

---

# **‘It’s Twitter, a bear pit, not a debating society’**

## **A qualitative analysis of contrasting attitudes towards social media blocklists**

**Dawn Wheatley, PhD\***

*School of Communications, Dublin City University*

**Eirik Vatnoey, PhD.**

*Department of Linguistics and Scandinavian Studies, University of Oslo*

---

### **Abstract**

This study of tweets ( $n = 2247$ ) explores discussions about a pro-choice blocklist (@Repeal\_Shield) used during the 2018 Irish abortion referendum campaign, capturing conflicting interpretations of engagement and political participation. Although qualitative Twitter studies bring methodological challenges, deep readings were needed to analyse arguments in favour and against the blocklist, and to consider what we can learn about users’ expectations of Twitter. Through deductive and inductive coding, opposing perspectives emerge on whether such lists are useful, democratic or regressive, but both sides share normative aspirations for Twitter to serve as a space for healthy debate, even if there is clear tension in how that is best achieved. Blocklists are traditionally cited as a harassment solution, facilitating participation from otherwise-excluded counterpublics. However, @Repeal\_Shield demonstrates how this affordance has evolved towards omitting broad spectrums of undesired content, while using blocklists – and being listed – can be a bold political statement in itself.

### **Keywords**

Blockbots, blocklists, counterpublics, harassment, rhetoric, social media, Twitter

\*Corresponding author: dawn.wheatley@dcu.ie

*Please note: this is a pre-print version of an article published in New Media & Society:*

<https://journals.sagepub.com/doi/10.1177/1461444819858278>

## Introduction

Social networks' role in changing the underlying conditions of political participation has received significant scholarly attention and garnered mixed conclusions (Castells, 2010; Couldry, 2015; van Dijck, 2006). Some argue that 'networked individualism' is empowering users (Rainie and Wellman, 2012), enforcing new forms of self-identity (Papacharissi, 2010) and changing how people relate to political causes (Bennett & Segerberg, 2013). Conversely, sceptics warn against the perils of 'echo chambers', in which opinions are rarely confronted but always reinforced (Sunstein, 2009, 2017). For 12 years, Twitter has provided a space for conversation and establishing connections, but this has raised concerns about some users' behaviour. Founder Jack Dorsey (2018) suggested Twitter has become divisive and prone to abuse, and on 15 May 2018, the company announced changes to achieve more 'healthy conversations', with contributions from those engaging in detracting 'troll-like behaviours' to be excluded from search results (Harvey and Gasca, 2018). Elsewhere that month, a US judge ruled that Donald Trump violated Americans' constitutional rights by blocking Twitter accounts. While these developments were garnering attention from those interested in Twitter as a civic space, Irish citizens were preparing for a landmark referendum on 25 May 2018 to repeal the effective total outlawing of abortion. At first glance, these events may seem unconnected, but the online debate in Ireland included one element illustrating growing divisions among users' expectations and attitudes regarding the content they encounter online and the available interventions.

Four months before the referendum, volunteers established a blocklist, @Repeal\_Shield, to exclude 'pro-life trolls' and 'bot/sock puppets [and] people who've demonstrably no interest in civilised nor factual debates', according to an early Tweet outlining its mission (Repeal\_Shield, 2018). Any Twitter user subscribing to such blocklists does not see content from accounts added to the list by administrators. This study analyses the contrasting attitudes towards this Irish blocklist and, more broadly, explores what these exchanges reveal about user expectations regarding social media's role in political debate. Early observations of tweets discussing @Repeal\_Shield's presence illustrated some of the diverse perspectives: *'Repeal shield seems to mainly be an attempt by the repeal [pro-choice] supporters to avoid confronting dissenting opinions – I don't think that's good for healthy debate'*, compared with others highlighting its apparent necessity: *'Arguing online can be incredibly draining. The constant onslaught of bullies, trolls and misogynists can eat away at your mental health. This blocklist can help people to protect themselves from that.'* Stark contrasts and tensions are clear as expressions of personal choice collide with sentiments about the need for wide-ranging perspectives, thus bringing our attention to what arguments were being made in favour and against the blocklist, and what the discussions tell us about user attitudes towards social media as a site for public debate and political participation. The research lens remains focused on the cultures and attitudes *within* the Twitter sphere, rather than positioning Twitter's broader civic role; the primary purpose is to understand how users *interpret* one particular Twitter affordance, which facilitates a network of exchanges more fragmented and closed than the network itself perhaps envisioned. Such potentially unintended consequences highlight how novel media formats and networks enable many new

practices. Blocklists are just one such intervention warranting consideration; in-depth studies of such phenomena are needed in the piecemeal mapping of how social media influence changing modes of political participation. The analysis presented here is built on an original framework of four categories of rhetorical strategies found in the tweets which can inform our understanding of Twitter users' attitudes towards these seemingly contentious blocking tools.

### **Blocklists and Fragmenting Online Spheres**

For some, the presence of abusive or unsavoury content is a salient downside of social media use (Mantilla, 2013) with no clear solution. One response, Benesch explains, may come not by curbing the original speaker, but instead through actions aimed at '[alleviating] the harm felt by people at whom [harassment] is directly aimed'(2017: 32). She cites blocklists, or blockbots, as one measure, allowing those targeted to protect themselves by bypassing the problematic content in the first instance: Twitter users will not see content from any account included on the list(s) to which they subscribe. On its website, BlockTogether.org, a leading platform hosting blocklists, says it can 'reduce the burden of blocking when many accounts are attacking you, or when a few accounts are attacking many people in your community' (BlockTogether.org, 2018). This third-party activity makes the response to harassment 'more efficient and more communal' (Geiger, 2016: 788), as individual blocking is extended into a group activity; such communal action, Maréchal surmises, helps users 'regain control of their social media experience' (2016: 5023). These notions of power and agency are commonplace in commentary about blocklists, reasserting the individual's ability to shape their online experiences. Maréchal also notes how blocklists facilitate 'the right of users to exist online free from harassment, without actually preventing the harasser from speaking', therefore positioning such applications as compromises which avoid potential infringements of other Twitter users' freedom (2016: 5023).

Blocklists garner passing mentions in works considering responses to harassment, mostly associated with the 2014 Gamergate controversy during which females in the online gaming community were targeted (Barnes, 2018; Benesch, 2017; Geiger, 2016; Golbeck et al., 2017), but empirical studies are more scarce. Nevertheless, research carried out thus far captures much of the tools' complexity. Geiger's work, primarily theoretical, describes these lists as a 'novel, user-generated, bottom-up, computationally supported moderation work' evolving in response to what he calls Twitter's hands-off moderation approach (2016: 788). He draws on Fraser's (1990) 'counterpublics', describing how these tools facilitate inclusive spaces for typically non-dominant communities who may otherwise avoid participating. Crucially, Geiger suggests these public spaces do not need to exist parallel to dominant spaces, as Fraser described, but remain within the 'hegemonic discursive environment' (794). Maréchal echoes this counterpublic narrative, concluding that blocklists support free expression by allowing participation in mainstream discourse (2016: 5025).

These aggregatory or filtering tools, therefore, bring notions of authority and control to the fore. What action is being taken, by whom, with what intention, is raised by Caplan and boyd (2016). Although not writing explicitly about blocklists, they draw on power in network technologies, declaring that 'no intervention is without externalities', thus highlighting the

influence such tools have in shaping public online spaces (2016: 15). They add that with any such intervention, 'it is essential to recognise the biases and assumptions that underpin any ideal and critically interrogate who benefits and who does not' (2016: 15). While many commentators see blocklists as a liberating force for those facing harassment, social media users have mixed attitudes with respondents from the gaming community overwhelmingly negative about the Gamergate list, ggautoblocker (Gosse and O'Meara, 2018). Its use was referred to as one of three transgressions of some so-called rules of online debate: (i) a form of employment discrimination (being listed could affect gaming industry employment); (ii) a form of slander (people taking issue with being branded a harasser); and (iii) a form of censorship (hindering free expression and suppressing ideological difference). In this latter, most prevalent, category, concerns were raised about how lists limit the potential for productive debate, alongside suggestions that people should have access to online public spheres without 'having to bend to others' (2018: 9). The authors conclude that ggautoblocker was a 'complex moment in digital culture, one with contradictory implications for participation in the public sphere' (2018: 9).

This complexity is at the crux of such applications: social media users with contesting values appear at odds in a space where many rules are not formalised beyond any official positions of the social networks. Twitter's apparent lack of action tackling abuse is one driving force for blocklists suggested by Jhaver et al. (2018), who interviewed some of those using and blocked by the ggautoblock list. While some were indifferent about being listed, others were annoyed, a frustration highlighted elsewhere (Auerbach, 2015; Gosse and O'Meara, 2018). Shaw (2017) acknowledges that users feeling unfairly blocked is a shortcoming, but suggests this is ultimately a shortcoming of Twitter's affordances more generally, suggesting a potential lack of perceived due process or recourse for users feeling slighted. Elsewhere in Jhaver et al.'s study, respondents make assumptions about the listed users (that they might be bigoted in other ways), while some believed socially curated blocklists were more objective than those algorithmically driven (ggautoblocker was automated), and some refer to blocklists' effectiveness in 'taming' and making Twitter cleaner (2018: 21). Some acknowledged their own privileges as not being targets for abuse, but noted that others – women, ethnic or religious minorities, feminists, women of colour, LGBT users – may be unable to freely participate without filters. No interviewee who used blocklists appeared concerned that accounts whom they would not have ordinarily blocked may be filtered out. This mass exclusion has been acknowledged by ggautoblocker founder Randi Harper, who admits the application is more like a sledgehammer than a nuanced tool, but worthwhile to stop hearing from a vocal few hundred (Woffard, 2014).

Disagreement over what constitutes harassment is also raised by Jhaver et al, resonating with a concluding point from Geiger; he describes how blocklists are evolving from their original purpose of excluding harassment, increasingly used for activity 'better described as incivility or trolling' (2016: 799). Given the literature's emphasis on Gamergate, there is clearly a need to extend blocklist study to new cases, in particular political contexts. A general dislike of viewpoints, or political motivations, could be the basis for mass blocking, but this appears empirically unexplored. This is particularly pertinent in an era of heightened polarisation online where contesting viewpoints may be easily avoided (Heatherly et al.,

2017; Sunstein, 2017). The terminology of echo chambers in particular has gained prominence, tied to the customisability of online space and how this limits exposure to conflicting ideas, thus deepening political polarisation, but such insulation may be unsurprising: '[If you] have strong convictions, you might want support, reinforcement, and ammunition, not criticism' (Sunstein, 2017: 64). However, despite concerns regarding echo chambers, questions remain over their presence and impact, with some suggesting fears are overstated as online exposure to diverse, cross-cutting information sources prevails (Bakshy et al., 2015; Dubois and Blank, 2018).

Yet, elsewhere, so-called digital 'safe spaces' have also provided havens for those fearful of participating in certain public arenas (Clark-Parsons, 2018; Maliepaard, 2017). The term has evolved over the past two decades, particularly through association with feminist movements and minority groups, related to physical spaces of safety, free expression and psychological freedom (Djohari et al., 2018). Yet their production involves polarising binaries at the core, between safety/danger, inclusivity/exclusivity, and public/private (The Roestone Collective, 2014), and these potential schisms have arguably contributed to a so-called safe-space culture becoming the target of attack. Some interpret safe spaces' existence as a violation of free speech, with critics often evoking the term with mocking or pejorative intent (Nicholson, 2016). This has been particularly prevalent in the US (Lukianoff and Haidt, 2018), and although the situation is not quite transferable in Ireland, some of the language deriding these safe spaces has been adopted (Hickey, 2016; Nolan, 2017; Power, 2018). Nevertheless, the vernacular of 'safe spaces' prevails among many advocates.

As safe spaces can be defined as a space free from violence and harassment (The Roestone Collective, 2014), such discussions may draw attention to conceptualisations of harm. Writing from a criminology perspective, the dual categories put forward by Henry and Milovanovic is a conducive lens (Milovanovic, 2006): *harms of reduction* result when offended parties experience a loss of some quality relative to their present standing, while *harms of repression* occur when limits or restriction prevent people from achieving desired positions or standings. In social psychology, these can be linked to Maslow's hierarchy of needs (Lanier, 2014): physiological, safety, love/belonging, esteem, and self-actualisation, thus demonstrating how harm, real or perceived, can affect individuals physically and psychologically.

Overall, blocklists can alter Twitter's boundaries to create more closed, personal spaces than what the network initially conceived, perhaps conflicting with its diverse, 'healthy' conversation goals (Harvey and Gasca, 2018). Yet the tools appear to provide a unifying dimension against unwanted or even harmful behaviour. Geiger suggests blocklists go beyond their technological affordance and are, in fact, a community, contributing to 'collective sensemaking' (2016: 789) regarding shared values, identities, and ideas of what constitutes harassment. This is particularly pertinent in the context of social movements and political debate in campaigns where there are collective decisions at stake and, in such contexts, it is relevant to view the use and implications of blocklists in light of political participation. The debate about how new digital media platforms influence political participation is broad and ongoing. A central question in this debate is to what extent such platforms also affect the modes of participation in a rapidly changing digital age (Fox, 2014).

Van Deth (2014) has suggested a set of decision rules to distinguish both known and yet unknown variants of political participation: actions should be considered political participation if they can be considered an activity or behaviour, if they are done by people in their role as citizens, if they are voluntary, and if they are placed within the sphere of the political. These criteria allow us to consider managing and subscribing to blocklists as more or less specific forms of political participation. Furthermore, focusing on the debate surrounding a particular blocklist allows us to not only investigate the potential of such lists, but also how this potential is perceived by users, and how this in turn affects the nature of participation.

### **The Irish Abortion Referendum**

The 1983 Eighth Amendment in Ireland recognised 'the right to life of the unborn' (Government of Ireland, 1983), effectively outlawing terminations in all circumstances; the constitutional dimension ensured no government could legislate without a referendum. By its nature, the abortion debate is typically polarising and, in this instance, took place in a traditionally Roman Catholic society. Yet the image of Ireland as a bastion of conservative values no longer holds, as evidenced in the 2015 referendum legalising same-sex marriage. While traditionally liberal and conservative voters' abortion stances were as expected, there was a strong middle ground: for example, those who support abortion, albeit in restricted circumstances, but their referendum support was needed to accommodate any change. The referendum on 25 May 2018 passed by 66.4%.

The referendum was announced on 29 January 2018 and the following day, the @Repeal\_Shield account was registered on Twitter, first tweeting on 31 January. In an early post, @Repeal\_Shield outlined its intentions: '*There are better ways to campaign for Repeal than to argue with Pro-Life trolls...we believe it's more productive/helpful to engage with people who want a genuine discussion in good faith, as opposed to accounts seen to troll, bot/sock puppets [and] people who've demonstrably no interest in civilised nor factual debates*'. Over the following four months, it added approximately 17,000 names to the blocklist. Approximately 4,500 accounts followed the @Repeal\_Shield Twitter account, but it is understood that less than a quarter of that actually subscribed to the BlockTogether list. @Repeal\_Shield was run by a group of seven men and women who could be generally described as pro-choice activists and heavy Twitter users themselves. While some of those involved would be known to those familiar with the Irish Twitter landscape, the individuals were not public figures. However, regarding their role with Repeal Shield, they were named and identifiable, openly speaking of their involvement in the administration of the blocklist, promoting it as a type of non-partisan, voluntary initiative. These administrators made manual additions to the blocklist based on both their own observations and others' suggestions.

### **Methodology and Value of Qualitative Twitter Research**

The value of large-scale quantitative Twitter studies is well established, proving particularly effective in mapping networks, hashtags and topic analyses. Yet concerns and limitations of

big data approaches have also been raised (boyd and Crawford, 2012), and deconstructing certain Twitter activity requires recognition of each contribution's complexity. The public declaration of stances or opinions on Twitter can be considered performance elements of the social network (Papacharissi, 2012), relevant for rhetorical engagement, and this focus, therefore, ensured that qualitative, reflexive approaches were needed

Marwick suggests that while large-scale pattern identification is useful, 'it can also overlook *how* people do things with Twitter, *why* they do them, and how they *understand* them' (2014: 119). Conversely, she argues, a qualitative approach can assist with meaning-making, placing technology 'into specific social contexts, places, and times' (2014: 119). Murthy echoes this, highlighting the need for inductive methods which allow us to dig deeper into the culture of digital spaces, as such 'messy' contexts and questions require approaches beyond deductive reasoning (2017: 560). These possibilities for more interpretive understandings of Twitter culture provide a solid basis for studies in which attitudes towards an affordance of a specific social network are explored, frameworks developed, and user perspectives are complex and heterogeneous. It is, therefore, an appropriate approach to address the study's research questions: RQ1 – What arguments were made in favour of the blocklist? RQ2 – What arguments were made against the blocklist? RQ3 – What can the discussions tell us about user attitudes and expectations towards social media as sites for public debate and political participation?

### **Data Collection and Sample**

The initial dataset was gathered using NCapture, a Chrome browser add-on that saves content via the Twitter API, collecting tweets including the words 'repeal shield'. Although this may not capture all discussions about the blocklist, it ensured that all tweets collected referred to this case. Searches were saved, on average, every three days: between 31 January 2018 (the date the @Repeal\_Shield account was first active) and 26 May 2018 (the referendum count), 38,063 tweets were gathered. Retweets were subsequently excluded to avoid duplications, reducing the sample to 22,684. This needed to be further refined to focus on tweets in which the blocklist's concept was discussed. Due to the nature of the exchanges, lack of suitable keywords, and the desire to avoid random sampling for fear of missing relevant consecutive tweets in clusters of conversations, the 22,684 tweets were manually assessed to determine relevance. This involved one researcher reading the total database on a simple inclusion/exclusion basis. To clarify this selection process further, the included tweets affirmatively answered this question: is the user sharing a general opinion on the use/existence of the @Repeal\_Shield account? This reduced the sample to 2,247 tweets, illustrated in Figure 1, with the majority excluded because either: i) The @Repeal\_Shield account was only tagged as part of more general referendum discussions; ii) the words 'repeal' and 'shield' appeared in discussions regarding other countries' laws; iii) a user tagged @Repeal\_Shield to attract the administrators' attention; iv) the tweet related to whether specific named users should be on the blocklist.

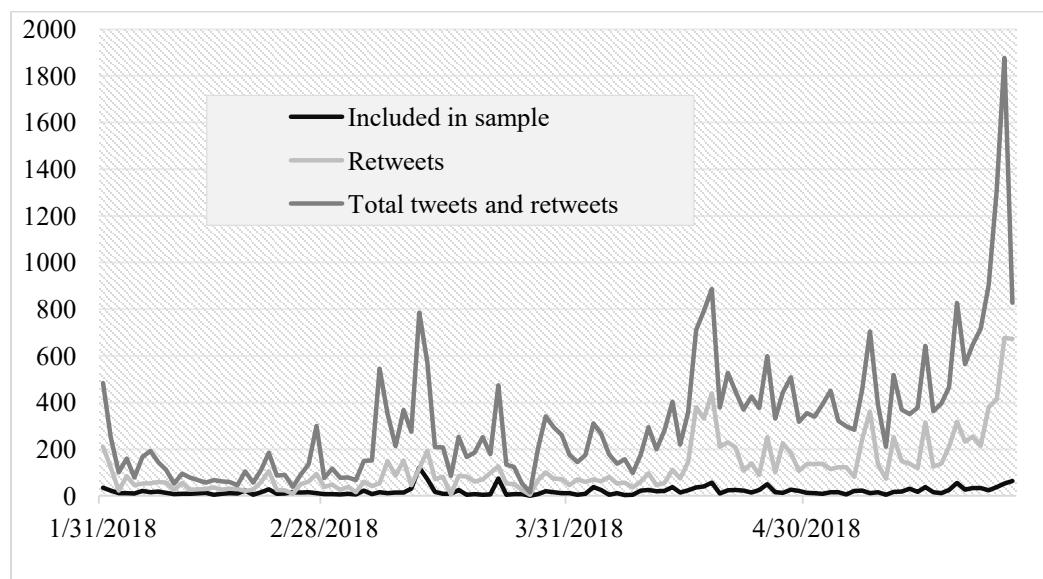


Figure 1: The total number of tweets gathered during the four-month period under observation ( $n=38,036$ ), also showing excluded retweets ( $n=15,379$ ), and the number included in the refined sample ( $n=2,247$ )

This smaller collection of tweets focused on discussions surrounding the merits, problems, and attitudes towards the blocklist's concept and implementation. Manual selection minimises representativeness concerns as the whole run-up to the referendum is covered. Furthermore, by resisting any temptation to focus on the most popular or influential users/tweets, the sample is not overshadowed by dominant users; the 2,247 tweets include contributions from 958 accounts; just 24 users had more than 10 tweets included, while 583 users had a single tweet in the sample. Notably, it is not this study's intention to make assessments about those blocked or those contributing to the discussions. Although bot accounts and co-ordinated tweeting is topical in online political campaigning, it was beyond our scope to draw conclusions about any user profiles. Furthermore, based on our in-depth reading, we do not believe the data is characterised by these accounts and do not factor this into the analysis.

### **Analysing the Material and Drawing up a Framework**

This smaller set of tweets underwent various stages of sorting and coding, outlined in Table 1, with all reading done manually, allowing for revealing interpretations of the content's performative elements (Papacharissi, 2012). The overall methodological approach was heavily influenced by Murthy's (2017) argument that qualitative Twitter approaches should maintain a clear focus through categorisation and solid coding schemes, while being open to reflexive amendment as the analysis develops. Coding in some Twitter studies has been informed by, and further builds on, patterns or frames of previous research (Vis, 2013). However, given the lack of blocklists research, the categories and themes initially had to be



generated inductively to start structuring the analytical framework, before refining and deductive sorting.

<b>Stage</b>	<b>Approach</b>	<b>N=</b>	<b>Description</b>
<b>(Pre-coding)</b>	Sorting for inclusion/exclusion	22,684	All tweets were read. Notes were taken but no formal categories created
<b>Stage 1</b>	Inductive	2,247	Tweets were read and coded as either an argument in favour or against RS, or as relevant to general social media use. Notes were taken and general themes identified
<b>Stage 2</b>	Retroductive	2,247	Tweets were re-read and examples for each theme/category found, and new categories created
<b>Stage 3</b>	Inductive & Deductive	2,247	Existing categories were grouped together by topic. Categories were further analysed and refined into four rhetorical strategies
<b>Stage 4</b>	Deductive	1,000	Sub-sample of tweets were coded into the rhetorical strategies and subcategories

*Table 1: The coding process including inductive, retroductive, and deductive approaches*

Throughout each stage, patterns and themes were identified with continued memo-taking. Central to stage 2, Murthy has put forward the notion of the 'retroductive' researcher, who 'probes the data regularly' (2017: 563) with some focus in what they are looking for, rather than purely inductive approaches. He cites Poole, describing how retroduction allows us to be stopped by a surprise and then try to comprehend it, highlighting an openness towards the data (2015, in Murthy 2017, p.563). Tandoc et al's (2018) method was also useful: in their study, two researchers independently carried out multiple rounds of reading a dataset (n=2,501), comparable in size to that in use here, to develop specific categories before comparing categories, discussing parameters, and integrating and re-sorting categories. A similar process was used here, with various stages and categorisations compiled, further refined and resolved through deliberation. Tandoc et al used this approach to create a conceptual model which resonated, as an original framework to capture various rhetorical activities was one of this study's aims, necessitating an open and reflexive approach.

Through stages 1-2, approximately 50 themes were identified, combined and restructured to form 13 categories and 31 subcategories for coding. These categories we label as 'rhetorical strategies', as they entail combinations of thematic orientation, the senders' stance on the issue, and performative functions. These 'strategies' thus comprise a holistic impression of the Tweets based on the coders' interpretations and discussions. Stages 3 and 4 were the culmination of the coding process resulting in the final identification of rhetorical strategies and designation into the four overarching categories that form the analytical framework. For Stage 4, approximately half the tweets (n=1,000), spanning the entire period,

were coded in NVivo, software which assists with the qualitative coding process and data management. The tweets were ordered chronologically and divided into groupings of 100, with every alternate grouping of 100 tweets selected for the next stage (ie, tweets 1-100, 201-300, etc). This generated 1,100 tweets at first, but 100 were removed after an initial reading due to the relatively weak nature of their contribution. Such a strategy ensured the duration of the period was well covered, with the reduced sample considered more manageable for close reading, and all tweet extracts quoted in the results and discussion are from this reduced sample. This round of reading and coding was also an effective way both to deductively test the framework and to get a sense of how widespread the different rhetorical strategies were, and thus how salient they were likely to be in the debate. These stages relied on the coders' ability to provide an accurate reading of the tweets which had been removed from their original context; three measures were taken to ensure reasonable interpretations of the rhetorical potential of each text: i) All coding stages were followed by elaborate discussions, about both themes and particular tweets, between the two researchers/coders; ii) The final coding stage aimed for consensus among both coders, with open discussions about discrepancies; and iii) Unclear cases, characterised by irony, intertextual references, conversational responses, etc. that were difficult to categorise, were excluded in the final coding stage.

## Results

Figure 2 gives an overview of the four dominant strategies identified and how these positions are represented in the smaller sample deductively coded in NVivo. The first, dominant category (~60%) consists of arguments for or against the *principle* of the blocklist; the second (~20%) relates to comments regarding the collective rationale behind blocking; the third category (~10%) consists of tweets criticising and defending people's inclusion on the list; and the final category (<10%) consists of tweets that, although not presenting refutable arguments or claims, express the sender's own position and in-groups/out-groups perceptions within the debate. This overview does not give an accurate impression of how salient different arguments might be in individual users' feeds but can give an impression of the extent to which the arguments dominated overall.

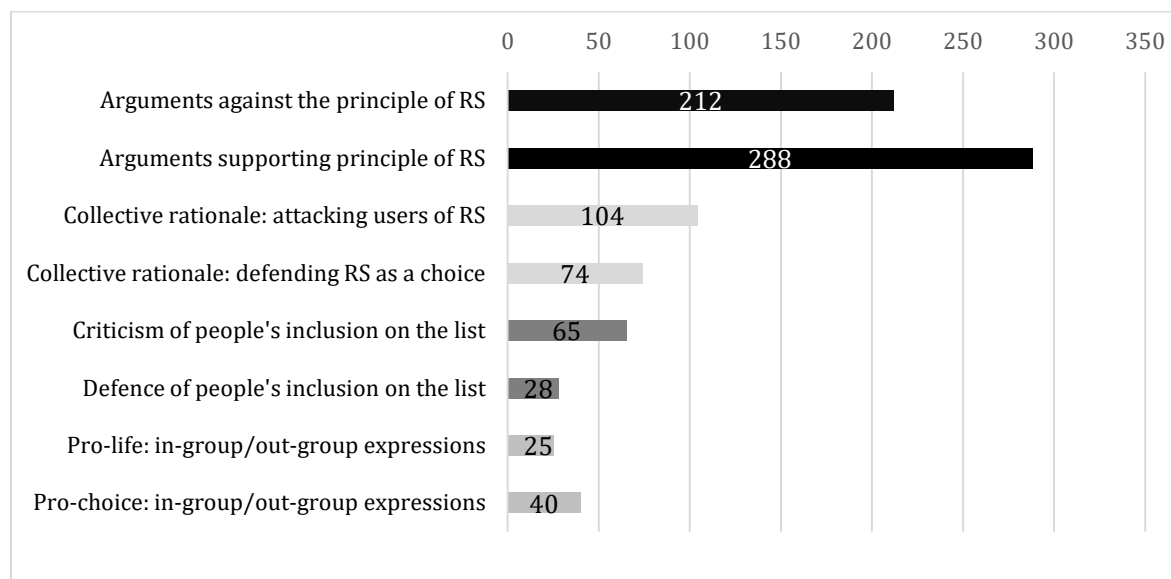


Figure 2: The four dominant categories of rhetorical strategies identified in the material, with two positions in each category

### 1: Arguments Dominating the Debate Surrounding Repeal Shield

The first category of rhetorical strategies is based on presenting *arguments for or against the principle of Repeal Shield (RS)*. These tweets are usually easy to sort into pro-choice and pro-life discourses based on the tweet’s content, rather than drawing conclusions from users’ biographical details. Table 2 gives an outline of the content categorised as principled arguments for or against Repeal Shield (RS):

	<i>Arguments for RS</i>	<i>Arguments against RS</i>	
<i>Pro-choice</i>	<ul style="list-style-type: none"> <li>- Improves user experience</li> <li>- Enables civic participation</li> </ul>		} <b>“Safe/sane space”</b>
<i>Pro-life</i>		<ul style="list-style-type: none"> <li>- Undermines user experience</li> <li>- Threatens civic participation</li> <li>- Undermines user experience</li> </ul>	
		<ul style="list-style-type: none"> <li>- Violates privacy</li> <li>- Stigmatises people</li> </ul>	

Table 2: Arguments for or against Repeal Shield from pro-choice and pro-life perspectives

The most common rhetorical strategies from the pro-choice side are arguments based on how RS *improves user experiences* and *enables participation*. These include claims about how RS protects users' mental health and makes Twitter a 'nicer place', ensuring users can resist the temptation of 'meaningless arguments', and facilitates a space for engagement without disruptive elements.

A central element in this rhetorical strategy is simply to praise how much 'nicer' or 'more peaceful' Twitter is with RS ('*You are all doing some great work. My timeline is a happier place now, thanks*'); it also becomes clear that users expect to now avoid abusive and aggressive material ('*No more trolls spying lies and abuse!*'). A central group here appears to be women either targeted themselves or who feel the nature, and scale, of tweets from anti-abortion activists is a personal burden.

Furthermore, RS is described as a way to *protect* users from trolls and abuse, or as a way for pro-choice voters and activists to 'stay safe' and 'take care' of themselves. These arguments are among the most salient, suggesting RS improves user experience by offering a 'safe space' to protect pro-choice people, in particular women. Many explicitly describe RS's benefits in *mental health* terms. These include both references to personal experiences that make confrontations stressful ('*Seriously, @repeal\_shield is great for my mental health, as a very #prochoice parent that did have an abortion when I was younger*'), and more general claims about the psychological toll of participating in conversations. In other cases, mental health references are probably better understood as general statements of how the blocked content is considered unpleasant and stressful ('*@repeal\_shield is really saving my sanity these days*'). This indicates that users see the blocklist as a preventative measure to avoid distress or harm, tied to the *harm of reduction* where users want to avoid ending up in a worse state because of Twitter content.

Drawing on this rhetorical strategy, pro-choice voices also make points about how RS makes it easier for users to *participate more meaningfully in the debate*, taking two distinct forms. The first is that RS helps users to resist the temptation to get into what they perceive as meaningless arguments ('*For me it is about getting rid of the urge I have to reply to every damn thing that one of them says*'). The second is that RS facilitates a space for engagement without disruptive elements ('*Don't waste your time with anti-choice trolls before the referendum seriously please just block them and focus your efforts on better things!*'). The general argument is that RS enables more productive participation by avoiding distracting encounters with people whose opinions are unlikely to be changed. This is also evident in those promoting RS, encouraging new users and campaigners to focus their efforts more productively ('*It gives people the strength to change minds where they can be changed: at home, in work, or on stalls and doorsteps*'). This suggests that users perceive differently how the blocklist assists them. Those referencing mental health still cite the idea of self-care, safety and 'safe spaces', even if that latter term is used in a derogatory sense from opponents. Yet for others, it appears to be more about working within a 'sane space' amid rational, calmer discourse.

The most common *rhetorical strategies from the pro-life side against the shield* are arguments based on how the blocklist *represents a threat to civic participation* and how it *undermines the user experience of Twitter* by obstructing dialogue. Central elements in this

rhetorical strategy are claims about how the blocklist is incompatible with democratic norms. Many describe RS as a form of censorship (*'It is outrageous. If that is not censorship then I don't know what your definition is'*) or obstruction of freedom of speech (*'These people fighting for 'bodily autonomy' and some kind of perceived liberation who want to shut out free speech and different opinions.'*). Tweets drawing on this rhetorical strategy are very often marked by strong language use, describing RS and its users as 'dumb', 'ignorant', or even comparing it to fascists and Nazi-like conduct (*'So-called 'liberals' against free speech, open, honest, and fair debate. Censorship. And, as you say, for wide range of people, who are mostly respectful. Pretty Nazi-like concept.'*). Others suggest that the blocklist supports anti-democratic impulses to shut down deliberation among citizens of opposing views (*'These people hate debate, they hate opinion that is not compatible with their own and, quite frankly, they just seem to be full of hate'*).

The other element of this rhetorical strategy consists of claims about how RS undermines Twitter as a political arena. A common claim is that RS is a purposeful introduction of echo chambers (*'Pitiful but predictable to see that @repeal\_shield is creating an echo chamber for such scare stories'*). Others approach the formation of such spaces as an inevitable consequence of RS, or RS fitting into a pattern already resembling an echo chamber (*'@repeal\_shield The servants of death want to turn the public space into a huge echo chamber where they can only hear their cries and the gnashing of teeth'*). All these principled elements align with the *harm of repression*, in which users interpret blocklists as a barrier to achieving their goal of, perhaps, appearing on opponents' timelines in a non-curtailed Twitter interface, or their desire for open participation and dialogue.

Another line of argument under this more general rhetorical strategy is that the blocklist *stigmatises people* and that lists presumably based on political opinions are a *violation of privacy*. Regarding stigmatisation, RS is frequently referred to as a 'blacklist' or 'ze list' (*'The List is compiled on the basis – supposedly - of someone holding a political opinion'*). Others express concerns that the blocklist labels citizens that hold different political beliefs as 'trolls' and abusive (*'They're a crew of demented self-appointed Stasi. There are twelve thousand accounts on the blacklist. Many are pro-choice. These are all slandered as "pro life trolls"'*). Remarks about privacy are made most frequently in relation to privacy laws or norms (*'Of course, the fact that everyone listed has data protection cases against the Repealers behind it will be months of fun...'*), or in relation to Twitter's policy (*'You should contact Twitter and ask them why they allow #Repeal-shield, preventing debate'*).

While most arguments *for* RS are clearly from the pro-choice side, and most arguments *against* RS are from the pro-life side, there are deviations, most notably those opposing the principle behind the blocklist from a pro-choice position. Like the pro-life side, these suggest that RS *undermines the user experience of Twitter*. Some express concerns about handing over control of their Twitter feed to others (*'I don't really like other people controlling who I block – it's great if some people want to use it, but I prefer to not'*). Others raise the general principle of blocking political opponents (*'It is a noxious instrument serving no purpose at all other than inciting hatred. This is playing right into the opponents' hands'*). Others still see it as a strategic advantage to monitor or 'keep an eye' on opponents (*'I would*

rather know what those people are thinking and saying and what moves them, and what has left them drawing a blank').

In sum, as suggested in table 2, the arguments form two opposing positions that understand RS in similar, but qualitatively very different, ways. One side sees RS as a *safe space*, a shield against behaviour that threatens genuine debate and isolates people from participating, as well as a *sane space* in which more productive conversation and participation can develop. The other side – comprising both pro-life and pro-choice voices – sees RS as a blatant attempt to promote an *echo chamber* and stifle debate between opposing viewpoints. Some comments illustrate the weaponisation of free speech, highlighting how those feeling aggrieved use a perceived slight like being added to a blocklist as a proxy for a violation of their rights and a basis for attack. This is believed to threaten genuine debate and isolate people from participating, while also raising concerns about potential stigmatisation and infringement of privacy rights.

## **2: Collective Rationale: Typical Liberal Behaviour or Right to Choose?**

The second category of rhetorical strategies is based on *criticising or defending peoples' rationale for using @Repeal\_Shield*; these tweets are oriented towards the act of shielding being understood as a form of group behaviour. Rather than presenting arguments for why RS is positive or negative, the strategy here is to attack or defend the collective motivation behind its use. These tweets are more difficult to sort into a pro-choice or pro-life discourse; the divide is more clearly drawn between those who support RS (including both pro-choice and pro-life views) and those who oppose RS (also including both discourses).

<i>In support of Repeal Shield</i>	<i>In opposition to Repeal Shield</i>
- Repeal Shield is choosing not to listen	- Repeal Shield is indicative of liberal behaviour - Public figures should not use Repeal Shield

Table 3: *Criticising or defending peoples' choice to use the Repeal Shield.*

A frequent rhetorical strategy in support of RS is arguing that it is *choosing not to listen to parts of the other side*. At the core is the claim that RS is not threatening the agency of those listed, but rather it is the list's users proclaiming their agency. This is often presented either as a counter-argument, asserting that RS does not qualify as censorship ('*Blocking you on twitter is not a fascist tactic. ... It's a choice not to listen to your bile*'), or as a clarification of their own position, emphasising the element of choice and how they have the right to block whomever they want ('*It is entirely voluntary, so how is that troubling? Just don't subscribe*'). This was often tied to the abortion issue itself, with users referring to the right to

choose, a common refrain by pro-choice activists (*'Those of us who signed up have made a personal choice due to the abuse we have received from no campaigners. But then, No folk aren't fans of having personal choice are they?'*)

The most common rhetorical strategy in opposition to RS is to argue that it is an example of typical liberal behaviour. The pro-choice side, and more generally 'the left' or 'liberals', are frequently described as 'sheltered', 'intolerant' and 'cowardly'; using RS is believed to fit into a pattern of avoiding opposing views (*'haha the left wants diversity in everything, apart from the diversity of opinions. 'Oh they are disagreeing'... Block!'*). A version of this are claims that the pro-choice side is afraid of 'the truth' or having their opinions challenged, and thus actively seek to shield themselves from opposing views (*'Sad thing is their not listening to any of this they have their weird censorship block out the truth repeal shield thingy on. Truth hurts I guess'*).

Another strategy in opposition to RS is to raise questions about public figures using or promoting RS. Critique is directed at famous people, journalists and the media (*'They are the most closed minded ... The agenda is to have abortion and nothing will stop them'*), and political officials (*'Many socialist politicians and groups run to promote abortion, dubious groups that don't want their content/opinions challenged'*). The underlying assumption is that these individuals have an obligation to interact with people of opposing views and beliefs.

In sum, this category of rhetorical strategies, drawing on *criticising or defending peoples' choice to use the Repeal Shield*, also cluster around two main positions. On one hand, RS users are approached as a metonym for the entire pro-choice campaign, the 'liberals' or 'intolerant' political left. On the other hand, blocklists are defended as a choice and a right not infringing on others' right to express themselves or participate. These arguments can often be read as attempts to reduce or minimise the implications of RS, by assuring that it will not actually silence opposing views.

### ***3: Responses to the Selection of People on the Blocklist***

The third category is based on *criticising or defending the selection of people on the Repeal Shield list*. The main question here is whether RS is reasonable given whom it blocks. Again, the divide is more easily drawn between those who support or oppose RS, as many users who clearly hold pro-choice sentiments express scepticism to whether the list reflects its intentions.

<i>In support of Repeal Shield</i>	<i>In opposition to Repeal Shield</i>
- Assumptions about people on the list	- Reactions to who's on the list - Reactions from people on the list

*Table 4: Criticising or defending selection of people on the blocklist*

The most common rhetorical strategy in support of RS is drawing links between those listed and their presumed attitudes on other issues, such as how they are often assumed to be racist and misogynistic (*'A pleasant side-effect is how MYSTERIOUSLY my feed is also seeing a vast reduction in racists, rape apologists and contempt for the poor.'*).

The most prominent rhetorical strategies here, however, are against RS. These are mainly reactions from people who think they are wrongfully put on the list themselves (*'I am active each day tweeting to repeal the Eighth. I ended up on the blocklist and am sure I am not the only one'*) or people reacting to blocking on other grounds than what is stated by RS. Some are also accusing RS of strategically blocking opposing viewpoints, an argument similar to the accusations that the pro-choice side is 'afraid of the truth' (*'\*If\* ze list was for what they said it was for ... I'd think it was fine. Instead the admins use it to block people willy fucking nilly'*).

#### 4: Celebrating Own Position

The final category of strategies is based on *in-group and out-group expressions* from the pro-choice/pro-life sides. These tweets do not present arguments for how RS should be understood, or whether RS is good or bad. Rather, they embrace and enforce implied group divisions, and appear to relish the provocation.

<i>Pro-choice</i>	<i>Pro-life</i>
- Satisfaction in frustrating pro-lifers	- Pride in being placed on RS
- Desire to have RS offline/other media	- Satisfaction that RS will hurt the pro-choice side

Table 5: *In-group and out-group expressions from the pro-choice and pro-life side*

The most common rhetorical strategies from the pro-choice side are expressions of satisfaction in frustrating the other side. In particular, RS users enjoy how those blocked are no longer listened to, or that they enjoy their protests (*'I use Repeal Shield as it pisses them off so much'*). In these cases, users are clearly satisfied by RS's ability to provoke those blocked. A similar argument can be drawn from users wishing for an 'offline Repeal Shield' or something for other social networks (*'Wish there was a Black Mirror-style real life @repeal\_shield that I could install for the next few weeks'*).

The most common rhetorical strategies from the pro-life side are expressions of pride in being listed or encouraging others to be proud, claiming it should be seen as a 'badge of honour' (*'Took a few days, but I managed to get on the Roll of Honour for Human Rights Heroes! :-)'*). Whether genuine or not, this can be read as a way to resist or redefine RS's sanctioning aspect.



On both sides, users drawing on these rhetorical strategies are less concerned with arguments for or against either RS, its use and composition, than with their own relation to it and what that can say about their political and social identity and position. Expressions of pride in being blocked and satisfaction in provoking opponents are clear manifestations of in-group out-group behaviour. The satisfaction the users seemingly experience is not only based on recognition from those sharing their beliefs, but also on the opponents' expressed rejection.

### Discussion & Conclusion

The diverse, animated contributions described throughout are indicative of how the structured, mass-blocking affordance of Twitter clearly exercised users on both sides. One contributor, supportive of the list, said in response to a critic: '*Odd that you think of a blocklist as an attack, as opposed to a defence*', and this remark reflects the core schism and tension. From those supporting the blocklist, patterns of individualism, personal choice, protection and self-care are to the fore. Conversely, more broad, principled, and combative arguments – positioned at the societal, rather than individual level, amid claims of democratic transgressions – are made by those who oppose the list. This also resonates with the harms of reduction and repression (Milovanovic, 2006): those supporting the list for individual protection and mental health used it to avoid that kind of harm of reduction, which could also be seen as an infringement on Maslow's physiological and safety needs. Conversely, those who objected to the list on principled levels interpret it as a violation and form of repression, whereby their esteem/reputation, and potential for self-actualisation are being compromised. The need for collectiveness and belonging was apparent in the in-group and out-group expressions evident from both sides. Such a multitude of considerations aligns with some previous findings on blocklists, such as concerns about slander and censorship (Gosse and O'Meara, 2018) and frustration with being blocked (Jhaver et al., 2018), arguments that contrast with those highlighting how blocklists make Twitter cleaner and easier for those facing harassment (Geiger, 2016; Jhaver et al., 2018).

Although these arguments from both sides appear distinct, there are clear commonalities. Both perspectives are striving to ensure Twitter is a healthy, deliberative space based on their respective visions of the social network's role in political debate. Both sides employ civic norms like participation and engagement, describing how these are needed to fulfil democratic ideals. Yet the central variances come from how they mobilise these norms and articulate such visions. Similarly to Gamergate, @Repeal\_Shield shows how participants actively negotiate what constitutes unacceptable behaviour, while demonstrating how motivations for using blocklists are more complex than simply muting abusive content. On an idealistic level, among those critical of RS, there were outcries over the apparent exclusion of some participants from this centralised public space. This contrasted with the claims from RS supporters that filters were, in fact, *necessary* to participate in such spaces, echoing the counterpublic narrative previously associated with blocklists (Geiger 2015, Maréchal 2016), especially relevant for women or minorities who may feel more of a need for such interventions (Jhaver et al., 2018). This group of RS supporters also argued that

civic participation in such online spheres was, in fact, open to all – unless users actively violated the apparent parameters and established norms of acceptable, civilised behaviour. This way, the affordances of Twitter that allow for the formation of blocklists might potentially have a formative effect on active norms, as blocklists are perceived as a way to sanction unwanted behaviour. On a more pragmatic level, those who used the list credited its role in facilitating their participation in other dimensions of the campaign, while those opposing the list frequently demeaned the other side, drawing on invective terms amid allegations of insecure convictions. These critics also drew on vocabulary common in the US regarding the derogation of 'safe spaces' and 'liberal snowflakes', as well as a weaponisation of the free speech concept.

Through the different rhetorical strategies identified we clearly see formations of communities in which the 'collective sensemaking' (Geiger, 2016: 789) is not only based on shared values, but also on continuous negotiations of what constitutes harassment. In the RS debate, participants make sense not only of the abortion issue, but also of the different positions offered by the issue, and the people who hold them. For some, using the blocklist is seemingly a bold statement of identity, a trend also demonstrated by some of those listed who saw inclusion as a badge of honour.

The desire to participate is clearly evident from both perspectives, reminding us that Twitter maintains an esteemed role in times of heightened political interest. Citizen-to-citizen exchanges are valued by both sides, but blocklists clearly create barriers; at a fundamental level, blocklist technology provides Twitter users with tools to easily exclude large swathes of accounts whom they may have previously encountered organically. Ultimately, the analysis here demonstrates how blocklists may be indicative of a shift towards Twitter becoming a more closed network for some, arguably challenging one of its distinctive features of openness. Consequently, as concerns grow about deepening socio-political divisions (Sunstein, 2017), questions arise regarding whether blocklists encourage the further fragmentation of group identities, thus questioning the ability for deliberation and collective sensemaking. Nevertheless, individual decision-making and empowerment is a clear motivating factor for blocklist users which cannot be dismissed, especially where questions of abusive content arise.

The analysis also suggests how curating and subscribing to blocklists can constitute a form of political participation. According to van Deth's (2014) conceptual map, initiating and curating a blocklist can most fittingly be described as a form of *targeted political participation*. This form of participation does not deal with matters of state directly, but is still lies in the political sphere. Although the most apparent objective of the curators is to *shield* individual Twitter users from unwanted content and abuse in their personal feed, the blocklist clearly arises out a situation of political choice regarding a referendum. Also, people on both sides tend to interpret RS as a strategic attempt to influence the referendum by giving less attention to the pro-life side.

Keeping in line with the understanding of van Deth's political participation, abstentions – like boycotting products or particular messages – can also be considered actions. Subscribing to blocklists can then be considered a form of *motivationally defined political participation* (van Deth, 2014). At the core is whether the *problem* that subscribing to a

blocklist aims to solve is perceived as personal or collective. As the analysis suggests, both subscribers and others see subscribing as an expression of political aims and intentions. At the same time, the most common rhetorical strategy in defence of the list is that it is a personal prerogative to improve one's own Twitter experience, indicative of the apparent need to 'tame' Twitter (Jhaver et al., 2018: 21). This way, the controversy surrounding the blocklist can be read as a response to the fact that engaging with lists are *indistinct* forms of political participation. Contributing to a major blocklist in a political referendum context for personal reasons is behaviour placed in the sphere of the political that insists on not being politically motivated. The case of RS demonstrates how, when the issue at hand is political, subscribing to a blocklist – even for personal reasons, like improving one's social media experience or avoiding unwanted content – is also perceived as politically motivated, and thus treated as a form of political participation by both opponents and supporters.

Elsewhere, Twitter's own action or inaction regarding the blocklists also requires further attention, especially as many users are seemingly employing such tools to improve troubling experiences on the platform. On its help section, Twitter supports blocking as a 'feature that helps you control how you interact with other accounts on Twitter' (Twitter Inc, 2018), providing instructions on how to block individual users and download and share your list of blocked accounts. Elsewhere, it considers 'add[ing] users to lists in a bulk or aggressive manner' as problematic in a spamming context. This may appear as somewhat of a contradiction, yet the social network appears to have resisted taking any clear stance on blocklists. It is also noteworthy that many opposed to being listed on RS highlighted their concerns through the lens of perceived injustice. In the case of RS, the administrators appeared to offer recourse to those who felt they were unfairly listed, by being allowed to appeal, but this attempt at fairness is at the discretion of every blocklists' administrators; such elements of due process are not necessarily built into the tools' framework, nor encouraged by Twitter itself.

In positioning this research internationally, there is no doubt that this Irish case is modest and making statements about its universality must be cautioned against. Nevertheless, it has provided a much-needed empirical case beyond the Gamergate list and reinforces Geiger's (2015) observation that blocklist use is expanding beyond cases of explicit harassment. Given the study's original framework, it would be valuable for future researchers to consider its applicability to other blocklists. It would also be useful to shift attention to those who administer blocklists as their motivations and reflections are underexplored. However, while @Repeal\_Shield was open and transparent about its administrators, this may not always be the case which could make this line of enquiry challenging, especially if some founders are difficult to identify and contact, or reluctant to engage. From the user perspective, attitudes about delegating control to administrators, and how this perhaps affects the autonomy of social media use, could also be insightful. Overall, due to the lack of empirical research on blocklists thus far, all efforts to contribute to this area, growing in prevalence as technological affordances evolve, would be valuable.

## REFERENCES

- Auerbach D (2015) Beware the Blocklists. Available at: [http://www.slate.com/articles/technology/bitwise/2015/08/twitter\\_blocklists\\_they\\_can\\_stop\\_harassment\\_and\\_they\\_can\\_create\\_entirely.html?via=gdpr-consent](http://www.slate.com/articles/technology/bitwise/2015/08/twitter_blocklists_they_can_stop_harassment_and_they_can_create_entirely.html?via=gdpr-consent) (accessed 24 July 2018).
- Bakshy E, Messing S and Adamic LA (2015) Supporting Materials for Exposure to Ideologically Diverse News and Opinion on Facebook. *Science* 348(6239): 1130–1132. DOI: 10.7910/DVN/LDJ7MS.
- Barnes R (2018) Lessons from #Gamergate. In: Barnes R (ed.) *Uncovering Online Commenting Culture: Trolls, Fanboys and Lurkers*. Cham: Springer International Publishing, pp. 93–111. DOI: 10.1007/978-3-319-70235-3\_5.
- Benesch S (2017) Civil Society Puts a Hand on the Wheel: Diverse Responses to Harmful Speech. *Perspectives on harmful speech online: A collection of essays* (August): 42–43. Available at: [https://dangerousspeech.org/wp-content/uploads/2017/08/2017-08\\_harmfulspeech.pdf#page=31](https://dangerousspeech.org/wp-content/uploads/2017/08/2017-08_harmfulspeech.pdf#page=31).
- Bennett, L. W., & Segerberg A (2013) *The Logic of Connective Action. Digital Media and the Personalization of Contentious Politics*. New York: Cambridge University Press.
- Bennett LW and Segerberg A (2012) The Logic of Connective Action. *Digital Media and the Logic of Contentious Politics. Information, Communication & Society* 15(5): 739–768.
- BlockTogether.org (2018) BlockTogether: A web app intended to help cope with harassment and abuse on Twitter. Available at: <https://blocktogether.org/> (accessed 26 July 2018).
- boyd danah and Crawford K (2012) Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information Communication and Society* 15(5): 662–679. DOI: 10.1080/1369118X.2012.678878.
- Caplan R and boyd d (2016) Who controls the public sphere in an era of algorithms: Case studies. *Data & Society*: 10.
- Castells M (2010) *The Rise of the Network Society. The information age ; v.1*. Oxford: Wiley-Blackwell.
- Clark-Parsons R (2018) Building a digital Girl Army: The cultivation of feminist safe spaces online. *New Media and Society* 20(6): 2125–2144. DOI: 10.1177/1461444817731919.
- Couldry N (2015) The myth of 'us': digital networks, political change and the production of collectivity. *Information, Communication & Society* 18(6): 608–626.
- Djohari N, Pyndiah G and Arnone A (2018) Rethinking 'safe spaces' in children's geographies. *Children's Geographies* 16(4). Taylor & Francis: 351–355. DOI: 10.1080/14733285.2018.1487032.

- Dorsey J (2018) March 1. @Jack. Available at: <https://twitter.com/jack/status/969234275420655616> (accessed 30 March 2018).
- Dubois E and Blank G (2018) The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information Communication and Society* 21(5). Taylor & Francis: 729–745. DOI: 10.1080/1369118X.2018.1428656.
- Fox S (2014) Is it time to update the definition of political participation? Political participation in Britain: The decline and revival of civic culture. *Parliamentary Affairs* 67(2): 495–505.
- Fraser N (1990) Rethinking the Public Sphere: a Contribution to the Critique of Actually Existing Democracy. *Social Text* (25/26). Working papers. University of Wisconsin-Milwaukee, Center for Twentieth Century Studies: 56–80.
- Geiger RS (2016) Bot-based collective blocklists in Twitter: the counterpublic moderation of harassment in a networked public space. *Information Communication and Society* 19. Taylor & Francis: 787–803. DOI: 10.1080/1369118X.2016.1153700.
- Golbeck J, Gnanasekaran RK, Gunasekaran RR, et al. (2017) A Large Labeled Corpus for Online Harassment Research. *Proceedings of the 2017 ACM on Web Science Conference - WebSci '17*: 229–233. DOI: 10.1145/3091478.3091509.
- Gosse CE and O'Meara VJ (2018) “Blockbotting Dissent”: Publics, Counterpublics, and Algorithmic Public Sphere(s). *Stream: Inspiring Critical Thought* 10(1): 3–11.
- Government of Ireland (1983) Eighth Amendment of the Constitution Act. Available at: <http://www.irishstatutebook.ie/eli/1983/ca/8/enacted/en/print> (accessed 24 November 2018).
- Harvey D and Gasca D (2018) Serving healthy conversation. Available at: [https://blog.twitter.com/official/en\\_us/topics/product/2018/Serving\\_Healthy\\_Conversation.html](https://blog.twitter.com/official/en_us/topics/product/2018/Serving_Healthy_Conversation.html) (accessed 14 August 2018).
- Heatherly KA, Lu Y and Lee JK (2017) Filtering out the other side? Cross-cutting and like-minded discussions on social networking sites. *New Media and Society* 19(8): 1271–1289. DOI: 10.1177/1461444816634677.
- Hickey M (2016) ‘Safe space’ now a tool of liberal censorship. *Irish Examiner*. Available at: <https://www.irishexaminer.com/viewpoints/analysis/safe-space-now-a-tool-of-liberal-censorship-422170.html> (accessed 6 March 2019).
- Jhaver S, Ghoshal S, Bruckman A, et al. (2018) Online Harassment and Content Moderation: The Case of Blocklists. *ACM Trans. Comput.-Hum. Interact* 25(2): 1–33. DOI: 10.1145/3185593.
- Lanier MM (2014) Harms of Reduction and Repression. In: Arrigo BA (ed.) *Encyclopedia of Criminal Justice Ethics*. Thousand Oaks, CA: SAGE Publications.

- Lukianoff G and Haidt J (2018) *The Coddling of the American Mind: How Good Intentions and Bad Ideas Are Setting up a Generation for Failure*. London: Allen Lane.
- Maliepaard E (2017) Bisexual Safe Space(s) on the Internet: Analysis of an Online Forum for Bisexuals. *Tijdschrift voor Economische en Sociale Geografie* 108(3): 318–330. DOI: 10.1111/tesg.12248.
- Mantilla K (2013) Gendertrolling: Misogyny Adapts to New Media. *Source: Feminist Studies* 39(2): 563–570. DOI: 10.2307/23719068.
- Maréchal N (2016) When bots tweet: Toward a normative framework for bots on social networking sites. *International Journal of Communication* 10(June): 5022–5031.
- Marwick AE (2014) Ethnographic and qualitative research on Twitter. In: Weller K, Bruns A, Puschmann C, et al. (eds) *Twitter and Society*. New York: Peter Lang, pp. 109–122.
- Milovanovic D (2006) Legalistic definition of crime and an alternative view. *The Annals of the Faculty of Law in Belgrade, international edition*: 48–86.
- Murthy D (2017) The Ontology of Tweets: Mixed-Method Approaches to the Study of Twitter. In: Sloan L and Quan-Haase A (eds) *The SAGE Handbook of Social Media Research Methods*. SAGE, pp. 559–572. DOI: 10.4135/9781473983847.n33.
- Nicholson R (2016) 'Poor little snowflake' – the defining insult of 2016. *Guardian.com*. Available at: <https://www.theguardian.com/science/2016/nov/28/snowflake-insult-disdain-young-people> (accessed 7 March 2019).
- Nolan L (2017) Free speech under threat on our college campuses. *The Irish Times [online]*. Available at: <https://www.irishtimes.com/news/education/free-speech-under-threat-on-our-college-campuses-1.3045149> (accessed 6 March 2019).
- Papacharissi Z (2010) *A Networked Self: Identity, Community and Culture on Social Network Sites*. New York: Routledge.
- Papacharissi Z (2012) Without You, I'm Nothing: Performances of the Self on Twitter. *International Journal of Communication* 6: 1989–2006. DOI: 1932–8036/20120005.
- Power J (2018) Conservatives on campus: 'We're imprisoned by liberals'. *The Irish Times [online]*. Available at: <https://www.irishtimes.com/news/education/conservatives-on-campus-we-re-imprisoned-by-liberals-1.3666673> (accessed 6 March 2019).
- Rainie L and Wellman B (2012) *Networked: The New Social Operating System*. Cambridge, MA: MIT Press.
- Repeal Shield (2018). January 31. @Repeal\_Shield. Available at: [https://twitter.com/repeal\\_shield/status/958726860573159424?lang=en](https://twitter.com/repeal_shield/status/958726860573159424?lang=en) (accessed 26 March 2018).
- Shaw A (2017) Encoding and decoding affordances: Stuart Hall and interactive media

technologies. *Media, Culture and Society* 39: 592–602. DOI: 10.1177/0163443717692741.

Sunstein CR (2017) *#republic: Divided Democracy in the Age of Social Media*. Princeton, New Jersey: Princeton University Press.

Tandoc EC, Ling R, Westlund O, et al. (2018) Audiences' acts of authentication in the age of fake news: A conceptual framework. *New Media and Society* 20(8): 2745–2763. DOI: 10.1177/1461444817731756.

The Roestone Collective (2014) Safe Space : Towards a Reconceptualization. *Antipode* 46(5): 1346–1365. DOI: 10.1111/anti.12089.

Twitter Inc (2018) How to control your Twitter experience. Available at: <https://help.twitter.com/en/safety-and-security/control-your-twitter-experience> (accessed 13 November 2018).

van Deth JW (2014) A conceptual map of political participation, *Acta Politica* 49(3): 349–367. *Acta Politica*.

van Dijck J (2006) *The Network Society. Social Aspects of New Media*. Thousand Oaks: SAGE Publications.

Vis F (2013) Twitter as a reporting tool for breaking news: Journalists tweeting the 2011 UK riots. *Digital Journalism* 1(1): 27–47. DOI: 10.1080/21670811.2012.741316.

Woffard T (2014) One woman's new tool to stop Gamergate harassment on Twitter. Available at: <https://www.newsweek.com/one-womans-new-tool-stop-gamergate-harassment-twitter-288008> (accessed 20 August 2018).