# “It’s a terrible way to go to work”: What 70 million readers’ comments on the Guardian revealed about hostility to women and minorities online

Becky Gardiner

*Department of Media and Communications, Goldsmiths College, University of London, London, UK*

b.gardiner@gold.ac.uk

Becky Gardiner worked as a journalist for 25 years. She was at the Guardian from 1998-2014, where she held several senior editorial positions including Women’s editor and Comment editor. She is now a senior lecturer in journalism at Goldsmiths College.

# “It’s a terrible way to go to work”: What 70 million readers’ comments on the Guardian revealed about hostility to women and minorities online

In 2006, the Guardian opened many of its articles to readers’ comments to encourage a “conversation” between journalists and their readers. Readers responded enthusiastically, and by 2016 they had posted 70 million comments on the site. However, from the outset many journalists complained about the quality and tone of comments. Female and BAME (Black, Asian and minority ethnic) journalists in particular complained that they were subject to more abuse than their male, white counterparts.

This study finds prima facie evidence to support the journalists’ claims. Using comments that had been blocked by moderators as a proxy for abuse and dismissive trolling, it was found that articles written by women did attract a higher percentage of blocked comments than those written by men, regardless of the subject of the article; this effect was heightened when the articles ran in a particularly male-dominated section of the site. There was also evidence that articles written by BAME writers attracted disproportionate levels of blocked comments, even though the research was not designed to reveal this.

Preliminary research findings were published in the Guardian (Gardiner et al, 2015) and readers were invited to comment on them. Guardian journalists’ experiences of comments were also surveyed. Both sets of responses are analysed here, in order to explore the contested nature of online abuse in an online news media environment, and to evaluate the potential of comments to ‘democratise’ journalism.

Keywords: readers’ comments, online abuse, sexism, racism, Guardian, journalism, moderation, democracy, participation

Subject classification codes: to follow

**Introduction: the Guardian and online comments**

In 2006, most Guardian articles were opened to comments. By 2016, 70 million comments had been left on the site, an average of 65,000 new comments were being posted every day, and 13 FTE moderators had been hired to keep them civil and constructive. In little more than a decade, Guardian journalists' interaction with readers had been transformed utterly – there was now a constant stream of instant reader response to every piece, much of it critical, some of it offensive.

Like other news organisations, the Guardian introduced readers' comments for both editorial and commercial reasons (Henry, 2006). Editorially, it was believed that the interactivity of the web could be used to enhance journalism by harnessing the potential power of readers as fact-checkers and sources of stories, and by providing a forum for civic discourse. Commercially, the hope was that comments would increase both readership and "engagement" with the content – two "metrics" which could attract advertising revenue. So, from the start, readers' comments were seen as simultaneously a democratic virtue, a journalistic resource, and a commodity which could create value.

The Guardian was not alone in this endeavour. By the mid-noughties, comments had become one of the most popular forms of user-generated content within news media (Hermida and Thurman 2008; Jönsson and Örnebring 2011). To some scholars this was evidence of the Internet's inherently democratising tendency, with comments representing a shift in power away from journalists as the sole producers of news**:** see Bruns's (2007) "produsage"; Rosen's (2006) "people formerly known as the audience"; Jarvis's (2006) "networked journalism"; Shirky's (2009) "here comes everybody"; and Gillmor's (2004) "readers know more than we do". Other theorists argued that these somewhat Utopian accounts were based on a misunderstanding of both the internet and media power (Curran et al, 2012; Phillips, 2018).

The Guardian's then editor-in-chief, Alan Rusbridger put himself firmly on one side of this debate, placing reader engagement at the centre of his vision not only for the Guardian, but for the industry as a whole, saying: "We are reaching towards the idea of a mutualised news organisation," (Rusbridger, 2010). In this sense, the Guardian can be seen as a test bed for a theoretical position, one which argues that the Internet is democratising, allowing readers and journalists to meet as equals, and debate diverse views.

The first challenge to that position came in the form of some of the comments themselves. Many journalists found them abusive; some said they had a chilling effect on their journalism. Female journalists complained that they were subject to more abuse "below the line"[i] than their male counterparts (Elliott 2007; Penny 2011; Thorpe and Rogers 2011; Lewis 2013), and BAME journalists spoke out against racist abuse (Hasan, 2012; Younge 2012; Bungawala et al 2012).

Internally, management attitudes shifted. In 2013, the then deputy editor reaffirmed the democratising role of comments, saying they had "wrecked" old hierarchies and helped create a "more levelled world" – but she also acknowledged that: "[S]ometimes readers will say things that are threatening and rude; certain groups, such a women and writers who are not white, can have a difficult time […] Some writers hate it, and it's hard to blame them." (Viner, 2013)

By the time this research was conducted in 2016, "Don't read the comments" had become a mantra throughout the media, and several news organisations had switched comments off altogether (Reagle 2015). This retreat from comments is remarkable when one considers the enthusiasm with which they were adopted.

As an editor on the Guardian from 1998 to 2014, I saw this transformation up close. So, although this current work has at its core a quantitative study of a large

corpus of readers' comments, together with small-scale qualitative studies of the attitudes of both commenters and journalists, it is also in part a broader autoethnographic study of one news media organisation's attempt to democratise media production. I have drawn on my "insider" status to show a group of people – Guardian journalists – "in the process of figuring out what to do" (Adams et al, 2015:2) in what was a period of rapid change, where journalists encountered large numbers of readers for the first time. This article focuses on the gendered and racialised aspects of Guardian comments and, in the words of its first female editor-in-chief, Katharine Viner, to address the "new torrents of racism and sexism"[ii] facilitated by the web. It considers evidence to support the belief that women experienced more abusive and/or dismissive comments than men and, finally, reflects on the implications of this evidence for the democratising potential of comments.

**Comments as a gendered public sphere**

Following Habermas and the critical review of his original theory, some scholars identified comment spaces as a new public sphere in which such deliberation of diverse views can take place (Papacharissi, 2004). Ruiz et al, found some comment sections (including the Guardian's) could "consolidate democratic processes in Western societies", and that online newspapers were becoming "the digital cafés of a Public Sphere 2.0" (2011, 464); Graham and Wright (2015) analysed 3,792 comments (also on the Guardian) and found them to be largely deliberative in nature.

These deliberations could be robust. Impoliteness – even anarchy – can be a valuable feature of these online debates (Papacharissi, 2004), and sometimes media power is challenged in ways that journalists might find confronting: Milioni et al (2012) found that commenters were able to resist journalists' hegemonic perspective; Craft et

al (2015) found that they acted as press critics, holding journalists accountable to social norms; Semmler & Semmler (2016) found that, by offering a wide range of views, comments provided valuable insights into public opinion on complex issues.

However, uncivil or offensive comments were found to be a threat to this democratic ideal, by preventing deliberative discussion (Papacharissi, 2004) and polarizing debate (Anderson et al, 2014). Journalists told researchers that comments were of poor quality (Viscovi and Gustafsson 2013), or abusive (Singer and Ashman 2009), and undermined professional standards (Hermida and Thurman 2008). The tone of comments was also found to deter some readers – particularly women – from participating in comment threads (Pierson, 2015; Martin 2015).

Far from being inherently democratising, the technological affordances of the Internet came to be seen as having amplified sexism and other hostilities (Citron, 2014). Although some accounts pointed out that "nastiness" was also found above the line (Anderson et al, 2014), and that mainstream media norms were the "soil" in which the "weeds" of offensive comments grow (Phillips, 2015), unwelcome readers' comments became the major focus of study.

"Incivility" (broadly, a violation of norms and a lack of respect for other people) received considerable attention (Muddiman and Stroud, 2017, Coe et al, 2014); others found evidence of "antisocial behaviour" (Cheng et al, 2015), or "toxic conversation" (Sood et al, 2012). It is notoriously hard to define online "abuse", but what unifies these accounts is the attempt to better identify problematic comments, in order to better regulate them – for example by automating detection of objectionable comments (Sood et al, 2012) or by identifying antisocial commenters at an early stage (Cheng et al 2015). In newsrooms, too, the regulation of comments became the primary concern, and there

was an industry-wide move towards more moderation, as well as other forms of gatekeeping (Hermida and Thurman 2008, Reagle 2015).

However, existing asymmetries between powerful and more marginalised actors complicate all attempts to regulate online "hate". For example, comment spaces are heavily gendered. In a review of 30 years' of academic literature, Jane (2014) found that, in academia, a "preoccupation with devising a fail-safe identification device for all online hostilities", has lead to a disregard for women's experiences online (2014, 539). Too often, she notes, it has been left to the women who have been the targets of this hostility to expose it, and to theorise the way in which it silences women's voices (Beard, 2017), often just as they are beginning to be heard (Sierra, 2014).

Although there has been significant scholarly work on the prevalence and impacts of misogyny on the Internet (Citron, 2014; Phillips, 2015; Mantilla, 2015), it has often had to rely on individual cases or surveys[iii] for evidence. Both have their limitations: individual accounts might be dismissed as exceptional, or the women were blamed for what had happened (for discussions of the sexist response to women who speak out, see Citron, 2014); surveys, meanwhile, are by definition based on self-reporting, which can lead to skewed results. By including the gender of the author whose work is being commented on as a variable in the analysis of a large corpus of Guardian comments, this is a gap in knowledge this paper hopes to fill.

But this paper also acknowledges that women are not the only group who struggle to be heard in comment spaces. In her feminist critique of Habermas's conceptualisation of the public sphere, Nancy Fraser (1992, 64) argues that "deliberation can serve as a mask for domination [in ways which] extend beyond gender to other kinds of unequal relations, like those based on class or ethnicity."

Shepherd et al (2015) find precedents for "online hate" in earlier mediated communications, and point toward "longstanding issues of exclusion and inequality in public speech". Comments, they argue, can be both a form of challenging existing power structures *and* reinforcing them: "The affordances of comments as […] spaces to speak back to power of course also contained the possibility of replicating existing structures of power in even more vitriolic forms that served to […] deny people's ability to be heard." (2014, 7)

If one is interested in widening participation in the media, the question therefore becomes how we can be both against sexism and racism and supportive of comments as a space where media power can be challenged.

## Research design and methodology

A mixed-mode of analysis was used to explore the research questions, drawing on both quantitative and qualitative methods and using three distinct data sets:

### *1.The analysis of the corpus of Guardian comments*

The analysis of the block rate patterns in the 70 million comments left on the Guardian site was carried out in March 2016. Blocked comments were used as a proxy for "abusive or dismissive" comments, in order to test the hypothesis that articles written by female journalists received more "abusive or dismissive" comments than those written by men.

At the Guardian, moderators blocked (ie removed from view) comments that violated the community standards[iv]: ad hominem attacks, disrespectful, profane, or discriminatory comments, and comments that are so off-topic they threaten to derail the conversation all violated the community standards, and were blocked if seen by

moderators. The limitations of using comments blocked by moderators as a proxy for a broad category of "abusive or dismissive" comments are acknowledged: the research findings are only meaningful if you accept both that the Guardian's community standards are fit for purpose, and that the moderators are reasonably skilled at applying them. However, there is no failsafe method for identifying offensive or unwelcome comments, and as Jane (2014) has shown, repeated demands for "univocal definitions and one-size-fits-all test mechanisms" has inhibited research in this area. Relying on the decisions of trained moderators has the advantage of treating offensive comments as an "emergent field" (Jane, 2014), and one that can only be investigated by taking into account context, apparent intent and likely impact – as Guardian moderators do.

The comments themselves were not read, but were treated as a quantitative data set, with the primary focus on two variables: the gender of the author of the article, and the proportion of comments blocked. The impact of other variables on the block rate, such as the subject of the article, or the section in which it appeared, was also measured. Nothing was known about the commenters themselves.

This part of the data set combined three elements:

(1) The 70m comments left on the site, including the 1.4m comments that had been blocked by moderators[v].
(2) The more than two million articles posted during this period, complete with key metadata, including subject keyword.
(3) A database of 12,655 authors (all those who had written at least two articles during this period), who were classified by gender[vi].

All comment threads were moderated, the vast majority post-publication[vii]. In most cases moderators reviewed comments that had been reported – usually by readers, sometimes by journalists; a filter tool was also used to flag comments containing

particular words (eg profanity). For this reason, not all comments that violate community standards at the Guardian were blocked.

In this study, "abusive or dismissive" comments were those that violated the community standards. According to moderators, extreme abuse such as threats to kill, rape or maim was extremely rare on the Guardian[viii]. Author abuse formed a significant proportion of the comments that were blocked. For example: a female journalist reports on a "pro-life" demonstration, and a reader responds, "You are so ugly that if you got pregnant I would drive you to the abortion clinic myself".

Ad hominen attacks were also common: comments like, "You are so unintelligent". So too were comments like, "Do you get paid for writing this?!" which dismissed the author.

Blocked comments were not always aimed at the writer – they might be directed at other commenters, for example. Nor are they always aimed at individuals. Hate speech as defined by law was rarely seen, but xenophobia, racism, sexism and homophobia were far more frequent. Take, for example, the following comment below an article on the mass drownings of migrants: "The more corpses floating in the sea, the better".

The moderators also blocked comments that were clearly off topic, and so were regarded as attempts to make constructive debate impossible. An example would be a thread about female genital mutilation in which there were multiple comments saying, "Why don't you write about male circumcision?" (a trope of men's rights activism). Taken alone, such comments appear neutral, but in context they might not.

A tiny minority of blocked comments were blocked for legal reasons, and it was not possible to exclude these. Spam was deleted, not blocked, and so were replies to blocked comments, which may or may not in themselves be in violation of the

community standards; neither spam nor replies to blocked comments were included in this research.

The findings reported here were all statistically significant to a significance level of 5%, using the chi square test of independence.

### *2. The staff survey*

372 Guardian journalists – all the staff writers, plus the most regular freelance writers, were emailed a link to an anonymous online survey[ix]; 183 (49%) responded. Of these, 100 were men, 80 were women[x], and 23 identified themselves as BAME. Fifteen were Jewish, and two were Muslim. Most questions were multiple choice, and were designed to discover what journalists regarded as abusive behaviour in comments and on social media, and to discover its frequency, severity and impact. Women and BAME people are over-represented in the responses, and this self-selection limits the reliability of the results. However, they are included here to provide contextual information.

### *3. The commenters' response to the preliminary findings*

The Guardian published a summary of an earlier analysis of the comment dataset on its website (Gardiner et al, 2016). The article was widely read (944,974 page views in the first week[xi]), and 2,458 comments were posted below the article within 24 hours (the thread was then closed). The third and final part of this work is a sentiment analysis of these comments.

The comments were manually coded "negative", "neutral" or "positive". Negative comments were further categorised in terms of the target of their criticism. Four main targets were identified: the research methodology; the journalists who experienced abuse; the moderators; and the Guardian's editorial stance. The coding was

done manually by a PhD student; a random sample of 100 of the comments were subsequently recoded by the author, and very few discrepancies were found.

## Findings

### *The Comments*

The majority of articles on the Guardian were written by men. While the number of articles published increased every year, this gender gap stayed fairly constant: in 2008, 27.52% of Guardian articles were written by women; by 2015 this had risen to 31.74%. Overall, during the period 2006-16, women wrote 27.6% of the articles, and men wrote 72.4%, a gender imbalance that is typical of mainstream media organisations in the UK and the US[xii].

Figure 1. Gender breakdown over time

*(to come)*

During the period under review, the number of comments per article rose year on year: in 2007, an average of 30 comments were left under each article; by 2015, 199. Overall the volume of comments increased dramatically – from under one million in in 2007 to 18 million in 2015 – and some individual articles attracted more than 2,000 comments. Although the moderation team also grew during this period (there were no dedicated moderators before 2007[xiii]; by 2015 there were 13 FTE moderators) it is clear that the volume of comments outstripped their capacity to effectively moderate them.

A small proportion of comments were blocked overall: 1.77% on average. However, the proportion of blocked comments increased over time, with a sharp rise

between 2009 and 2011 (from under 0.5% to 2.4%). This supports previous research showing that the tone of comments generally was becoming less civil (Anderson et al, 2014).

The proportion of comments blocked varied significantly between threads. Although the underlying trends – more comments, and a higher proportion of blocked comments – were true for articles written by both genders, articles written by women consistently attracted a significantly higher proportion of blocked comments (2.16%) than articles by men (1.62%). This gender imbalance in the block rate varied slightly year to year, but was always significant. This finding adds to previous scholarship by giving quantitative evidence of the gendered nature of online hostility.

Figure 2. Block rate by gender, year by year

*[to come]*

The data did not include the race, ethnicity or religion of the writers. However, as we have seen, the Guardian's BAME writers reported that they got more abuse than their white counterparts. This research supported this, even though it was not designed to test it.

Of the ten regular writers on Comment is free[xiv] who had the highest proportion of deleted comments, eight were women (four white, four BAME) and the other two were black men. Two of the women and one of the men were gay. And of the eight women in the "top ten", one was Muslim and one Jewish. And to drive this unexpectedly stark finding home: the ten regular Comment is free writers with the lowest proportion of deleted comments were all men. This disparity was particularly stark given that, in the period under review, only 28% of all writers on Comment is free

were women. The proportion of BAME and LGBTQ writers is not recorded, but make up a tiny proportion of the whole.

The findings were further affected by the section in which the story was placed. Although there were fewer women writing than men on the site overall, some sections were more male-dominated than others: Sport had the smallest proportion of articles written by women, but Technology and Film also had a pronounced gender gap; a few sections had slightly more women than men writing (Money; Life and Style), but the only subject section that had significantly more articles written by women was Fashion.

Chart 1: Gender gap and block rate by section

| **Section** | **Articles by women** | **Overall block rate** | **Female block rate** | **Male block rate** |
|---|---|---|---|---|
| **Fashion** | 69.72% | 3.74% | 3.67% | 3.98% |
| **Life & Style** | 54.22% | 2.45% | 3.01% | 1.41% |
| **Money** | 53.70% | 0.91% | 0.94% | 0.89% |
| **World news** | 27.11% | 2.85% | 3.05% | 2.77% |
| **Technology** | 13.37% | 0.91% | 1.69% | 0.85% |
| **Film** | 12.48% | 1.60% | 2.20% | 1.48% |
| **Sport** | 3.95% | 0.80% | 2.06% | 0.75% |

The gender imbalance in blocked comments increases in those sections where the writers' gender gap was significantly bigger than average (Sport, Technology and Film). In those sections where the writers were more gender-balanced, such as Money,

the block rate tended to be more balanced too – though Life and Style bucked that trend[xv], with articles written by women getting many more blocked comments than articles written by men. The *only* subject section of the site where articles written by men got a significantly higher block rate than those written by women was Fashion – also the only section on the site where there were many more women writing than men.

This relationship between the gender imbalance in each section and the gender imbalance in block rates gives quantitative support to the argument that women are more likely to experience abuse when they are perceived to be intruding on "male" spaces, and that abuse can therefore be seen as a reaction to a perceived loss of power (Shepherd et al 2015; Sierra 2014; Beard 2017, Mantilla 2015). The outlier – Fashion – is interesting: it may be that men are seen as the intruders in this "female" space, or perhaps men who write for fashion are targeted for being "unmasculine". The sentiment analysis needed to explore this further was beyond the scope of this research.

The data also included a "subject tag" for every article. Articles by women attracted more blocked comments regardless of the subject: of the 745 most used subject tags (all had more than 1,000 articles), only 14 showed no significant difference in the block rate for men and women writers (four of these subjects, incredibly, had no articles written by women, so there was no difference to measure; in several of the remaining 10, too few women had written articles to produce a statistically significant result).

Where a subject had a block rate of more than 4% overall, we labelled it as "sensitive". Looking at subjects where more than 1,000 articles had been written, and more than 1,000 comments left, I produced a list of the 27 most "sensitive" subjects. In all but five of these, articles written by women had a higher block rate than those written by men, despite the fact that 20 of these "sensitive" subjects were gender-neutral. For example, articles about South Africa had a differential block rate of 5.27%

(women) and 3.53% (men) and articles about the Leveson Inquiry had a differential block rate of 4.78% (women) and 3.84% (men). So when an article is already likely to attract a disproportionate number of blocked comments as a result of its subject, the gender of the author appears to act as a second independent variable.

The remaining seven sensitive subjects were directly about gender, race or sexuality – "Rape"; "Feminism"; "Race issues"; "Women"; "Gender"; "Sexuality"; "LGBT rights" – suggesting that subjects that were inherently gendered or racialised were more likely to attract comments that were subsequently blocked. In all these subjects, however, articles by women still got more blocked comments.

To summarise: by using blocked comments as a proxy for abusive or dismissive comments, I found that articles written by women attracted a significantly higher percentage of comments that were subsequently blocked than those written by men, regardless of the subject of the article. This effect was heightened when the articles ran in a particularly male-dominated section of the site. I also found evidence that articles written by BAME writers attracted disproportionate levels of blocked comments, even though the research was not designed to reveal this.

***The survey: 'Some hate it, and it's hard to blame them'***

The survey of Guardian journalists provides an insight into how journalists were affected by comments, both on the site and on social media (for those targeted, it is hard to separate the two, as it can feel like the attack is coming from everywhere – see Citron, 2014, for a discussion of this). Of the journalists who responded to the survey, 80% said they had experienced comments on the Guardian site or on social media which they felt "went beyond acceptable criticism of their work to become abusive". This was slightly higher for women (85%) and BAME respondents (83%) than for men (78%). Older journalists were less likely to experience it than younger ones (70% of

those over 55 had experienced it, compared to 89% of those aged 45-55). Opinion writers were the most likely to experience abuse, with 90% of those who write for Comment and Debate saying they had.

Of those who reported having abusive comments, 58% said it had happened more than 20 times and 29% more than 100 times. On average, this had happened to journalists 49 times. The frequency was similar for both men and women. 34% said they had received repeated criticism that left them feeling harassed, mobbed, or bullied in the comment threads. Here, age was a significant factor: this has happened to 50% of those aged 25-34, compared to 20% of those aged 55+.

As one female journalist put it:

> Imagine going to work every day and walking through a gauntlet of 100 people saying, 'You're stupid', 'You suck', 'I can't believe you get paid for this'. It's a terrible way to go to work.'

Ridicule was by far the most common form of abuse, with 84% of journalists saying they had experienced this – with no significant difference between men and women. However, whereas 57% of the women had received abusive comments that focused on their body, private life or sexuality, only 17% of male journalists had experienced this. Only 11% of the respondents overall had experienced comments that focused on their race, ethnicity or religion, but all five of the Black respondents said they had experienced this, as did both Muslim respondents and over half the Jewish respondents.

Of those who said they had experienced abuse, 68% felt angry as a result, 43% felt depressed, and 37% said they had anxiety symptoms. As one gay Black journalist said:

> Even if I tell myself that somebody calling me a nigger or a faggot doesn't mean anything […] it has an emotional effect, it takes a physical toll.

The experience had led to behavioural change: 53% had stopped reading comments, 33% said they stayed away from public debate, and 14% had seriously considered leaving journalism. Reader abuse was also seen as having a chilling effect on journalism itself – a quarter had subdued or changed angles in stories, and 20% had refused assignments as a result of abuse. Only 27% said reader abuse hadn't affected their journalism in any way.

Overall, the picture that emerged from this comprehensive survey of the newsroom was far from the idealised vision of a mutually beneficial conversation between journalists and their readers: most journalists were affected, and young, female and BAME journalists were affected disproportionately.

***The "below the line" response to the preliminary findings***

2458 comments were left in the thread below the preliminary findings. 130 of these had been left by moderators or other Guardian staff, and were not included in the sentiment analysis. Half the comments (1,235, or 50.24%) left below the published research results were coded as negative; 294 (11.96%) were positive; 799 (32.51%) were neutral.

212 (17.17%) of the negative comments criticised the research methodology[xvi], mainly on one of two grounds: either claiming that the research had failed to take moderator bias into account (this is discussed below), or that it had failed to consider the quality of the articles (for example, they said that articles by women may be more "worthy of complaint"). The study did not control for article quality, but assumed that, taken as a whole, articles written by women are not of poorer quality or otherwise more "deserving" of abusive or dismissive responses than articles written by men. In the author's view, this "methodological" criticism is an implicit form of victim-blaming.

A further 277 (22.43%) of the negative comments were overtly victim-blaming. Some asserted that female and/or black journalists in general were more likely to write poor quality or controversial articles – for example, "I would hardly say that all woman writers write daft things. But a lot of them do", or "Male author: Neutral / economic / sport / war / politics (general) articles; Female author: More click-bait / anti-male articles / feminist articles". Others blamed individuals for the abuse they received – for example, "Thrasher[xvii] gets negative feedback because he racebaits, not because he's black". These commenters failed to engage with the finding that the gender disparity was not confined to a few individuals, but was seen across the entire corpus, or that articles written by women got more blocked comments regardless of the subject they were writing about, and that this proportion *increases* when they write on subjects traditionally regarded as "male".

418 (33.85%) of the negative comments to the preliminary research findings attacked the moderators. Many suggested there was a "politically correct" bias which made reasonable comments liable to "censorship". For example: "[…] on race and gender articles […] certain ideas and even facts are deemed taboo."

I was unable to control for moderator bias in this study. However, it is unlikely to explain the overall findings, given the volume of comments, the consistency of the data, the training of the moderators, and the fact that most blocked comments came to the moderators' attention because readers had flagged them.

Some commenters argued that all moderation was a de facto attack on free speech, and what the Guardian sees as self-evident – that commenters should abide by the community standards – was far from being universally accepted. This points to a fundamental breakdown between the assumptions of the Guardian and a significant cohort of its commenters, and will complicate any attempt to manage comments.

This lack of shared assumptions is also evident in the 193 (15.63%) of the negative comments that criticised the Guardian's editorial stance more generally, such as this:

> Seriously, we all know what sort of Internet the Guardian wants. One where a feminist writer can write what she wants with no critical feedback unless it's in the affirmative, one where a black writer can play the race card with no negative feedback and one where even a slight inoffensive joke about the LGBT community is considered hate speech. […].

Most commenters on the Guardian are men[xviii], but the gender of individual commenters is not known. However, many of the negative comments were gendered male – that is, "they present themselves as conventionally masculine" in socially constructed ways (Phillips, 2015:42). Well-known tropes of men's rights activism (MRA) were evident – for example, that male freedom of speech is being curtailed, and inconvenient facts suppressed, by feminists – and many comments positioned the commenter as one of an "us" pitted against a feminist/black/LGBT "them". In this way, many of the negative comments simultaneously "do and deny sexism", in the same way (and often using the same arguments) as Benton-Greig et al (2017) found in their study of responses to a feminist campaign in New Zealand. And in these Guardian comments, some commenters do and deny racism too.

However, it is important to note that some commenters' hostility towards the Guardian's editorial stance was not gendered or racialized, but was expressed in class or anti-elitist terms. Some examples include:

> MPs' and columnists' vilifications of the powerless are reported without question, it is the powerless' vilification of MPs and columnists which is subject to questioning ….

> The Guardian despises ordinary people […] and wants to erase them from the debate. They want to sanitise and civilise until only approved opinions by London-based middle-ranking professionals with impeccable manners remain.
>
> […] maybe you'd be better off asking why it is that the comments are so displeasing […] Maybe the fact that many Guardian journalists are now utterly out of alignment with a great many of their own readership has something to do with it.

In her exploration of the relationship of interactivity to the organisation of social power, Kylie Jarrett (2008) argues that, far from being a genuinely democratising experience, "[t]he interactive user … encounters their own absence of agency and freedom in the free expression … offered to them within Web 2.0 sites". However, despite this consolidation of pre-existing power relations, there may be moments "when the control of capital and of elites has slipped" (Phillips, 2018, forthcoming). This is perhaps what is on display here. What the Guardian regarded as "abuse" below the line was seen by some commenters as a response to the elitism of individual journalists or of the Guardian as institution. And whereas the Guardian saw the comment threads as open to those who want to participate on their terms, some commenters saw them as a "public space", which the Guardian has no right to "police" – a perception which the Guardian arguably fostered by launching the comment space as a democratic forum that would "mutualise" journalism.

What is striking, however, is how often this feeling of powerlessness in relation to media elitism was expressed in anti-feminist and/or racist terms, or as a disavowal of the existence of sexism and racism. For some, it is as if it is black and female journalists who wield media power, and white men who are cowed, despite all the evidence to the contrary.

## Conclusion

By analysing block rates in 70m comments left below Guardian articles, using blocked comments as a proxy for abusive or dismissive comments, this study extends previous scholarship by providing quantitative evidence, which was not reliant on self-reporting, to support what journalists had long complained of: that articles written by women and people of colour attract a disproportionate amount of abusive and dismissive comments, regardless of what they were writing about.

The Guardian is typical of UK and US media in publishing fewer articles by women than men. In this research, there was evidence that the more male-dominated the section, the more blocked comments the articles by women got. This suggests that the lack of diversity within the media impacts on the quality of online conversations. The results of the survey show that the abusive comments female journalists receive are more personal in nature, and that women are more likely to change their behaviour as a result of them. Thus both employment practices and hostile comments combine to keep women's voices out of the media. This silencing of women's voices is repeated throughout public life, and this research supports the view that it may be both a cause and an effect of an environment in which abuse flourishes (Beard, 2017).

Guardian commenters, the subject of this research, responded to it in ways that exposed a seam of hostility not only to individual journalists, moderators and the Guardian as a whole, but to the "liberal, metropolitan, media class" in general. Sometimes, though by no means always, this anti-elitist hostility was expressed in sexist and racist terms – as if it was female and black journalists who wielded media power, and white men who were left powerless.

It would be easy to dismiss this anti-elitist strand of hostility given the current focus on sexism and racism. However, as well as being majority male and white,

Guardian journalists are also predominately from privileged class backgrounds, and this too is typical of the wider industry: according to the British government's 2012 report on social mobility, journalism is more "dominated by a social elite" than politics, law, or any other profession. Because of this, and because comments do challenge journalists' hegemonic perspectives (Milioni et al, 2012; Craft et al, 2015; Semmler & Semmler, 2016), this form of hostility should be of interest to anyone concerned with fostering wider participation in the media, and is an area that deserves further study.

It also complicates the idea of the Internet as democratising – old inequalities and hostilities do not disappear in this new digital space. Women and people of colour face the same sexism and racism here as they do in the physical world, with the same consequences. And arguably, class inequalities remain too: comments are left in their millions, a proportion of which challenge journalistic hegemony and media power – but is anyone listening to what they say?

Since the publication of the preliminary research on the Guardian's site, the organisation has made a number of changes. Overall, the main thrust of the institutional response has been on how, through more effective governance and the creation of new technological tools, abuse can be more efficiently prevented or blocked. Fewer articles are now opened to comments, and the moderation team has been strengthened; moderators have been given new tools, including a machine learning tool which flags potentially abusive comments; the Guardian is also exploring how to best authenticate commenters' online identities, and prevent banned users from returning.

Action has also been taken to make moderation decisions as consistent and transparent as possible; moderators now select a reason for blocking a comment from a drop-down menu, and their decisions are spot-checked. There are plans to email this

reason to commenters if their comment is blocked in order to increase awareness of the rules governing the space.

In addition, staff have been given written guidance on digital safety, and steps are being taken to introduce stronger institutional support for those who do experience abuse online.

All this to be commended, but some real challenges remain. First, the tension between the editorial and commercial motives for inviting reader participation persists: if comments are to have editorial value, journalists need to be able to surface the useful comments and enter into genuine dialogue with readers; but if they are to have commercial value, comments need to be high in volume, making this impossible. In a crowded comment space, individuals have to shout to be heard, or even to exist as online subjects (Shepherd, 2015). In this way, high volume threads encourage antagonistic behaviour. The desire to "monetise" comments may be incompatable with both the desire to deal with online hatred and misogyny, and the desire to democratise journalism.

Moderation is not endlessly scaleable, and although technologies (better filters, machine learning tools and so on) will be an important part of the solution, they will not be enough. What is needed is a change of culture. If comment threads are to be diverse and inclusive, media organisations need to create small, curated comment spaces where journalists can genuinely engage with what is said, even when it is critical[xix]; they will also need to develop anti-racist and feminist strategies to counter racist and sexist speech, and offer stronger institutional support to journalists and others who do experience this.

Secondly, this research indicates that the hostility to women and people of colour below the line mirrors a historical institutional hostility to women and people of

colour "above the line" – the discriminatory hiring and commissioning practices over many decades that have left them struggling to get published at all. It is encouraging that the new editor-in-chief has stated her commitment to publishing journalists from more diverse backgrounds[xx]. We must not let the very real problem of online abuse divert attention away from other more systemic forms of sexism, racism and class discrimination, which this research suggests may be the soil in which that abuse grows.

Acknowledgments

The analysis of the comment data would have been impossible without the collaboration of the Guardian's then Head of Data Science, Mahana Mansfield. I would also like to thank Katharine Viner, who initiated and supported this project, and gave me access to the data, Helena Bengtsson, who helped design the analysis, Chay Woodford, who checked the results and created the graphs, Meghan McCarthy, who helped design the staff survey, and to all the Guardian staff who gave generously of their time. Annie Kelly coded the readers' comments. Thanks also to Angela Phillips, Des Freedman, Mirca Madianou and Lisa Blackman, for their advice and support, and to my reviewers for their invaluable feedback.

---

[i] 'Below the line' refers to the comment thread; articles are sometimes referred to as being 'Above the line'.

[ii] https://www.theguardian.com/media/2016/jul/12/how-technology-disrupted-the-truth

[iii] The US National Violence Against Women Survey estimated that 60% of "cyber-stalking victims" were women; Working to Halt Online Abuse found that 72.5% of the 3,000 reports of "cyber harassment" were from females; The Pew Research Center's 2014 report found that women were more likely to experience sexual harassment and stalking (all cited in Citron, 2014)

[iv] https://www.theguardian.com/community-standards (accessed 5 December 2017)

[v] The data for 2016 was not for a full year, so the findings given here are for the years 2006-2015, unless otherwise stated. Only comments made on the Guardian's website are included, not those left on Facebook or other social platforms.

[vi] This classification was done using genderize.io, which predicts gender based on first names. This is the best currently available approach (Wais, K, 2016). The 746 names that could not be classified using this method were classified manually by the author, using the individual's Guardian profile, or google search.

[vii] In a tiny minority of cases moderators either "watchlisted" (actively monitored) or pre-moderated a thread (if, for example, the author was seen as vulnerable).

[viii] Several Guardian journalists were subject to such abuse on social media.

[ix] Sent in February 2016.

[x] Three people did not record their gender.

[xi] 41.5% of these were regular Guardian readers, ie they had visited the site at least eight times with a gap of no more than a week; 68% were from the UK or US.

[xii] The Women's Media Center releases an annual report on the status of women in the US media: in 2017, they found that 62% of bylines across all media are male, while only 38% are female[xii]. A 2011 Women in Journalism survey of bylines in five national papers recorded a 78:22 split in favour of men[xii].

[xiii] Before 2007, ad hoc moderation would have been carried out by desk editors and the community team.

[xiv] A "regular" writer is defined as someone who has written more than 100 times for Comment is free. Comment is free was the Guardian's online opinion section.

[xv] It is possible that this is because Life & Style is a large and rather broad section, including a lot of disparate "light" subjects such as Family, Food and Drink, Health and Fitness etc.

[xvi] There were also a number of criticisms of the methodology which were expressed in a polite or neutral tone (eg as a straightforward question) or which were valid: these were labelled neutral, not hostile.

[xvii] Stephen Thrasher, a US-based, gay African American columnist.

[xviii] Internal research in 2016 found that 69% of those who have posted comments are male, compared to 31% who are female, a commenter profile in line with that of other news sites (see Pierson, 2015; Martin 2015).

[xix] For clarity, I do not mean that journalists should engage with abusive speech, which can cause real harm.

[xx] See her article, "A Mission for Journalism in a time of crisis" (The Guardian, 16 November 2017) https://www.theguardian.com/news/2017/nov/16/a-mission-for-journalism-in-a-time-of-crisis.

**References**


Adams, T.E., Holman Jones, S., Ellis, C. (2015) *Autoethnography:Understanding qualitative research*. New York, NY: Oxford University Press.

Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A. and Ladwig, P. 2014. The "Nasty Effect:" Online Incivility and Risk Perceptions of Emerging Technologies. J Comput-Mediat Comm, 19: 373–387.

Beard, Mary. 2017. "Women in Power". *London Review of Books*: 39 (6). Retrieved at: https://www.lrb.co.uk/v39/n06/mary-beard/women-in-power

Benton-Greig, P., Gamage, D., & Gavey, N. 2017. Doing and denying sexism: online responses to a New Zealand feminist campaign against sexist advertising, Feminist Media Studies, DOI: 10.1080/14680777.2017.1367703

Bergström, Annika, and Ingela Wadbring. 2015. “Beneficial yet crappy: Journalists and audiences on obstacles and opportunities in reader comments”. European Journal of Communication 30 (2): 137-51.

Bruns, A. (2007). Produsage: Towards a Broader Framework for User-led Content Creation. In Proceedings of the 6th ACM SIGCHI Conference on Creativity & Cognition. New York: ACM.

Bungawala, Inayat and Huma Qureshi, Simon Woolley, Nadiya Takolia, Bim Adewunmi. 2012. “Online racist abuse: we’ve all suffered it too.” https://www.theguardian.com/commentisfree/2012/jul/11/online-racist-abuse-writers-face (accessed 11 June 2017)

Cheng, J., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2015, April). Antisocial Behavior in Online Discussion Communities. In *ICWSM* (pp. 61-70)

Citron, D.K. 2014. *Hate Crimes in Cyberspace*. Cambridge MA: Harvard University Press.

Coe, K., Kenski, K. and Rains, S. A. (2014), Online and Uncivil? Patterns and Determinants of Incivility in Newspaper Website Comments. J Commun, 64(4): 658–679

Stephanie Craft, Tim P Vos, J David Wolfgang (2015) “Readers comments as press criticism: Implications for the journalistic field”. (2015) Journalism, 1-17

Curran, J., Fenton, N., & Freedman, D. (2012). *Misunderstanding the Internet*. London: Routledge.

Elliott, Cath. 2007. Speaking truth to power. *Guardian* https://www.theguardian.com/commentisfree/2007/nov/28/speakingtruthtopower (accessed 2 June 2017)

Fraser, N. (1992) 'Rethinking the Public Sphere: a Contribution to the Critique of Actually Existing Democracy', in C. Calhoun (ed.) Habermas and the Public Sphere, pp. 109–42. Cambridge, MA: MIT Press

Gardiner, Becky, and Mahana Mansfield, Ian Anderson, Josh Holder, Daan Louter and Monica Ulmanu. 2016. "The dark side of Guardian comments." *Guardian*. https://www.theguardian.com/technology/2016/apr/12/the-dark-side-of-guardian-comments (accessed 11 June 2017)

Henry, Georgina (2006) "Welcome to Comment is Free". *Guardian*. https://www.theguardian.com/commentisfree/2006/mar/14/welcometocommentisfree (accessed 2 June 2017)

Gillmor, Dan. 2004. *We the Media, Grassroots Journalism by the People, for the People.* Sepastopol, CA: O'Reilly Media

Graham, Todd and Scott Wright . 2015. "A Tale of Two Stories from 'Below the Line': Comment Fields at the Guardian". The International Journal of Press/Politics 20 (3): 317-338.

Hasan, Mehdi. 2012. "We mustn't allow Muslims in public life to be silenced." *Guardian*. https://www.theguardian.com/commentisfree/2012/jul/08/muslims-public-life-abuse?commentpage=all#start-of-comments (accessed 11 June 2017)

Hermida, Alfred and Neil Thurman. 2008. “A Clash of Cultures: The Integration of User-Generated Content within Professional Journalistic Frameworks at British Newspaper Websites.” Journalism Practice 2 (3):343-56.

Jane, E. A. (2014) “Your a Ugly, Whorish, Slut”, Feminist Media Studies, 14:4, 531-546

Jarrett, K (2008) Interactivity is Evil! A critical investigation of Web 2.0. *First Monday, 13 (3).* Retrieved from http://firstmonday.org/article/view/2140/1947

Jarvis, Jeff. 2006. “Networked Journalism”. *BuzzMachine* http://buzzmachine.com/2006/07/05/networked-journalism/ (accessed 2 June 2017)

Jönsson, Anna Maria, and Henrik Örnebring. 2011. ”User-Generated Content and the News: Empowerment of Citizens or Interactive Illusion?” Journalism Practice 5 (2): 127-44.

Lewis, Helen. 2013. “You should have your tongue ripped out: the reality of sexist abuse online”. *New Statesman* http://www.newstatesman.com/blogs/helen-lewis-hasteley/2011/11/comments-rape-abuse-women (accessed 2 June 2017)

Mantilla, K. 2015. *Gendertrolling: How Misogyny Went Viral.* Santa Barbara, California: Praeger

Martin, F. 2015. Getting my two cents worth in: Access, interaction, participation and social inclusion in online news commenting. #ISOJ Journal (5), 1.

McKay-Semmler, K., Semmler, S. 2016. “The Unfinished work of feminism: An analysis of online user generated comments responding to public opinion polls about gender equality.” In *Gender and Work: exploring intersectionalisty, resistance, and identity,* eds Miglena Sternadori and Carrie Prentice, 111-126. Newcastle upon tyne. Cambridge Scholars Publishing.

Milioni, D L, Vadratsikas, K & Papa, V. (2012). “Their two cents worth”: exploring user agency in readers’ comments in online news media. *Observatorio, 6*(3), 21-47.

Muddiman, A. and Stroud, N. J. (2017), News Values, Cognitive Biases, and Partisan Incivility in Comment Sections. Journal of Communication, 67(4), 586-609

Papacharissi, Z. (2004). Democracy online: civility, politeness, and the democratic potential of online political discussion groups. New Media & Society, 6(2), 259–283.

Penny, L. 2011. “A Woman’s opinion is the mini-skirt of the internet”. Independent. http://www.independent.co.uk/voices/commentators/laurie-penny-a-womans-opinion-is-the-mini-skirt-of-the-internet-6256946.html (accessed 2 June 2017)

Phillips, Angela. 2018. The Technology of Journalism, in Vos, Tim P. ed, (2018) *Handbooks of Communication Science: Journalism*. De Gruyte (forthcoming)

Phillips, W. 2015. This is Why We Can't Have Nice Things: Mapping the Relationship Between Online Trolling and Mainstream Culture. Cambridge, MA: MIT Press.

Emma Pierson. 2015. Outnumbered but Well-Spoken: Female Commenters in the New York Times. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW).

Reagle, J.M. 2015. *Reading the Comments: Likers, Haters, and Manipulators at the Bottom of the Web.* Cambridge, Mass: MIT Press

Rosen, Jay. 2006. The People Formerly Known as the Audience. *PressThink*. http://archive.pressthink.org/2006/06/27/ppl_frmr.html (accessed 2 June 2017)

Ruiz, Carlos, David Domingo, Josep Lluı´s Mico´ , Javier Dı´az-Noci, Koldo Meso, and Pere Masip. 2011. "Public Sphere 2.0? the Democratic Qualities of Citizen Debates in Online Newspapers." International Journal of Press/Politics 16 (4): 463–487.

Rusbridger, A (2010) "The Hugh Cudlipp Lecture: Does Journalism exist?" *Guardian* https://www.theguardian.com/media/2010/jan/25/cudlipp-lecture-alan-rusbridger (accessed 2 June 2017)

Shepherd, T, and Alison Harvey, Tim Jordan, Sam Srauy and Kate Miltner. (2015) "Histories of hating". Social Media + Society 1 (10)

Shirky, Clay. 2009. *Here Comes Everybody: How Change Happens When People Come Together* Penguin UK.

Sierra, Kathy. 2014. "Trouble at the Koolaid Point" *SeriousPony* http://seriouspony.com/trouble-at-the-koolaid-point/ (accessed 2 June 2017)

Singer, Jane B, and Ashman, I. 2009. "'Comment Is Free, but Facts Are Sacred"" User Generated Content and Ethical Constructs at the Guardian." *Journal of Mass Media Ethics* 24 (1): 3-21.

Sood, S., Antin, J., and Churchill, E. 2012. Profanity use in online communities. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12). ACM, New York, NY, USA, 1481-1490.

Thorpe, V., and Rogers, R. 2011. "Women bloggers call for a stop to 'hateful' trolling by misogynist men." *Observer*. https://www.theguardian.com/world/2011/nov/05/women-bloggers-hateful-trolling?CMP=twt_gu (accessed 2 June 2017)

Viner, K. 2013. "The rise of the reader: Journalism in the age of the open web" *Guardian*. https://www.theguardian.com/commentisfree/2013/oct/09/the-rise-of-the-reader-katharine-viner-an-smith-lecture (Accessed 2 June 2017)

Viscovi, D., and Gustafsson, M. 2013. "Dirty work: why journalists shun reader comments." In *Producing the Internet: Critical Perspectives of Social Media*, ed Tobias Olsson, 85-101. Goteborg: Nordicom.

Wais, K. 2016. Gender Prediction Methods Based on First Names with genderizeR https://journal.r-project.org/archive/2016/RJ-2016-002/RJ-2016-002.pdf (accessed 1 Nov 2017)

Younge, Gary. 2012. "Who thinks about the consequences of online racism?". *Guardian*.https://www.theguardian.com/commentisfree/2012/jul/12/consequences-of-online-racism (Accessed 11 June 2017).