (Midworth, 2012)

The posting was ambiguous about the extent of its seriousness, repurposing a typo-strewn cultural artifact of the Facebook platform, and yet the sense of displeasure was clear. This social aloofness and unwillingness to engage ‘normally’ or sincerely with hate speakers signalled by flarf also works to send an implicit message to many other potential or actual *sincere* commenters, who may well share the same concerns about the potentially racist nature of the discussion or the Facebook page itself, and the same objections to the hate speech. The message it sends to these potential allies is *do not waste your time*, and particularly *do not bother engaging* (or at least, do not bother engaging in good faith since it will only cost you time, energy or emotional resources). This can be doubly important given the ambiguous context in which much internet discussion in the quasi-public space of Facebook now occurs. As previously mentioned, the classic ‘tells’ of trolling no longer denote anything more than what Phillips (2013) describes as ‘a basic familiarity with memes’ leaving earnest individuals keen to challenge, address or in some way take responsibility for online cultures of hate, without any clear indication of what is worth spending their time or energy engaging with. When it is used in this way, Facebook flarf can communicate to other earnest and well-meaning critics of hate speech that there are people who are willing to ‘play’ with these hate speakers (sincere or otherwise), and to ‘outdo’ potential trolls. Flarf displays a willingness to take on, and even enjoy playing with commenters on these objectionable pages, so those for whom this is an emotionally charged issue (as much hate speech is for its intended recipients) do not have to. In this way it offered, for the period that it was most active, a way for some users of the site to take charge and take responsibility for the online spaces they inhabited, in much the way Banks (2010) calls for.

A significant limitation on flarf’s practical efficacy, then, is that it largely relies upon

individual actions – to achieve much of the effects mentioned above (particularly the practical hiding of hate speech behind a flood of blue text) it requires a relatively sizeable and active community relative to the offending hate speakers. Those wishing to be effective with flarf and make a significant contribution to online culture on Facebook must be willing and able to respond to hate speech in a timely manner, since much of the impetus behind this goal is preventing the offending hate speech from being seen. All of flarf’s positive effects – from the drowning out of hate speech, to expressing opprobrium towards it, as well as contributing to online norms – are significantly diminished when deployed by only one or a small number of individuals. In other words, on the “neutral” Facebook platform which the service provides, Facebook flarf still commonly relies for success on strength in numbers and the organisation of communities, which I only observed happening in an ad hoc and small-scale way in the community of users deploying it.

In the McDonald’s thread, the ‘drowning out’ function was limited in that it was only effective for the period flarfers were active and available (and invested enough in the thread) to keep commenting, which petered out as the discussion went on into the late evening and people drifted off to bed or other activities. Furthermore, a significant amount of the initial flarf that was left was later deleted over the next hours or days, presumably by Facebook itself, likely in response to other users ‘flagging’ the comments as spam (perhaps done by commenters expressing the sexist speech themselves). The flarf itself may have fallen foul of Facebook’s community standards, the most likely culprit being Facebook’s community standards regarding ‘phishing and spam,’ yet the description of what counts as such only currently refers to unsolicited commercial contact. It is conceivable, however unlikely, that the comments that went missing were removed by the operators of the McDonald’s Facebook page itself, however there was no other indication that the thread was ever observed or moderated by the page owners who, much like Facebook themselves, would probably prefer to avoid liability or responsibility for what is said on their page. If it *was* the administrators of the McDonald’s Australia Facebook page, they deemed it fair to leave some flarf but not others (including some flarf containing profanities), and as far as it was possible to tell they also removed none of the sexist comments. An alternative explanation altogether is that some flarf posts tripped automated spam detection and prevention algorithms, which possibly automatically flagged them for review by Facebook’s moderation staff, explaining their later disappearance.

This is yet another example of the regulatory environment that Facebook presents to users, one that is often opaque and that needs to be constantly negotiated by users wishing to explore what freedom and constraints they face when attempting to take responsibility for online culture and challenge hate speech on Facebook. In the following section I turn to discuss in more detail this regulatory environment and some attempts that

have been successful at pressuring Facebook into changing it. This informs both how we view Facebook flarf's history and potential, and why it may have emerged when it did. It also illuminates details of the broader landscape facing users and groups invested in the elimination of the worst aspects of hate speech online.

Facebook's Regulatory Environment and its Relationship to Individual Action

Facebook flarf can be interpreted, at least partially, as a response to Facebook's historically lacklustre and inconsistent policing of its aforementioned community standards – a set of regulations with greater room for interpretation (even misinterpretation) and more lackadaisical enforcement than the stricter (and often legally binding) terms of service. As of late 2014, Facebook's own community standards:

... [do] not permit hate speech, but distinguishes between serious and humorous speech. While we encourage you to challenge ideas, institutions, events, and practices, we do not permit individuals or groups to attack others based on their race, ethnicity, national origin, religion, sex, gender, sexual orientation, disability or medical condition. (Facebook Inc., 2014)

As mentioned earlier in Frydman and Rorive's (2002) discussion of intermediaries, Facebook does not take an active role policing the content of expressions uttered on its platform, as it is not legally required to. This is because Facebook is (presumably, though this has likely not been tested in court) protected by the same safe harbour provisions applicable to internet service providers and internet hosts (of which Facebook would be considered to be one example). To a degree this is explained by Gillespie's (2010) argument about the rhetoric of the 'platform' and the way the phrase allows sites like Facebook to downplay their own liabilities and responsibilities for everything from the illegal to the simply sordid material that circulates through the service. Gillespie (2010: 12) reminds us that 'platform':

... is a valuable and persuasive token in legal environments, positing their service in a familiar metaphoric framework – merely the neutral provision of content, a vehicle for art rather than its producer or patron – where liability should fall to the users themselves.

Instead of an active policing of the material that passes through its servers, and because of this cultural-discursive positioning as a neutral facilitator of content, Facebook can take a reactive approach to policing content on its platform. It relies on user-generated reports via the flagging of particular content, such as comments, statuses and pages. This introduces a certain permissive dynamic to the social regulation of hate speech on the site, a kind of 'quasi-privacy' that lets users in their own semi-private spaces get away with transgressions that might receive social or legal sanction if they did not avoid the notice of other users. One consequence of the reactive approach to regulation of hate speech on Facebook is it becomes possible and even likely that concerned users fail to notify Facebook of instances of hate speech in time to prevent its intended recipients seeing it. Indeed, the burden of 'reporting' instances of hate speech to some extent *presumes* that recipients will see it (and then report it). In my conversations with some of the flarf activists I observed on Facebook it is clear that many users do not even bother to report hate speech to begin with, particularly given that until recently Facebook's implementation of its community standards policy was notably unpredictable and highly permissive when faced with the issue of public hate speech online.

One needs only to recall the numerous hate speech promoting pages that existed on the site with the tag 'controversial humour' to note this lax enforcement at work. For instance, the Facebook page 'Aboriginal memes,' which garnered a great deal of public attention in August of 2012 in the Australian media, following a series of grassroots campaigns organised to put pressure on Facebook to close down the page – including two separate online petitions that attracted several thousand signatures (Oboler, 2012: 10). The page itself, which posted image macros (photos with captions) that played upon some of the worst racial stereotypes of indigenous Australians was reported by many Facebook users, with the Australian Online Hate Prevention Institute documenting in their 2012 report on the page their own attempts to have the page removed. Facebook subsequently declined to deactivate the page, merely prepending the warning phrase "[controversial humour]" to the page's name, a practice that it employed across the site with numerous "controversial" sites that promoted and propagated hate speech against various groups. The Australian based Online Hate Prevention Institute's report described the decision as 'creating an attitude where people feel racism is acceptable.' (Oboler, 2012: 56) The report's author takes a markedly different approach to the 'neutral platform' discourse, noting that 'Facebook is not a neutral player, but is actively promoting this shift based on their "Facebook Principles."' (Oboler, 2012: 56) This is a sentiment clearly at odds with the site's positioning of itself, as having little or no responsibility for the content on it. The Australian Human Rights Commissioner met several times with the company about the site, and it was further referred to the Australian Communications and Media Authority. (Oboler, 2012: 10, 58) The Aboriginal Memes page was eventually taken down, but only after the threat of state intervention and regulation became apparent.

In response to this and the many other instances of lax and uneven enforcement of Facebook's own community standards, in May of 2013 a coalition of groups, led by the 'Women, Action and the Media' activist group, the Everyday Sexism project, and more than 100 affiliated women's and activist groups, petitioned Facebook 'to take concrete, effective action to end gender-based hate speech on its site.' (Women, Action and the Media, 2013) Their actions were in response particularly to a long string of incidences and the widespread perception that Facebook did not treat hate speech directed at women as seriously as other forms of hate speech, for instance, anti-Semitism and racism. Journalist Dara Kerr (2013), reporting on the action, highlights the extremity of some of the content that would often be reviewed and, without the looming threat of state intervention or regulation, would not be removed from Facebook instead receiving the "controversial humour" tag, despite being in clear breach of the site's standards. Kerr (2013) notes that:

... several Facebook pages have popped up that encourage or make a joke of violence against women, pages like Fly Kicking Sluts in the Uterus, Violently Raping Your Friend Just for Laughs, and Raping your Girlfriend.

The Women, Action and the Media petition which was sent to advertisers to notify them that their advertisements were running alongside such objectionable material was a modest success, with the coalition of groups managing to raise awareness of Facebook's permissive acceptance of hate speech directed at women with both the public and, perhaps just as importantly, with several major advertisers on the site (Kerr, 2013). Facebook's response to the petition, and the growing pressure it faced to address the issue, came in the form of a statement from Marne Levine (2013), VP of global public policy at the site. Crucially she acknowledges Facebook's failures in enforcement, specifically –

In recent days, it has become clear that our systems to identify and remove hate speech have failed to work as effectively as we would like, particularly around issues of gender-based hate. In some cases, content is not being removed as quickly as we want. In other cases, content that should be removed has not been or has been evaluated using outdated criteria.

As a result, many fewer pages on Facebook exist with the 'controversial humour' tag and the result has been fewer visible instances of public hate speech on the platform since. However this is only true for pages and groups, the more public and visible spaces of the site, and the same cannot be said for individual comments expressing hate speech, the reporting processes for which remain largely confined to hiding the content from the sight of those objecting to it. An important notice at the very bottom of Facebook's community standards page reminds users that:

... it's possible that something could be disagreeable or disturbing to you without meeting the criteria for being removed or blocked. For this reason, we also offer personal controls over what you see, such as the ability to hide or quietly cut ties with people, Pages, or applications that offend you.

Facebook's regulatory environment, then, still has significant gaps and weaknesses in its approach to preventing and addressing hate speech, continuing in places to defer responsibility for hate speech to individual users via 'personal controls.' Despite the success in ensuring Facebook more actively addresses some of the more public components of the site like the offensive pages, the fundamental structure remains one of non-intervention by the site at the level of individual user speech, and the deferment of responsibility to individuals "offended" by hate speech remains in place. User driven practices like Facebook flarf may contribute to addressing the remaining gaps that allow for hate speech in these other areas of the site that are less able to attract the kind of mainstream attention required to instigate significant structural change. Users can contribute, as we saw earlier, in practical ways, such as by drowning out and making hate speech less visible, and by saving the emotional energy of those who would engage in counter speech. They can also contribute in more discursive or normative ways, contributing to 'taking back the web' as Banks (2010) suggests, through personalisation and discursive activism. Facebook flarf outlines the sense in which a space exists between the competing regulatory regimes of Facebook and state based legislation, a space that leaves significant room for user-led creative responses to online problems like hate speech.

Since its peak in or around late 2012 and early 2013, the amount of flarf has, anecdotally, reduced in the circles where I observed that it was once fairly common. This may be partly attributable to the wearing off of the novelty of the activity, as much as to other changes in Facebook's enforcement of its community standards, following the WAM! Coalition and others' efforts to confront Facebook and force it to take greater responsibility for the public hate speech promulgated on its platform. The WAM! Coalition group's success in attaining Facebook's redoubled commitment to enforce its own community standards via regulation may have also contributed to the reduced need for Facebook flarf, since reporting public pages that promote hate speech through Facebook's reporting tools should (in theory) now prove more effective. This is in spite of Facebook's largely deferential approach to individual utterances of hate speech (in comment threads in particular) which it remains difficult to flag and report, relying largely on "personal controls" that do little to change what is visible to the public.

This discussion serves to underscore the negotiated space in which Facebook flarf and other individual responses to hate speech exist, and the limited avenues that are available to restrict and suppress hate speech within the larger regulatory regimes of both the platform itself and state based legislations. There are significant costs involved with deploying flarf individually and in small communities, in terms of the time and effort it takes to suppress hate speech, further reinforcing the importance of larger scale interventions, public pressure and state level responses, as both Levin (2010) and Banks (2010) have argued.

Conclusion

This paper has aimed to inform our understanding of the problem of hate speech online and the unique constraints and opportunities for intervention by individual users on Facebook, principally via a small case study of the creative activist practice of Facebook flarf. I began by discussing the broad international context that hate speech occurs within, highlighting difference between United States and European comfort with the regulation of hate speech and questions around state intervention. Just as important as these historical and cultural differences, however, are practical issues with online hate speech interventions, with governments and state agencies often unwilling or unable to regulate online hate speech, and with intermediary web services like social media similarly reluctant. Intermediary services like Facebook and Twitter have often sought to position themselves as a 'neutral platform' the better to avoid liability for the material that passes through their services.

Into this context I placed flarf, beginning with a history of the practice in found-text poetry and the way in which it came to be reconfigured and repurposed to employ the Facebook tagging algorithm. I then elaborated the specific utility I saw in Facebook flarf activism as observed during its peak in 2012, arguing that Facebook flarf presents a useful case study for theories of regulating and responding to hate speech online. I argued that Facebook flarf has some ability to drown out hate speech practically and aesthetically, but perhaps more importantly it can serve to communicate social opprobrium and community limits on acceptable discourse online. Facebook flarf represents an encouraging attempt by users to 'take responsibility' for online hate speech and online culture in the spaces they frequent, through personalisation and the performance of an expertise within the platforms affordances. It also communicates a meta-textual and reflexive awareness of the medium of communication itself. I situated the practice of Facebook flarfing for activist ends within a contemporary context of ubiquitous memes and the uncertainty around the sincerity of online comments and discourse, viewing flarf as an example of discursive activism that

repurposes the tropes and practices of troll culture.

Yet despite all this, flarf for all its promise remains constrained in a number of significant ways by the larger Facebook regulatory context, with important structural features that preclude individual user responses on their own constituting a satisfactory response to hate speech on the platform. Specifically, when compared to the effects of public and advertiser pressure on Facebook to better implement and enforce its own community standards policies in public spaces, flarfing seems inadequate to this particular type of problem. And yet, even after significantly improving how Facebook enforces its community standards policy, gaps in Facebook's regulatory regime leave responsibility for reporting and responding to individual hate speech utterances up to users.

The space between these constraints and the possible utility in employing user-led strategies like Facebook flarf leads me to affirm the perspective that finds that, in Banks (2010: 234) words, 'a broad coalition of government, business and citizenry is likely to be most effective in reducing the harm caused by hate speech.' How activists, researchers and users of social media sites can realise a more effective coalition of responses to hate speech lies beyond the scope of this paper, but if Facebook flarf is any indication, there will likely remain a role in any larger regulatory framework for the responsible actions of individual users and communities to challenge hate speech, enforcing the standards they wish to see online.

Biographical Note

Benjamin Abraham is a PhD candidate at the University of Western Sydney, researching network cultures, the philosophy of non-human objects, and digital games. He has published on the internet activist technique of 'Fedora Shaming as Discursive Activism' (2013) as well as on internet communities of video game critics. His current research project is on video game depictions of climate change.

References

- Arnott, Kristina. 'Is it really necessary to include stupid young...?' *Facebook*. 18 January (2013). <https://www.facebook.com/McDonaldsAU/posts/494412147268390>
- Banks, James. 'Regulating Hate Speech Online.' *International Review of Law, Computers & Technology* 24.3 (2010): 233–239.
- Barnett, Brett A. *Untangling the Web of Hate: Are Online "Hate Sites" Deserving of First Amendment Protection?* (Cambria Press, 2007).
- Berlet, Chip. 'When Hate Went Online.' *Northeast Sociological Association Spring Conference in April* (2001): 1–20.
- Brazeal, Gregory. 'How Much Does a Belief Cost?: Revisiting the Marketplace of Ideas.' *Southern California Interdisciplinary Law Journal* 21 (2011).
- Butler, Judith. 'Ruled Out: Vocabularies of the Censor.' in Post, Gary (ed.). *Censorship and Silencing. Practices of Cultural Regulation* (Getty Publications, 1998): 247–260.
- Danet, Brenda. *Cyberpl@y: Communicating Online* (Oxford: Berg, 2001).
- Facebook Inc. Facebook Community Standards. *Facebook*. (2014) <https://www.facebook.com/communitystandards>
- Fine, Michelle. 'Passions, Politics, and Power: Feminist Research Possibilities,' in Fine, Michelle (ed.). *Disruptive Voices: The Possibilities of Feminist Research* (Ann Arbor, MI.: University of Michigan Press, 1992): 205–231.
- Foxman, Abraham H., and Wolf, Christopher. *Viral Hate: Containing Its Spread on the Internet* (Palgrave Macmillan, 2013).
- Frank, Jenn. 'How to Attack a Woman Who Works in Video Gaming.' *The Guardian*. 1 September (2014). <http://www.theguardian.com/technology/2014/sep/01/how-to-attack-a-woman-who-works-in-video-games>
- Frydman, Benoît, and Rorive, Isabelle. 'Regulating Internet Content Through Intermediaries in Europe and the USA.' *Zeitschrift für Rechtssoziologie* 23.1 (2002): 41–59.
- Goldman, Alvin I., and Cox, James C. Speech, Truth, and the Free Market for Ideas. *Legal Theory* 2.1 (1996): 1–32.
- Guardian staff. 'Two jailed for Twitter abuse of feminist campaigner.' *The Guardian*. 25 January (2014). <http://www.theguardian.com/uk-news/2014/jan/24/two-jailed-twitter-abuse-feminist-campaigner>
- Hildebrand, Caleb. 'A few thoughts on facebook flarf.' *Flaneur in Pajamas*. May 6, 2012. <http://flaneurinpajamas.tumblr.com/post/22538991313/a-few-thoughts-on-facebook-flarf>
- Kelly, Erika, Davis, Blake, Nelson, Jessica, and Mendoza Jorge. 'Leader Emergence in an

Internet Environment.' *Computers in Human Behavior* 24.5 (2008): 2372–2383.

Kerr, Dara. 'Facebook Pulls Pages Depicting Violence Against Women.' *CNET*. 29 May (2013). http://news.cnet.com/8301-1023_3-57586781-93/facebook-pulls-pages-depicting-violence-against-women/

Leaver, Tama. 'Olympic Trolls: Mainstream Memes and Digital Discord?' *The Fibreculture Journal* 22 (2013): 216–233. <http://twentytwo.fibreculturejournal.org/fcj-163-olympic-trolls-mainstream-memes-and-digital-discord/>

Levin, Abigail. *The Cost of Free Speech: Pornography, Hate Speech and Their Challenge to Liberalism* (Palgrave Macmillan: 2013).

Levine, Marne. 'Controversial, Harmful and Hateful Speech on Facebook.' *Facebook*. 29 May (2013). <https://www.facebook.com/notes/facebook-safety/controversial-harmful-and-hateful-speech-on-facebook/574430655911054>

Lewis, Helen. 'Game Theory: Making Room for the Women.' *The New York Times*. 25 December (2012). <http://artsbeat.blogs.nytimes.com/2012/12/25/game-theory-making-room-for-the-women/>

Manivannan, Vyshali. 'Tits or GTFO: The logics of misogyny on 4chan's Random - /b/.' *The Fibreculture Journal* 22 (2013): 109–132. <http://twentytwo.fibreculturejournal.org/fcj-158-tits-or-gtfo-the-logics-of-misogyny-on-4chans-random-b/>

Meyer, Robinson. 'The Existential Crisis of Public Life Online.' *The Atlantic*. 30 October (2014). <http://www.theatlantic.com/technology/archive/2014/10/the-existential-crisis-of-public-life-online/382017/>

Midworth, Luke. 'CAnt wait until the year 2135 and the husk of...' *Facebook*. 13 January (2013). <https://www.facebook.com/SamKekovich/posts/10151366346445939>

Oboler, Andrew. 'Aboriginal Memes and Online Hate'. *Online Hate Prevention Institute*. Report IR12–2, October (2012). <http://ohpi.org.au/aboriginal-memes-and-online-hate/>

Phillips, Whitney. 'What an Academic Who Wrote Her Dissertation on Trolls Thinks of Violentacrez.' *The Atlantic*. 15 October (2012).* *<http://www.theatlantic.com/technology/archive/2012/10/what-an-academic-who-wrote-her-dissertation-on-trolls-thinks-of-violentacrez/263631/>

Polanyi, Karl. *The Great Transformation: The Political And Economic Origins Of Our Time* (Beacon Press, 2001; 1944).

Roggenbuck, Steve. 'facebook commenting and wall-post flarf.' *YouTube*. 19 April (2011a). <https://www.youtube.com/watch?v=bhg4wOqfkFc>

Roggenbuck, Steve. 'introduction to flarf poetry by steve roggenbuck.' *YouTube*. 2 October (2011b). https://www.youtube.com/watch?v=8Pe_x_BkroM

Rose, Nikolas. *Powers of Freedom: Reframing Political Thought* (Cambridge University Press, 1999).

Sarkeesian, Anita. 'Tropes vs Women in Video Games.' *Kickstarter*. (2012) <https://www.kickstarter.com/projects/566429325/tropes-vs-women-in-video-games/posts>

Sarkeesian, Anita. 'User: FeministFrequency.' *YouTube*. (2014) <https://www.youtube.com/user/feministfrequency>

Schauer, Frederick. 'The Ontology of Censorship.' in ed. Gary Post. *Censorship and Silencing. Practices of Cultural Regulation* (Getty Publications, 1998): 147–68.

Shaw, Francis. 'The Politics of Blogs: Theories of Discursive Activism Online.' *Media International Australia* 142 (February, 2012): 41–49.

Sparrow, Robert and Goodin, Robert E. 'The Competition of Ideas: Market or Garden?' *Critical Review of International Social and Political Philosophy* 4.2 (2001): 45–58.

Sullivan, Gary. 'Flarf: From Glory Days to Glory Hole.' *The Brooklyn Rail*. 4 February (2009). <http://www.brooklynrail.org/2009/02/books/flarf-from-glory-days-to-glory-hole>

Wingfield, Nick. 'Feminist Critics of Video Games Facing Threats in 'GamerGate' Campaign.' *The New York Times*. 15 October, 2014. <http://www.nytimes.com/2014/10/16/technology/gamergate-women-video-game-threats-anita-sarkeesian.html>

Women, Action and The Media. *WAM!* 28 May (2013). <http://www.womenactionmedia.org/fb agreement/>



The LOCKSS System has the permission to collect, preserve and serve this open access Archival Unit



This Issue of the Fibreculture Journal by The Fibreculture Journal Incorporated is licensed under a Creative Commons Attribution 4.0 International License.



OPEN HUMANITIES PRESS

The Fibreculture Journal is published by The Fibreculture Journal Incorporated in partnership with Open Humanities Press.