

RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

A Fringe Mainstreamed, or Tracing Antagonistic Slang between 4chan and Breitbart before and after Trump

Peeters, Stijn; Willaert, Tom; Tuters, Marc; Beuls, Katrien; Van Eecke, Paul; Van Soest, Jeroen

Published in:
How Misinformation Propagates on Social Media

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for pulished version (HARVARD):
Peeters, S, Willaert, T, Tuters, M, Beuls, K, Van Eecke, P & Van Soest, J 2023, A Fringe Mainstreamed, or Tracing Antagonistic Slang between 4chan and Breitbart before and after Trump. in R Rogers (ed.), *How Misinformation Propagates on Social Media: A Cross-platform Analysis*. Amsterdam University Press, Amsterdam, pp. 165-185.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

8 A fringe mainstreamed, or tracing antagonistic slang between 4chan and Breitbart before and after Trump

Stijn Peeters, Tom Willaert, Marc Tuters, Katrien Beuls, Paul Van Eecke and Jeroen Van Soest

Abstract

We studied whether the vernaculars of the extremely vitriolic, “politically incorrect” sub-forum of 4chan/pol/ have crossed over to the comment section of Breitbart News, a right-wing news website that was found in earlier research to have played a significant “agenda-setting” role in the 2016 U.S. presidential elections. We study if crossover exists around both the 2016 and 2020 elections. In our analysis, we find evidence suggestive of such crossover, centered around the presence first on 4chan and later Breitbart of a series of racist, antagonistic and otherwise extreme terms. This crossover of 4chan/pol/s vitriolic vernacular marks an expansion of hyper-antagonistic “alt-right” politics to Breitbart’s more mainstream right-wing populist audience.

Keywords: Alt-right, 4chan, Breitbart, vernacular crossover, extreme speech

Research questions

Can we find evidence of language originating on 4chan that propagates to the comment sections of Breitbart News around the time of the 2016 U.S. elections? How to characterize the words used on 4chan as compared to Breitbart around that time? Does the use and change in use of language on both platforms suggest a spread of extreme political thought? Can we observe similar dynamics of language propagation between both platforms around the 2020 U.S. presidential elections?

Essay summary

Over the past decade a diverse and increasingly influential far-right online media sphere has emerged. It has raised concerns that parts of this sphere may function as incubators for radicalization. In particular, the 2016 presidential elections in the United States were marked by the coarsening of the tone of political discourse, with candidate and eventual winner Donald Trump slandering his opponents, spreading conspiracy theories and provoking xenophobia. Alongside Trump's insurgent takeover of the Republican party, his election campaign during 2015 and 2016 marked the emergence of the "alt-right" political movement, which perceived Trump as an alternative to establishment conservatism.

As a libertarian movement with a strongly xenophobic, often racist stance towards immigration, the alt-right was also characterized by its use of antagonistic vernacular. We can think of this antagonistic slang as "memes," a concept typically used to refer to user-generated shared images that seem to spread across platforms and between communities, but which can also be used to refer to any "building blocks of complex cultures" online, including words and phrases (Shifman, 2011, p189). Indeed, in the analysis on offer here, we view specific phrases and tokens as such memetic building blocks that seem to propagate within and between distinct environments online. A platform of interest in this context is the far-right image board 4chan, which has been positioned as a "birthplace of memes" (Ludemann, 2018), an incubator of conspiracy theories like QAnon (De Zeeuw et al., 2020), and a place of rapid innovation of oftentimes antagonistic language (Peeters et al., 2021). It might therefore be expected that antagonistic alt-right slang incubated on the platform has the potential to spread to a wider audience, with 4chan acting as a breeding ground. To study this hypothesis, we look at 4chan as well as a more mainstream platform that has been associated with the alt-right, Breitbart News.

The questions are particularly relevant as the alt-right is a relatively unique, insurgent far-right political movement that rose to international attention in 2015 with remarkably little in the way of a centralized organizational structure, and for whom the circulation of memes and internet jargon was fundamental to its success (Hawley, 2017). Most emblematically, the memetic subcultural icon of "Pepe the Frog" became notoriously associated with this school of thought during the first half of the 2010s and achieved widespread attention (Lobinger et al., 2020). Arguably, however, among the alt-right's most significant accomplishments was the extent to which their antagonistic slang succeeded in framing political discussion.

Illustrative of this pattern was for example the expression “cuckservative” which emerged in early 2015 on (now deplatformed) alt-right websites such as My Posting Career, The Right Stuff as well as on 4chan’s notorious /pol/ forum (Bernstein, 2015). In its original far-right subcultural usage the term referred to a genre of often racialized pornography thereby connecting a critique of establishment republicanism with the far-right’s longstanding preoccupation with masculinity and miscegenation. By the end of the year, the prolific alt-right author Vox Day had self-published a track with the title *Cuckservative: How “conservatives” betrayed America*, and this alt-right meme had effectively worked its way into political discussion amongst mainstream Trump voters. It is this type of “propagation” of politically extreme vocabulary that is under study in this chapter.

Considering these recent events, there is a legitimate concern that the subculture associated with sites at the “bottom” of the internet could insinuate itself (or has already done so) with an extreme and conspiratorial discourse into the American political debate across a continually evolving range of platforms. There are indications that it has already transpired in the more recent 2020 U.S. election campaign. The QAnon persona, central to a right-wing conspiracy theory positing, among other things, that prominent members of the Democratic Party are part of a Satan-worshipping cannibalistic cult, started on 4chan but has since become a major factor in mainstream U.S. politics and as such is now discussed on a wide variety of platforms (De Zeeuw et al., 2020; Stanley-Becker, 2020). The polarized language we study reflects this rift in recent American political discourse.

An understanding of the internet as having a “bottom” implies the existence of further “layers.” Along these lines, at the top we would find big media conglomerates, often rooted in “legacy media” such as major newspapers as the *New York Times*, cable broadcasters as CNN, and newer online-first outlets like Vox. As one moves “down,” platforms grow more obscure, with a smaller reach and less clear editorial or content policies. At the bottom, one finds “fringe” sites, with obscure subcultures; this “deep vernacular web” (De Zeeuw and Tuters, 2020) can appear culturally baffling as well as offensive to the uninitiated. Sites in this stratum usually have a relatively small number of visitors, compared to mainstream sites. 4chan is particularly relevant here, as a fringe platform that has nevertheless been scrutinized for its production of internet memes (Bernstein et al., 2011), peculiar subcultural practices (Nissenbaum and Shifman, 2017) as well as language innovation (Tuters and Hagen, 2020; Peeters et al., 2021).

Our findings are based on datasets centered on the 2016 and 2020 U.S. elections, collected from 4chan/pol/ and from the comment section of

Breitbart News, a conservative, right-wing American news website especially popular during the first period as a staunch supporter of eventual winner Donald Trump. Although it has been described as “factually dubious” (Guess et al., 2018), Breitbart News occupied a crucial place in the political media ecosystem at the time. Benkler et al. (2018) offered an in-depth study of Breitbart’s “agenda-setting” role in that election. Their analysis shows how Breitbart “anchored” a network of other similarly dubious right-wing news sites such as Daily Caller, Gateway Pundit and Infowars. Though no formal or editorial association between these sites exists, they provide a similar brand of content characterized as a mix of “paranoid conspiracy interpretations around a core of true facts” (Benkler, 2018, p. 34). Together they occupied a crucial position in the media ecology around the 2016 elections.

In this ecology Breitbart is a particularly interesting site for several reasons. One is that, at the time, Breitbart was the largest of these sites with approximately 10% of the entire general news audience according to one estimate (Malone, 2016). Founded by the deceased Andrew Breitbart, formerly a reporter for the Huffington Post, under the more recent editorship of Steve Bannon the site championed the right-wing libertarian Tea Party and a strongly American populist, civic nationalist agenda (cf. Burley, 2017). Receiving substantial financial support from the billionaire Mercer family, who initially backed Ted Cruz in the 2015 U.S. election campaigning, Breitbart would develop into a nakedly partisan branch of the Trump campaign while at the same time Bannon famously claimed that he considered the site to be a “platform for the alt-right” (Posner, 2016). With a background in both high finance and documentary filmmaking, Bannon is a self-styled public intellectual noted for his interest in an obscure branch of far-right political philosophy known as Traditionalism, which also had a readership on 4chan/pol/ (Teitelbaum, 2020; Tuters and OILab, 2020). Bannon would later join the Trump campaign as its chief strategist (Green, 2017). In 2016, Breitbart published an article entitled *An establishment conservative’s guide to the alt-right*, co-authored by the notorious alt-right provocateur Milo Yiannopoulos. An investigative report later revealed it to have been written with the participation of known alt-right ideologues (Bernstein, 2017). As such, the site combines a clear, alt-right editorial position and explicit ties to the Trump campaign with a relatively wide reach.

Earlier analyses of Breitbart, including Benkler et al.’s, were limited to the editorial content of the site. We instead study the comment sections of Breitbart’s articles that routinely receive thousands of comments, many only tangentially related to the article’s subject. These appear to be moderated loosely, if at all. A 2017 report cites Disqus, which provides the technology on

which Breitbart's comment section runs, promising that Breitbart "[wants] to work with us to figure out ways to minimize [hate speech]" (Captain, 2017). In this permissive setting, the comment section of Breitbart's London section was characterized as "a malignant swamp of race-baiting, nativism and antisemitic conspiracy," even accused of providing a platform for notorious alt-right celebrities (Mulhall et al., 2017). Appearing to function like a largely unmoderated discussion forum, the comment threads can thus serve to study the political views and discourse of the readership of a highly active element of far-right politics that moved increasingly to the center of the American Republican party around the 2016 elections.

In this same period /pol/, the self-described "politically incorrect" discussion board of anonymous imageboard 4chan, overtook /b/ (the "random" board) as the site's most active discussion forum. Previous quantitative research on /b/ has noted how the site was an "excellent venue for studying innovation diffusion," due in part to the fact that it was generally considered as "the source of many online memes" (Bernstein, 2011, p. 56). While in the earlier period in which /b/ had been more popular 4chan was the source of innocuous memes such as LOLcats, /pol/ memes were far more toxic, including offensive depictions of Pepe the Frog as well as the antisemitic triple parentheses phrasal meme (Tuters and Hagen, 2020). While there has been some quantitative research into the diffusion of toxic /pol/ memes to other web communities (Zannettou et al., 2018), to our knowledge there is no previous empirical work focused specifically on the crossover of vernacular language from /pol/ to another such threaded discussion forum.

This chapter, then, adds to a growing body of work focused on the "mainstreaming" of previously "fringe" web spaces like 4chan as the source of a "neoreactionary" style of political discourse (Nagle, 2017; Wendling, 2018; Beran, 2019; Woods, 2019). 4chan and Breitbart represent two parts of the media ecosystem that are particularly interesting to study in the context of the polarized and increasingly extreme U.S. political landscape. As such, we investigate whether 4chan's discourse resonates beyond its own borders around the time of the 2016 and 2020 U.S. presidential elections. Since so much of the discourse on 4chan's political discussion board, /pol/, can be characterized as conspiratorial, racist or otherwise extreme (cf. Tuters and Hagen, 2020), its later occurrence on other platforms is of great interest to those studying the mainstreaming of extremism and misinformation. While our analysis is primarily focused on the 2016 election campaign in which the alt-right movement first gained prominence, we also provide an initial analysis of the 2020 campaign for comparison.

We found that there are far more terms that appear only in the language of 4chan/pol/ than in the language of Breitbart comments. Additionally, of the terms that over time are prominent first in one dataset and later in both, those that first appear on 4chan are often highly political and furthermore can be characterized as anti-Muslim and xenophobic (e.g., “germanistan”), homophobic or transphobic (e.g., “xhe”) or otherwise extreme (e.g., “shitlibs”). These extreme terms are then later observed on Breitbart. Though a direct relationship is difficult to ascertain, our initial findings suggested that 4chan, an active but non-mainstream niche site, had an outsized impact that reaches beyond its own confines.

We reflect on these findings, concluding that for the period 2015–2017 4chan/pol can be considered an originator or incubator of extreme discourse, where extreme idioms appear before propagating to the more mainstream discussion space of Breitbart News. Additionally, our observations indicate that this propagation of idioms between 4chan and Breitbart News seem to be less intense around the time of the 2020 U.S. presidential elections, and that consequently, studies of extreme discourse and misinformation should consider and monitor other platforms as the main sites of the mainstreaming of such terms. We end with a brief section on our data collection and analytical methods.

Implications

Our findings indicate that around the time of the 2016 U.S. elections antagonistic, highly political and problematic words that also can be characterized as xenophobic (e.g., “germanistan”); transphobic (e.g., “xhe”) or otherwise extreme (e.g., “shitlibs”) first observed on 4chan later entered the discourse in the comment section of Breitbart News, a more mainstream platform with important connections to the Trump presidential administration. While earlier research has investigated the crossing over of particular ideas (e.g., conspiracy theories), our study provides empirical data that suggests that this crossing-over also occurs on the level of language and is not bound only to specific theories or ideas. The findings further support previous observations about the sustained connection between 4chan/pol/ and Breitbart’s comment section during this period.

One possible explanation for the propagation of extreme “chan” vernacular towards Breitbart around the 2016 elections is that some 4chan posters also frequent Breitbart’s comment section. It would not be surprising if they used the language they were familiar with, which could explain their

occurrence in both spaces. Tracing whether actors move between these platforms is difficult because 4chan is designed as an anonymous platform (Knutti, 2011). 4chan posters are notoriously derisive of “mainstream media” and typically dismiss Breitbart as inadequately extreme. Although Breitbart has been described as having an “extreme right-wing bias” (Media Bias/Fact Check, 2021), it is seen as a place for “normies.” In the vernacular, “normies” are those who follow mainstream media and otherwise adhere to common social norms (De Zeeuw et al., 2020). Nevertheless, it is possible that some 4chan posters may also frequent Breitbart News, which would be one explanation for the appearance of 4chan-like vernacular there. It would be the manner for both this vernacular as well as the extreme political positions to which it implicitly and explicitly refers to spread to a new “normie” audience.

Though a direct relationship between both platforms remains difficult to ascertain, our initial findings suggest that 4chan, the active but non-mainstream niche site, had an outsized impact that reaches beyond its own confines. As such, we conclude that for the period 2015–2017 4chan/pol can be considered an originator or incubator of extreme discourse, where extreme idioms appeared before they propagated to the more mainstream discussion space of Breitbart News. Additionally, our observations indicate that this propagation of idioms between 4chan and Breitbart News seems to be less intense around the time of the 2020 U.S. presidential elections, and that consequently studies of extreme discourse and misinformation should consider and monitor other platforms as the main sites of the mainstreaming of such terms.

A key implication of our work, then, is that 4chan /pol/ might give an early impression of problematic discourse that may become used by a wider audience at a later stage. As such, continued observation of the language disseminated through these fringe platforms—for which we offer one methodological blueprint by addressing its propagation towards Breitbart News—might benefit journalists, researchers and policy makers seeking to signal the emergence of new extreme discourses on emerging platforms such as Parler (cf. Floridi, 2021) and others that have more recently gained prominence in the 2020 U.S. election campaign.

More fundamentally, our findings speak to the much-debated relationship between the “bottom” of the internet—consisting of niche, often politically extreme sites—and more mainstream sites. The observation and study of this “bottom” has acquired urgency as ideas and vernacular that originate in these parts have been implicated in several far-right terrorist attacks in the United States, Canada and New Zealand. Furthermore, sites like

4chan serve as incubators for various impactful conspiracy theories, e.g., “Pizzagate” (Tuters et al., 2018) and the figure of QAnon (De Zeeuw et al., 2020). Indeed, while for many years the effects of the web were framed in terms of the democratic promise of participatory media (Jenkins, 2006; Benkler, 2006), the last half decade has shaken that narrative to its core with the emergence of “dark participation” in the context of online political discussion (Quandt, 2018). The role of the upstart Breitbart in anchoring a right-wing news ecosystem that set the agenda for the 2016 U.S. election may be seen as the fruition of earlier concerns over the fragmentation of the web into personalized spheres (Pariser, 2011), which continue apace with the emergence of the alt-tech ecosystem that has benefited from social media platforms’ “deplatforming” of the Trump movement (Rogers, 2020b). Given the “fringe” quality of some of these sites we have good reason to believe that their vernacular subculture will overlap with that of 4chan, as this study showed for the “normie” website, Breitbart News, in the midst of the 2016 U.S. election.

Findings

For both the 2015–2017 and 2020–2021 periods, we split up the 4chan and Breitbart posts and comments in terms; each word, after filtering out hyperlinks and punctuation, is a term. For each term we can then classify on which of the platforms it occurs on a per-month basis, resulting in a propagation pattern for each term (see Figure 8.1).

Our findings suggest that around the 2016 U.S. presidential elections, the political vocabulary associated with extreme right-wing politics consistently appears on 4chan first, and then on the more mainstream Breitbart News later, potentially representing one strand of this propagation dynamic. We also observed that this dynamic becomes less prominent around the 2020 elections, suggesting that the locus for this extreme idiom’s propagation from the “bottom” of the internet has again shifted. In particular, the analysis of these propagation patterns allows for the following observations:

Finding 1: Around the time of the 2016 U.S. presidential elections, the language of 4chan/pol/ contains more unique terms than that in the Breitbart comment sections. Our analysis shows that there are more terms unique to the /pol/ dataset than there are terms unique to the Breitbart dataset. Of the 67,605 terms, 19,346 (28.6%) were classified as occurring in the 4chan/pol/ dataset only, while 2,857 (4.2%) were classified as occurring only in the Breitbart dataset (see Figure 8.2). 4chan has previously been described as a

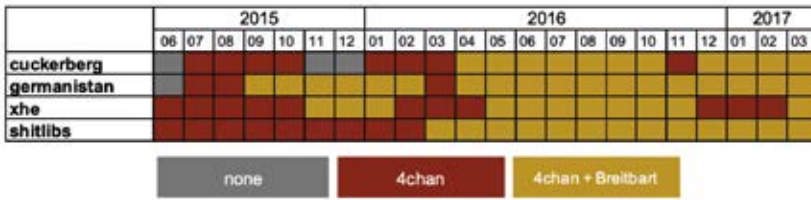


Figure 8.1 Visualization of the monthly occurrence of the terms “cuckerberg,” “germanistan,” “xhe,” and “shitlibs” around the time of the 2016 U.S. presidential elections, between June 2015 and March 2017. For each month, terms are classified (color-coded) based on a comparison of their relative frequencies in 4chan/pol posts and in the comments on Breitbart News. These words represent the political vernacular found within the 4chan dataset.

source of subcultural and linguistic innovation (Nissenbaum and Shifman, 2017). This finding empirically confirms the observation, at least concerning the unique use of language on the forum. As English-language datasets, both are concerned with informal political discussion focused primarily on the United States context. Thus, while some variation may be expected; in principle, one might expect the language used to be similar between both, but this is only partially the case.

4chan’s vernacular has been referred to as “chanspeak”: “peculiar in-group misspellings” characterized by “shortening, simplifying and cutting down words” (Fiorentini, 2013; Herring, 2012). While this is perhaps true for the broader 4chan vernacular, the /pol/ slang we found is not adequately captured by this description. This can be attributed to the rapid linguistic innovation on this forum (Peeters et al., 2020). The terms we find are more adequately described as “phrasal memes,” highly self-referential “remixes” of words, e.g., “cuckerberg” (a combination of “cuck” and “Zuckerberg”). While a proper linguistic analysis of this vernacular is outside the scope of this article, the dataset on offer here could in the case of /pol/ serve as a starting point for such a study.

Finding 2: During the same period, a substantial number of terms are first only observed in the language on 4chan, the fringe platform, but later also on Breitbart, the more mainstream platform, suggesting propagation of this vocabulary. Terms that occur on one platform first and later on another platform or both platforms can be observed in both “directions”; some occur first on Breitbart while others occur first on 4chan/pol/. In total, 2,043 terms (3%) follow such a pattern. Of these, 932 (45.6%) occur on 4chan first, while 1,111 (54.4%) occur on Breitbart first. This seems counterintuitive; it would imply that terms are first anchored in the language of Breitbart and only later in that of 4chan, which is difficult to reconcile with 4chan’s reputation as a more innovative linguistic space as established in Finding

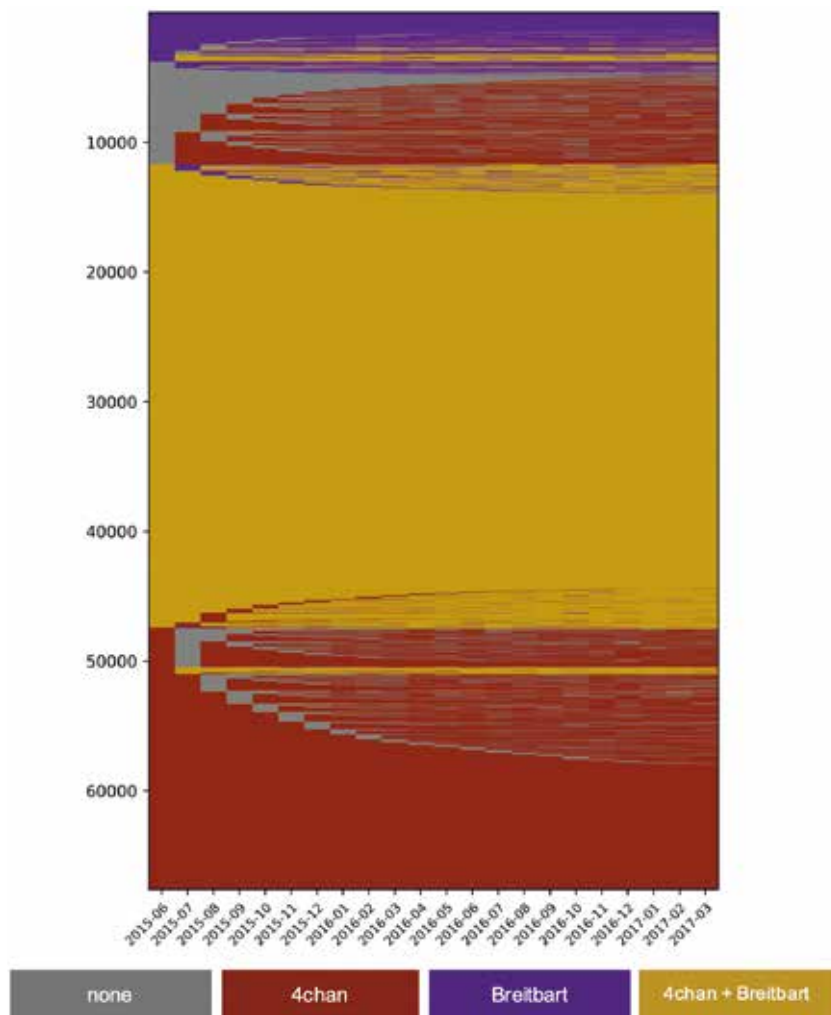


Figure 8.2 Classifications of all words in the 2015–2017 data ($N = 67,605$) over time. The consecutive occurrences of each word are represented as a single row of per-month squares that are color-coded for the occurrence of the word in 4chan/pol/ posts and Breitbart comments respectively.

1. On the other hand, Breitbart is a far larger and arguably more influential platform, and agenda-setting power may be attributed to it in that capacity. From this perspective, the fact that a substantial number of terms occur on 4chan first at all is significant and suggests that the terms might indeed “propagate,” with language spreading—directly or indirectly—from 4chan eventually to Breitbart.

A closer look at these terms reveals that they can be divided into two broad categories—“named entities” and “neologisms.” Linguistically, named

entities refer to all terms that are proper names, for example, countries and people. The other category, “neologisms,” are words that are neither common English nor otherwise used in “normal” discourse. In practice, these terms are mostly various slurs and part of a memetic vocabulary that is associated with 4chan discourse.

The named entities cannot reasonably be assumed to originate on either platform. Instead, the likely explanation for the occurrence of these terms is that they refer to people, places or organizations that were discussed because they were relevant to a current event or news item. This indicates that Breitbart users discussed these topics before 4chan, which is interesting insofar as it provides insight into the type of topics discussed by both forums and how rapidly they enter the discourse. The “neologisms” (including the examples in Figure 8.1) on the other hand are likely to originate in the vernacular of online platforms (Peeters et al., 2021). As such, the fact that they appear on 4chan and later on Breitbart suggests that they do propagate from the one to the other, either directly or via another intermediary platform.

Finding 3: Around the time of the 2016 U.S. presidential elections, many terms that seem to propagate from 4chan/pol/ to Breitbart reflect an extreme far-right politics. Of these terms that can be assumed to originate on 4chan/pol/, most are implicitly or explicitly related to far-right and conspiratorial theories or ideas. This is not surprising, since 4chan/pol/ itself has been associated with the “Pizzagate” political conspiracy (Tuters et al., 2018) and has been described as a “kind of petri dish for concocting extreme and extremely virulent forms of right-wing populist antagonism” (Tuters and Hagen, 2020, p. 2223). Of the words that appear first on 4chan (see also Figure 8.1) several are emblematic of an extreme political discourse, such as “cuckerberg” (a jab at Facebook owner Mark Zuckerberg combined with the slur “cuckold”; other variations found were “cuckservative(s),” “cucktard,” “cucky,” and “cuckery”), anti-Muslim terms such as “germanistan” and “britainistan,” words like “xhe” (used mockingly to insult transgender people), and various slurs aimed at liberal U.S. voters like “shitlibs” and “berniebots.”¹

While 4chan/pol/ is well-known as a far-right discussion space (Hine et al., 2017; Ludemann, 2018), our data and analyses show that the vocabulary associated with this discourse is not contained to this “fringe” platform but after initial usage it also appears on more mainstream platforms. More specifically, the various xenophobic or otherwise extreme slurs and phrasal memes that are developed and incubated on 4chan/pol/ in some cases see

1 The full list of terms that propagate may be found in the dataset available from Zenodo at <https://doi.org/10.5281/zenodo.5535341>.

uptake in the comments on Breitbart News. As most of this language is unambiguous, and hard to mistake for anything else than derogatory, it raises concerns that not only the language but also the extreme political discourse associated with it is shared across sites.

Finding 4: Around the time of the 2020 U.S. presidential elections, the aforementioned mainstreaming of extreme chan vernacular seems to be less outspoken. Around the time of the 2020 U.S. presidential elections, some of the vernacular that propagated in 2015–2017 remains shared between 4chan and Breitbart, with notable examples including “cuck” and its derivatives, such as “cucked.” Further analysis shows, however, that comparatively fewer new terms propagate from 4chan to Breitbart News around this time. As observed in our 2020–2021 dataset, only 347 terms out of 57,602 (or 0.6% compared to 3% for the 2015–2017 dataset) actually move from one platform to the other, and of those, only 124 were classified as moving from 4chan/pol to Breitbart (see Figure 8.3). Closer inspection of these moving terms reveals few original vernacular terms, even though the data suggest that during this period, 4chan in and of itself remains an incubator for extreme vernacular. Examples that do point towards a continued mainstreaming of 4chan terminology and memes concern the terms “coomer” (which refers to the 4chan meme of the “20-year-old coomer”), and “libshits” (an inversion of the previously discussed term “shitlibs”), but in comparison with the 2015–2017 period, the language propagation dynamics between both platforms seems much less outspoken.

Any comparison between the two datasets is necessarily tentative as we are yet to capture as much of the post-election period as we did for the 2016 elections. As such there remains a possibility that the propagation dynamic lags in this case, or that Breitbart’s comment space has later become milder for other reasons—perhaps 4chan’s interest in Breitbart has diminished, which may be found in a subsequent analysis. Nevertheless, the discourse around both elections may be assumed to reach its zenith in the months surrounding the election date. As such, the data gathered around the 2020 elections should provide a representative impression of the discourse around that election, even if it is quantitatively smaller and of a shorter duration than the earlier dataset.

These empirical observations strengthen previous assertions in the literature that the period 2015–2017 was one characterized by an intensified and salient “mainstreaming” of harmful vernacular between 4chan/pol and Breitbart News. Possible explanations for the relative decline in the propagation of idioms from 4chan to Breitbart around the 2020 U.S. presidential elections include the fact the site experienced a precipitous

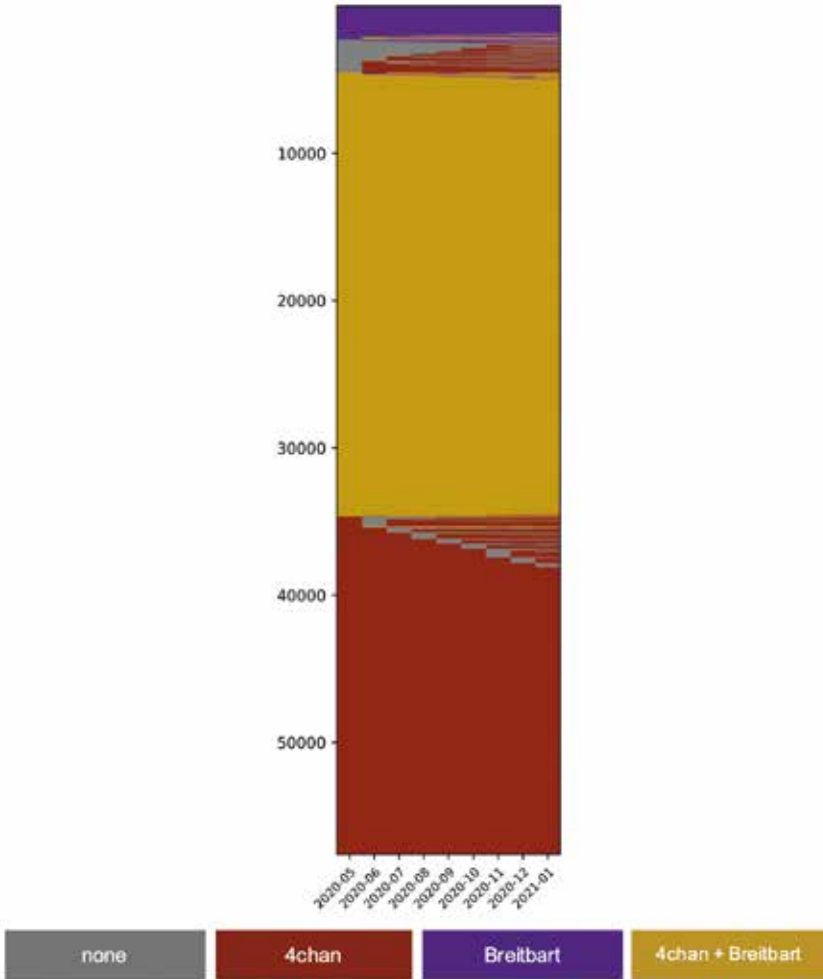


Figure 8.3 Classifications of all words in the 2020–2021 data (N = 57,603) over time. The consecutive occurrences of each word are represented as a single row of per-month squares that are color-coded for the occurrence of the word in 4chan/pol/ posts and Breitbart comments respectively.

decline after Trump took office, ultimately losing as much as three quarters of its total audience (Ellefsen, 2019). According to Steve Bannon the site appears to have struggled financially following an advertiser boycott, which began in 2016 and was organized by Sleeping Giant to protest the site’s bigotry and sexism (Klayman, 2019). The billionaire Mercer family sold their shares in the site in 2017 and are currently majority stakeholders of the alternative social media site Parler, connected to the 2021 storming of the DC Capitol (Lerman, 2021). The growth in significance of such

“alt-tech sites” can be seen as one of the bi-products of the “deplatforming” of alt-right figureheads from social media—including eventually Trump himself (Rogers, 2020b).

Methods

4chan data was collected with 4CAT, a forum analysis toolkit (Peeters and Hagen, 2021) that contains a dataset comprising 4chan /pol/ data from 2013 to the present. This data is collected continuously (as it is posted on 4chan) by the tool itself and, for the period prior to 2018, supplemented with data from 4plebs.org, a third-party 4chan archive which publishes semi-regular data “dumps” on the Internet Archive, containing all posts made on a number of 4chan’s boards, including /pol/. (Merged 4plebs’ and 4CAT’s datasets have been used in other research on 4chan, too (Tuters and Hagen, 2019; Voué et al., 2020; Jokubauskaitė and Peeters, 2020).) Notably, posts are included even if they are later deleted from the site, as all posts eventually disappear from 4chan itself, as threads are deleted after a period of inactivity.

The 2015–2017 Breitbart data was collected between September 2–9, 2019 using a custom scraper written in Python which first crawled breitbart.com for internal links to create an index of all articles posted on the site, and then collected all comments for all articles posted between June 2015 and March 2017, using the Disqus API. The resulting dataset reflects the state of the comment section as it was at the moment of scraping. There is a possibility that some comments were removed between the moment of posting and the moment of scraping, up to 4 years later; however, as mentioned earlier, Breitbart’s moderation policy seems to have been lax during the period we study, and it is unlikely that later policies were enacted retroactively. We therefore assume that the data is a reasonably accurate reflection of what the comment threads would have looked like closer to the date the comments were posted. The 2020–2021 Breitbart data was collected with the same technique, between February 17 and March 3, 2021.

Our first dataset thus spans the period between the announcement of Donald Trump’s candidacy for the U.S. presidential election (June 2015) and his first months in office, whereas the second dataset comprises a smaller interval around the 2020 U.S. elections in which Trump was again a candidate (and lost). As data capture of this nature is cumbersome, we were unable to gather a dataset comprising an equal timespan so shortly after the 2020 elections; a more direct comparison would be a fruitful avenue for future work.

Before analyzing the captured data, for both datasets we cleaned the scraped comments and posts by applying case folding and removing punctuation, URLs, HTML tags (in Breitbart comments) and comment referral numbers such as ">>280207128" (in /pol/ comments).

Analysis

This chapter addresses the questions of (1) whether we can empirically identify terms that are first prominent in the language on 4chan/pol/ and later also in the language of the Breitbart comment sections around the time of the 2016 U.S. presidential elections, (2) how to characterize the language used on 4chan/pol/ compared to that of Breitbart's comment sections at that time, (3) whether the nature of these identified terms indicates a spread of extreme political thought, and (4) whether we can identify similar dynamics between both platforms around the time of the 2020 U.S. presidential elections. We expect that we can observe this pattern for terms associated with far-right thought, and that it constitutes a mainstreaming of fringe, taboo or otherwise extreme political concepts.

Quantitative analysis

We investigate corpora of posts and comments using methods from natural language processing to empirically identify terms that occur first on one platform and then on another, and to quantify the propagation patterns of these terms between both platforms (Willaert et al., 2020; Willaert et al., 2021). We collected two datasets for both platforms, a first set comprising posts from June 2015 through March 2017, and a second set containing data from May 2020 through January 2021. These texts were then tokenized (split into individual terms). For both platforms, the monthly frequencies of each term were counted, and those terms with an absolute frequency of less than 200 were removed, as these were mostly less germane and included typos. Next, the relative monthly frequency of each term was calculated for both /pol/ and Breitbart. Relative frequencies were used because we are interested in the prominence of the terms in the language of each platform, and we aim to compare this prominence between platforms. We then compare these relative frequencies and classify each term into one of four classification bins, indicating whether for a given month the term:

- occurred neither on /pol/ nor on Breitbart,
- occurred on both /pol/ and Breitbart,

- occurred exclusively on /pol/, or
- occurred exclusively on Breitbart.

For each term, this analysis results in a sequence of classification bins. In order to reduce the influence of very low frequency terms, a term is only assigned to the Breitbart bin or /pol/ bin if it had a relative frequency of at least 0.00001%. If not, its frequency is considered to be 0 for that month. This filtering resulted in a classification sequence for each term, which was visualized using color coding (Figure 8.2).

Qualitative analysis

The initial quantitative approach yielded a subset of terms for both periods that warranted further scrutiny; we are particularly interested in those terms that were first observed as prominent on 4chan/pol/ and later also observed on Breitbart. Our approach here was to first remove any obvious named entities (people, countries, institutions) from the list as well as common English language. The remaining tokens could then be analyzed in more detail via a closer reading, in which the context and occurrence of the token on 4chan/pol/ as well as on other platforms is studied via 4plebs (the searchable archive of /pol/) and 4CAT (the modular web platform scraping tool). Here we retained words with a clear political (sub)text, similar to those shown in Figure 8.1.

As such, we have employed a quali-quantitative approach (Venturini and Latour, 2010), where we combine an initial computational analysis of a large dataset to extract a relevant subset of the corpora at hand, which we then analyze further with a more interpretative qualitative approach of this subset.

References

- Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. Yale University Press.
- Benkler, Y., Faris, R., and Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press.
- Beran, D. (2019). *It came from something awful: How a toxic troll army accidentally memed Donald Trump into office*. St. Martin's Publishing Group.
- Bernstein, J. (2015, July 27). Behind the racist hashtag that is blowing up Twitter. *BuzzFeed News*. <https://www.buzzfeednews.com/article/josephbernstein/behind-the-racist-hashtag-some-donald-trump-fans-love>.

- Bernstein, J. (2017, October 5). Here's how Breitbart and Milo smuggled Nazi and white nationalist ideas into the mainstream. *BuzzFeed News*. <https://www.buzzfeednews.com/article/josephbernstein/heres-how-breitbart-and-milo-smuggled-white-nationalism>.
- Bernstein, M., Monroy-Hernández, A., Harry, D., André, P., Panovich, K., and Vargas, G. (2011). 4chan and /b/: An analysis of anonymity and ephemerality in a large online community. In *Proceedings of the International AAAI Conference on Web and Social Media*, 5(1), Article 1. <https://ojs.aaai.org/index.php/ICWSM/article/view/14134>.
- Burley, S. (2017). Disunite the right: The growing divides in the pepe coalition. Political Research Associates. <https://www.politicalresearch.org/2017/09/19/disunite-the-right-the-growing-divides-in-the-pepe-coalition>.
- Captain, S. (2017, March 8). Disqus grapples with hosting toxic comments on Breitbart and extreme-right sites. *Fast Company*. <https://www.fastcompany.com/3068698/disqus-grapples-with-hosting-toxic-comments-on-breitbart-and-extreme-right-sites>.
- Day, V. and Eagle, J. R. (2016). *Cuckservative: How "conservatives" betrayed America*. Castalia House.
- De Zeeuw, D., Hagen, S., Peeters, S., and Jokubauskaite, E. (2020). Tracing normification. *First Monday*. <https://doi.org/10.5210/fm.v25i11.10643>.
- De Zeeuw, D. and Tuters, M. (2020). Teh internet is serious business: On the deep vernacular web and its discontents. *Cultural Politics*, 16(2), pp. 214–232. <https://doi.org/10.1215/17432197-8233406>.
- Ellefson, L. (2019, August 7). Breitbart's audience has dropped 72% since Trump took office—As other right-wing sites have gained. *The Wrap*. <https://www.thewrap.com/breitbart-news-audience-dropped-steve-bannon-72-percent/>.
- Florentini, I. (2013). "ZOMG! Dis is a new language": The case of lolpeak. *Newcastle Working Papers in Linguistics*, 13(1), pp. 90–108.
- Floridi, L. (2021). Trump, Parler, and regulating the infosphere as our commons. *Philosophy & Technology*, 34(1), pp. 1–5. <https://doi.org/10.1007/s13347-021-00446-7>.
- Green, J. (2017). *Devil's bargain: Steve Bannon, Donald Trump, and the nationalist uprising*. Penguin.
- Guess, A., Nyhan, B., and Reifler, J. (2018). Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 U.S. presidential campaign [Working paper]. <http://www.dartmouth.edu/~nyhan/fake-news-2016.pdf>.
- Hawley, G. (2017). *Making sense of the alt-right*. Columbia University Press.
- Herring, S. (2012). Special internet language varieties: Culture, creativity & language change [Paper]. The II LETiSS Workshop Language Go Web: Standard and Nonstandard Languages on the Internet, Pavia.

- Hine, G., Onaolapo, J., Cristofaro, E. D., Kourtellis, N., Leontiadis, I., Samaras, R., Stringhini, G., and Blackburn, J. (2017). Kek, cucks, and God emperor Trump: A measurement study of 4chan's politically incorrect forum and its effects on the web. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1), Article 1. <https://ojs.aaai.org/index.php/ICWSM/article/view/14893>.
- Jenkins, H. (2006). *Convergence culture: Where old and new media collide*. New York University Press.
- Jokubauskaitė, E. and Peeters, S. (2020). Generally curious: Thematically distinct datasets of general threads on 4chan/pol/. In *Proceedings of the International AAAI Conference on Web and Social Media* (pp. 863–867), 14.
- Klayman, A. (2019). The Brink [Feature documentary; Digital film]. <https://alisonklayman.com/the-brink>.
- Knuttila, L. (2011). User unknown: 4chan, anonymity and contingency. *First Monday*. <https://doi.org/10.5210/fm.v11i10.3665>.
- Lerman, R. (2021, February 24). Major Trump backer Rebekah Mercer orchestrates Parler's second act. *Washington Post*. <https://www.washingtonpost.com/technology/2021/02/24/parler-relaunch-rebekah-mercero/>.
- Lobinger, K., Krämer, B., Venema, R., and Benecchi, E. (2020). Pepe—Just a funny frog? A visual meme caught between innocent humor, far-right ideology, and fandom. In B. Krämer and C. Holtz-Bacha (Eds.), *Perspectives on populism and the media* (pp. 333–352). Nomos. <https://doi.org/10.5771/9783845297392-333>.
- Ludemann, D. (2018). /pol/emics: Ambiguity, scales, and digital discourse on 4chan. *Discourse, Context & Media*, 24, pp. 92–98. <https://doi.org/10.1016/j.dcm.2018.01.010>.
- Malone, C. (2016, August 18). Trump made Breitbart great again. *FiveThirtyEight*. <https://fivethirtyeight.com/features/trump-made-breitbart-great-again/>.
- Media Bias/Fact Check. (2021). Breitbart. <https://mediabiasfactcheck.com/breitbart/>.
- Nagle, A. (2017). *Kill all normies: Online culture wars from 4Chan and Tumblr to Trump and the alt-right*. Zero Books.
- Nissenbaum, A. and Shifman, L. (2017). Internet memes as contested cultural capital: The case of 4chan's /b/ board. *New Media & Society*, 19(4), pp. 483–501. <https://doi.org/10.1177/1461444815609313>.
- Papasavva, A., Zannettou, S., Cristofaro, E. D., Stringhini, G., and Blackburn, J. (2020). Raiders of the lost kek: 3.5 years of augmented 4chan posts from the politically incorrect board. *Proceedings of the International AAAI Conference on Web and Social Media*, 14, pp. 885–894.
- Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin.
- Peeters, S. and Hagen, S. (2021). The 4CAT Capture and Analysis Toolkit: A Modular Tool for Transparent and Traceable Social Media Research (SSRN Scholarly

- Paper ID 3914892). Social Science Research Network. <https://doi.org/10.2139/ssrn.3914892>.
- Peeters, S., Tuters, M., Willaert, T., and de Zeeuw, D. (2021). On the vernacular language games of an antagonistic online subculture. *Frontiers in Big Data*, 4(65). <https://doi.org/10.3389/fdata.2021.718368>.
- Quandt, T. (2018). Dark participation. *Media and Communication*, 6(4), pp. 36–48. <https://doi.org/10.17645/mac.v6i4.1519>.
- Rogers, R. (2020b). Deplatforming: Following extreme internet celebrities to Telegram and alternative social media. *European Journal of Communication*, 35(3). <https://doi.org/10.1177/0267323120922066>.
- Shifman, L. (2012). An anatomy of a YouTube meme. *New Media & Society*, 14(2), pp. 187–203. <https://doi.org/10.1177/1461444811412160>.
- Stanley-Becker, I. (2020, August 1). How the Trump campaign came to court QAnon, the online conspiracy movement identified by the FBI as a violent threat. *Washington Post*. https://www.washingtonpost.com/politics/how-the-trump-campaign-came-to-court-qanon-the-online-conspiracy-movement-identified-by-the-fbi-as-a-violent-threat/2020/08/01/d DOEa9b4-d1d4-11ea-9038-af089b63ac21_story.html.
- Teitelbaum, B. R. (2020). *War for eternity: The return of traditionalism and the rise of the populist right*. Penguin.
- Tuters, M. and Hagen, S. (2020). (((They))) rule: Memetic antagonism and nebulous othering on 4chan. *New Media & Society*, 22(12), pp. 2218–2237. <https://doi.org/10.1177/1461444819888746>.
- Tuters, M., Jokubauskaitė, E., and Bach, D. (2018). Post-truth protest: How 4chan cooked up the Pizzagate bullshit. *M/C Journal*, 21(3), Article 3. <https://doi.org/10.5204/mcj.1422>.
- Tuters, M. and OILab. (2020). Esoteric fascism online: 4chan and the Kali Yuga. In L. D. Valencia-García (Ed.), *Far-right revisionism and the end of history: Alt/histories* (pp. 287–303). Routledge.
- Venturini, T. and Latour, B. (2010). The social fabric: Digital footprints and qualitative methods. *Proceedings of Futur En Seine 2009: The Digital Future of the City*. Futur en Seine 2009.
- Voué, P., De Smedt, T., and De Pauw, G. (2020). 4chan & 8chan embeddings. *ArXiv:2005.06946 [Cs]*. <http://arxiv.org/abs/2005.06946>.
- Wendling, M. (2018). *Alt-Right: From 4chan to the White House*. Pluto Press.
- Willaert, T., Van Eecke, P., Beuls, K., and Steels, L. (2020). Building social media observatories for monitoring online opinion dynamics. *Social Media + Society*, 6(2). <https://doi.org/10.1177/2056305119898778>.
- Willaert, T., Van Eecke, P., Van Soest, J., and Beuls, K. (2021). A tool for tracking the propagation of words on Reddit. *Computational Communication Research*, 3(1), pp. 117–132. <https://doi.org/10.5117/CCR2021.1.005.WILL>.

Woods, A. (2019). Cultural Marxism and the cathedral: Two alt-right perspectives on critical theory. In C. M. Battista and M. R. Sande (Eds.), *Critical theory and the humanities in the age of the alt-right* (pp. 39–59). Springer. https://doi.org/10.1007/978-3-030-18753-8_3.

Zannettou, S., Caulfield, T., Blackburn, J., De Cristofaro, E., Sirivianos, M., Stringhini, G., and Suarez-Tangil, G. (2018). On the origins of memes by means of fringe web communities. *ArXiv:1805.12512 [Cs]*. <http://arxiv.org/abs/1805.12512>.

Funding

The authors received funding from the ODYCCEUS project within the European Union's Horizon 2020 program, grant agreement number 732942.

Data availability

Datasets with the monthly term counts for Breitbart comments and 4chan/pol/ counts for the periods under investigation are available from Zenodo at <https://doi.org/10.5281/zenodo.5535341>. The data do not contain any personal information or post-level metadata.

About the authors

Stijn Peeters, PhD, is Assistant Professor in Media Studies at the University of Amsterdam, Technical Director of the Digital Methods Initiative, and a co-investigator of the CAT4SMR project, where he has (co-)developed several research tools such as 4CAT. His research interests include the media-archaeological analysis of fringe communities on social media.

Tom Willaert, PhD, is a postdoctoral researcher in digital methods at the Vrije Universiteit Brussel. His research bridges methodological gaps between data science and humanities interpretative practice, with a focus on methods for the analysis of online (mis)information.

Marc Tuters, PhD, is a Senior Lecturer in Media Studies at the University of Amsterdam where his current research examines radical visual subcultures at the bottom of the Web together with colleagues at the Open Intelligence Lab as well as the Digital Methods Initiative.

Katrien Beuls, PhD, is Assistant Professor in artificial intelligence in the Faculty of Computer Science at the University of Namur. Her main research interests include emergent communication and language, and computational construction grammar.

Paul Van Eecke, PhD, is Assistant Professor in the Artificial Intelligence Laboratory at the Vrije Universiteit Brussel. His main research topics include the emergence and evolution of language through communicative interactions and computational construction grammar and its applications.

Jeroen Van Soest, MSc, is a developer and member of the Evolutionary and Hybrid AI team in the Artificial Intelligence Laboratory at the Vrije Universiteit Brussel (VUB). He has developed NLP tools and data science applications for the Horizon 2020 ODYCCEUS project and is VUB's lead developer on the imec-ICON Trendify project.