



This is a repository copy of *Which politicians receive abuse? Four factors illuminated in the UK general election 2019.*

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/162172/>

Version: Published Version

---

**Article:**

Gorrell, G. [orcid.org/0000-0002-8324-606X](https://orcid.org/0000-0002-8324-606X), Bakir, M.E., Roberts, I. et al. (2 more authors) (2020) Which politicians receive abuse? Four factors illuminated in the UK general election 2019. *EPJ Data Science*, 9. 18.

<https://doi.org/10.1140/epjds/s13688-020-00236-9>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:  
<https://creativecommons.org/licenses/>

**Takedown**






If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>



# Which politicians receive abuse? Four factors illuminated in the UK general election 2019

Genevieve Gorrell<sup>1\*</sup> , Mehmet E. Bakir<sup>1</sup> , Ian Roberts<sup>1</sup> , Mark A. Greenwood<sup>1</sup>  and Kalina Bontcheva<sup>1</sup> 

\*Correspondence:

[g.gorrell@sheffield.ac.uk](mailto:g.gorrell@sheffield.ac.uk)

<sup>1</sup>Department of Computer Science,  
Sheffield University, Regent Court,  
211 Portobello, Sheffield, UK

## Abstract

The 2019 UK general election took place against a background of rising online hostility levels toward politicians, and concerns about the impact of this on democracy, as a record number of politicians cited the abuse they had been receiving as a reason for not standing for re-election. We present a four-factor framework in understanding who receives online abuse and why. The four factors are prominence, events, online engagement and personal characteristics. We collected 4.2 million tweets sent to or from election candidates in the six week period spanning from the start of November until shortly after the December 12th election. We found abuse in 4.46% of replies received by candidates, up from 3.27% in the matching period for the 2017 UK general election. Abuse levels have also been climbing month on month throughout 2019. Abuse also escalated throughout the campaign period. Abuse focused mainly on a small number of high profile politicians, with the most prominent individuals receiving not only more abuse by volume, but also as a percentage of replies. Abuse is “spiky”, triggered by external events such as debates, or certain tweets. Some tweets may become viral targets for personal abuse. On average, men received more general and political abuse; women received more sexist abuse. Conservative candidates received more political and general abuse. We find that individuals choosing not to stand for re-election had received more abuse across the preceding year.

**Keywords:** UK general election; Online abuse; Cyber-bullying; Twitter; Politics

## 1 Introduction

Following the announcement of the UK parliamentary election, held in December 2019, a striking number of UK MPs, mostly women, cited the abuse and hostility they had been receiving in the course of their work as a factor in their decision not to stand for re-election [1]. This brought concerns about representation to the forefront; a political environment increasingly regarded as hostile and even dangerous may differently impact different sections of the community. The impact of social media on politics is also a cause for concern, not only with regards to incivility but also disinformation and polarisation [2]. Dialogue on the subject takes the form of media [3] and government engagement [4], re-

© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

search effort and innovation from the platforms [5]. Yet there is much work still to be done in understanding the causes of online toxicity and in forming an effective response.

In 2016, the UK voted to withdraw from the European Union in a close referendum that left parliament, as well as the nation, divided. In the context of heightened national feelings regarding “Brexit”, it will come as no surprise to many that online abuse toward politicians in the UK has increased ([6] and below), with strong feelings on both sides of the fence. Yet this is only one factor in the abuse we saw towards politicians in connection with the election. In this work we propose the following heuristic framework for understanding the abuse politicians receive on Twitter, which is discussed in more detail in the findings section below:

- *Prominence*: First and foremost, attention and therefore abuse focuses on a limited number of individuals most in the public eye.
- *Event surge*: Secondly, events may result in a surge in attention/hostility toward particular individuals—for example a political event or a media appearance.
- *Engagement*: Thirdly, an opinionated tweet by a politician provides a focus for any ill feeling towards their viewpoint or them as individuals that may be present on the Twitter platform.
- *Identity*: Fourthly, politics, gender, ethnicity and other personal factors affect the opinions that individual may express without incurring abuse (“norm violations” [7] or political intimidation) as well as the form that abuse is likely to take.

This study draws on longitudinal data spanning the general election in 2019 and also the previous UK general election in 2017. Using natural language processing, we identify abuse and the recipient of the abuse, and type it according to whether it is political or sexist abuse, or abuse of any kind. This enables a large-scale quantitative investigation. The context of the general election and the event-rich nature of a campaign provide opportunities to test the robustness of observations across multiple similar contexts. Work of this nature is strengthened by comparing findings across multiple periods, in which factors such as events and individuals have changed, demonstrating the robustness of generalisations.

The main contributions of the paper are threefold. Firstly, we present evidence that online abuse toward UK politicians is rising, and quantify this. We then present findings organised in terms of illustrating the above framework, beginning with the way abuse distributes itself among individuals and moving on to the way events produce variation in this, the tweets likely to precipitate out abuse, and finally the way genders and political parties are treated differently, with reference also to the broader picture of discrimination against groups. Finally, we present the first comparison of the abuse toward MPs who chose to stand again vs. those who chose not to, in order to inform the topic of how abuse affects political careers and therefore, ultimately, representation.

*Warning* In describing our work, we make use of strong, offensive language and slurs. This may be distressing for some readers.

## 2 Related work

As the effect of abuse and incivility in online political discussion has come to the fore in public discussion, the subject has begun to be seriously investigated by researchers [8, 9]. Binns and Bateman [10] review Twitter abuse towards UK MPs in early 2017. Gorrell *et al.* [11] compare similar data from both the 2015 and 2017 UK general elections. Ward

*et al.* [12] explore a two and a half month period running from late 2016 to early 2017. Greenwood *et al.* [13] extends work presented by Gorrell *et al.* [11] to span four years. This study is the first of its kind to incorporate the 2019 UK general election. In the political sphere, findings are affected by the events of the time, which unfold over years, making it hard to draw general conclusions. By situating the exploration of the four factors in this new recent context, yet drawing on earlier work, we are able to make a much stronger case than can be achieved with an isolated, shorter study.

Women and ethnic minority MPs say that they receive worrying abuse [14], and abuse toward women has emerged as a topic of particular concern [8, 15–17]. Pew [18] find that women are twice as likely as men to receive sexist abuse online, and are also more likely to perceive online abuse as a serious problem. Gorrell *et al.* [6] present findings for the first three quarters of 2019, with an emphasis on racial and religious tensions in UK politics. They find ethnic minority MPs, in addition to receiving more racist abuse, also receive more sexist abuse, and that women receive more sexist abuse. Rheault *et al.* [8] find incivility toward women politicians increases with visibility, which they suggest relates to the extent of gender norm violations. Broadly speaking, the emerging picture is one in which women in politics are generally treated somewhat more politely than men, but within that, subjected to a lesser but more sinister volume of misogyny specific to them, and that women reasonably feel more distressed by the abuse they receive. In this context, our quantitative contribution on the subject of how receiving abuse interacts with choosing to stand for re-election is pertinent.

Gorrell *et al.* [2, 19] highlight the greater vocality of the pro-Brexit group on Twitter during the 2016 EU membership referendum, a group that tends to be associated with the (centre right) Conservative party. Similarly Vidgen *et al.* [20] explore UK far-right Islamophobia on Twitter. Yet Gorrell *et al.* [6] find that generally speaking, UK Conservative politicians are attracting more abuse on Twitter, a finding supported by earlier work [11]. Such findings show that no one side of the political spectrum is silenced in the UK. However, in other countries, political intimidation can be a much bigger problem. Chaturvedi [21] describes India's state-sponsored online intimidation operation from the inside. A body of work describes how online intimidation is used to silence political opponents in various countries (e.g. [22, 23]).

The quantitative work presented here depends on automatic detection of abuse in large volumes of Twitter data. A significant amount of work exists on the topic of automatic abuse detection within the field of natural language processing, often in the context of support for platform moderation. Schmidt and Wiegand [24] provide a review of prior work and methods, as do Fortuna and Nunes [25]. Whilst unintended bias has been the subject of much research in recent years with regards to making predictive systems that do not penalize minorities or perpetuate stereotypes, it has only just begun to be taken up within abuse classification [26]; unintended bias, such as an increased false positive rate for certain demographics, is a serious issue for sociological work such as ours. For that reason and others we adopt a rule-based approach here, as discussed below. More broadly, a biased dataset is one in which it is possible to learn classifications based on features that are unconnected to the actual task. Wiegand *et al.* [27] share performance results for several well known abuse detection approaches when tested across domains, giving a more accurate impression of the state of the art with regards to actual abuse detection.

### 3 Corpus and methods

Our work investigates a large tweet collection on which a natural language processing has been performed in order to identify abusive language, the politicians it is targeted at and the topics in the politician's original tweet that tend to trigger abusive replies, thus enabling large scale quantitative analysis. It includes, among other things, a component for MP and candidate recognition, which detects mentions of MPs. Topic detection finds mentions in the text of political topics (e.g. environment, immigration) and subtopics (e.g. fossil fuels). The list of topics was derived from the set of topics used to categorise documents on the gov.uk website [28], first seeded manually and then extended semi-automatically to include related terms and morphological variants using TermRaider [29], resulting in a total of 1046 terms across 44 topics. This methodology is presented in more detail by Greenwood *et al.* [13], with supporting materials also available online, as indicated at the end of the paper. However abuse detection has been extended since previous work, and is therefore explained in the next section.

#### 3.1 Identifying abusive texts

A rule-based approach was used to detect abusive language. An extensive vocabulary list of slurs, offensive words and potentially sensitive identity markers forms the basis of the approach. The slur list contained 1081 abusive terms or short phrases in British and American English, comprising mostly an extensive collection of insults, racist and homophobic slurs, as well as terms that denigrate a person's appearance or intelligence, gathered from sources that include <http://hatebase.org> and Farrell *et al.* [30].

Offensive words such as the "F" word don't in and of themselves constitute abuse, but worsen abuse when found in conjunction with a slur, and become abusive when used with an identity term such as "black", "Muslim" or "lesbian". Furthermore, a sequence of these offensive words in practice is abusive. 131 such words were used; examples include "f\*\*king", "sh\*t" and "fat". Similarly, identity words aren't abusive in and of themselves, but when used with a slur or offensive word, their presence allows us to type the abuse. 451 such words were used. Word lists are available online, as discussed in "Availability of data and materials" below.

On top of these word lists, 53 rules are layered, specifying how they may be combined to form an abusive utterance as described above, and including further specifications such as how to mark quoted abuse, how to type abuse as sexist or racist, including more complex cases such as "stupid Jew hater" and what phrases to veto, for example "polish a turd" and "witch hunt", that a naive application of the lists would find abusive. Making the approach more precise as to target (whether the abuse is aimed at the politician being replied to or some third party) was achieved by rules based on pronoun co-occurrence. In the best case, a tight pronoun phrase such as "you idiot" or "idiot like her" is found, that can reliably be used to identify whether the target is the recipient of the tweet or a third party. Longer range pronoun phrases are less reliable but still useful. However, large numbers of insults contain no such qualification and are targeted at the tweet recipient, such as for example, simply, "Idiot!". Unless these are plurals, we count these. The approach is generally successful, but where people make a lot of derogatory comments about a third party in their replies to a politician, for example racist remarks about others, there may be a substantial number of false positives.

Data from Kaggle's 2012 challenge, "Detecting Insults in Social Commentary" [31], was used to evaluate the success of the approach. The training set was used to tune the terms

included. On the test set, our approach was shown to have an accuracy of 80%, and a precision/recall/F1 of 0.72/0.47/0.57. This precision is considered sufficient for empirical work (being greater than 0.7 [32]). However there is a long tail of linguistically more complex abuse that is hard to identify with sufficient precision, and therefore recall is low. As a rule of thumb, the method finds about half of the abuse. Therefore the results can be seen as an indicator of a more pervasive problem.

To compare this to the current state of the art, we refer to Wiegand *et al.* [27], who demonstrate that data-driven classification approaches leverage bias in the dataset to obtain an inflated result. The median F1 they find for a set of well-known systems, tested across domains to reduce this bias, is 0.617, showing that our performance is in keeping with the current state of the art. Furthermore our approach carries a much reduced risk of unwanted bias, such as more false positives for ethnic minorities or women [33–35], that might reduce confidence in the findings presented here, since we don't use indiscriminate features.

The resulting system is publicly available, as discussed below under “Availability of data and materials”.

### 3.2 Collecting tweets

The corpus was created by collecting tweets in real-time using Twitter's streaming API. We began immediately to collect any candidate who had been entered into Democracy Club's database [36] who had Twitter accounts. Some of these are members of the previous parliament who are standing for re-election, or for other reasons are well-known politicians, and others were not members of the previous parliament, and possibly have little in the way of a previous public profile. We used the API to follow the accounts of all candidates over the period of interest. This means we collected all the tweets sent by each candidate, any replies to those tweets, and any retweets either made by the candidate or of the candidate's own tweets. Note that this approach does not collect all tweets which an individual would see in their timeline, as it does not include those in which they are just mentioned. We took this approach as the analysis results are more reliable due to the fact that replies are directed at the politician who authored the tweet, and thus, any abusive language is more likely to be directed at them. Data were of a low enough volume not to be constrained by Twitter rate limits.

### 3.3 Corpus

A total of 4,192,027 tweets were collected for the 2019 six-week period, comprising 184,014 original (authored) tweets from 2581 individual politicians, 334,952 retweets by them, 131,292 replies by them and 3,541,844 replies *to* them. Table 1 gives the party and gender breakdown of the individuals. We can see that aside from in the Labour party, men were better represented.

Additionally we utilised matching data collected in the campaign period of the 2017 UK general election. Corpus statistics for that election can be found in Gorrell *et al.* [11]. The data was analysed using the same (updated since then) version of the abuse detection application described in this work.

## 4 Findings

We proceed with an overview of the election, providing a landscape of environmental factors affecting all candidates.

**Table 1** Sample statistics for the individuals studied. All candidates for the UK 2019 parliamentary election are included, where we were able to obtain a Twitter account for them

Party	Female	Male	Non-binary	Total	Not male
Labour Party	310	275	1	586	53.07
Conservative Party	162	367	0	529	30.62
Liberal Democrats	168	352	1	521	32.44
Green Party	150	211	2	363	41.87
The Brexit Party	43	153	0	196	21.94
Independent	23	86	0	109	21.10
Scottish National Party	20	39	0	59	33.9
Plaid Cymru	8	22	0	30	26.67
Other	50	138	0	188	26.60
Total	934	1643	4	2581	36.38

**Table 2** Number of tweets, retweets and replies by candidates are given, alongside replies received, of which abusive, and the percentage thereof. The first line shows findings for the six week election campaign period. The second line contrasts this with the corresponding period for the 2017 UK general election

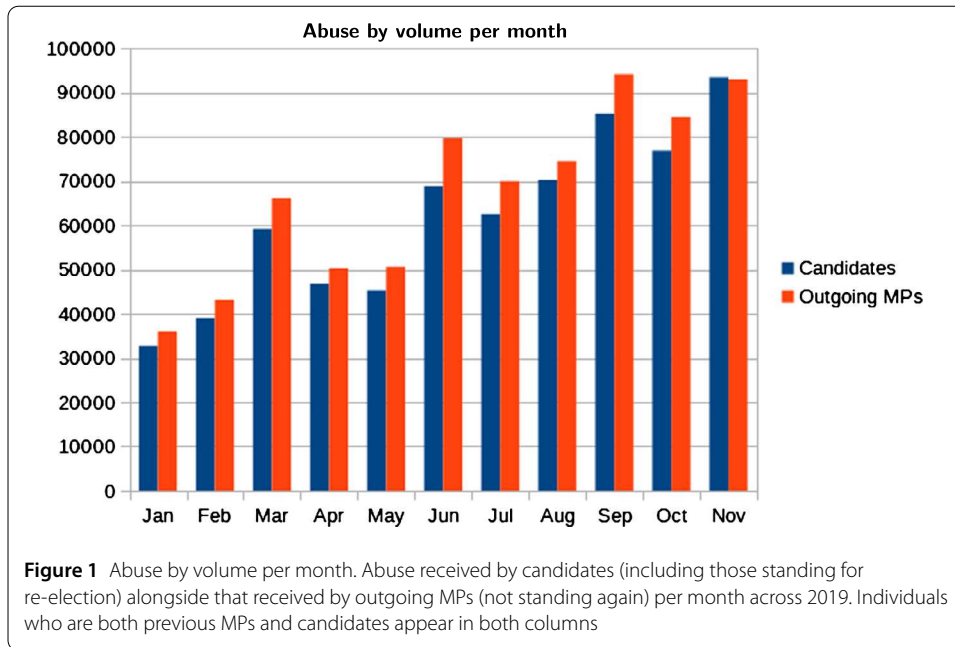
Period	Original MP tweets	MP retweets	MP replies	Replies to MPs	Abusive replies to MPs	% abusive
3 Nov–15 Dec 2019	184,014	334,952	131,292	3,541,769	157,844	4.46
29 Apr–9 Jun 2017	126,216	245,518	71,598	961,413	31,454	3.27

#### 4.1 Online abuse toward politicians is increasing

Table 2 gives overall statistics of the corpus. 3,541,769 replies to politicians were found, of which abuse was found in 4.46%. The second row gives statistics for the matching 2017 general election period, allowing comparison to be made. It is evident that the level of abuse received by political candidates has risen in the intervening two and a half years, a change which is statistically significant ( $p < 0.001$ , Fisher's exact test).

In Fig. 1 we show abuse received per month by individuals who are running for election in 2019, alongside that received by individuals who were MPs in the previous parliament ("outgoing", though many will be re-elected).<sup>a</sup> Outgoing MPs who are standing again therefore appear in both columns. Most abuse is received by a handful of these prominent individuals appearing in both columns, so naturally they are similar, but showing both allows us to contextualise and remove any doubt about the finding as regards the transition at end of the period. The graph shows that abuse toward politicians by volume has risen steeply across the year. As a percentage of replies received, there has been an increase of around 1% as calculated on outgoing MPs. In November, candidates who aren't previous MPs are likely to have stepped up their engagement level, and outgoing MPs who aren't standing for re-election, the opposite, making it the only month where candidates receive more abuse. In the previous months, we are counting abuse for individuals that are only "candidates" in hindsight—they didn't actually announce their candidacy until November.

Echoing the rise in percentage of abuse seen on longer timescales, across the six week campaign period we see a rising level of abuse toward candidates, as shown in Table 3. Figure 2 shows that for the majority of the period, this was due to rising abuse toward Conservative candidates, which was not echoed in responses to either Labour or Liberal Democrat candidates.<sup>b</sup> Twitter users have tended to show a bias toward the left of the political spectrum (see Gorrell *et al.* [37] for further discussion of the Twitter population) so greater hostility toward Conservative candidates is not a surprise; however it is interesting



**Table 3** Tweet statistics on a per-week basis for the six week campaign period

Period	Original MP tweets	MP retweets	MP replies	Replies to MPs	Abusive replies to MPs	% abusive
Nov 3–Nov 9	18,633	40,683	14,456	464,473	17,854	3.84
Nov 10–Nov 16	19,845	40,110	14,651	444,045	20,742	4.67
Nov 17–Nov 23	30,445	57,764	19,372	547,748	22,007	4.02
Nov 24–Nov 30	35,254	62,688	23,674	572,976	27,666	4.83
Dec 1–Dec 7	37,615	65,601	24,237	590,781	28,151	4.77
Dec 8–Dec 14	42,222	78,106	34,902	921,746	41,421	4.49

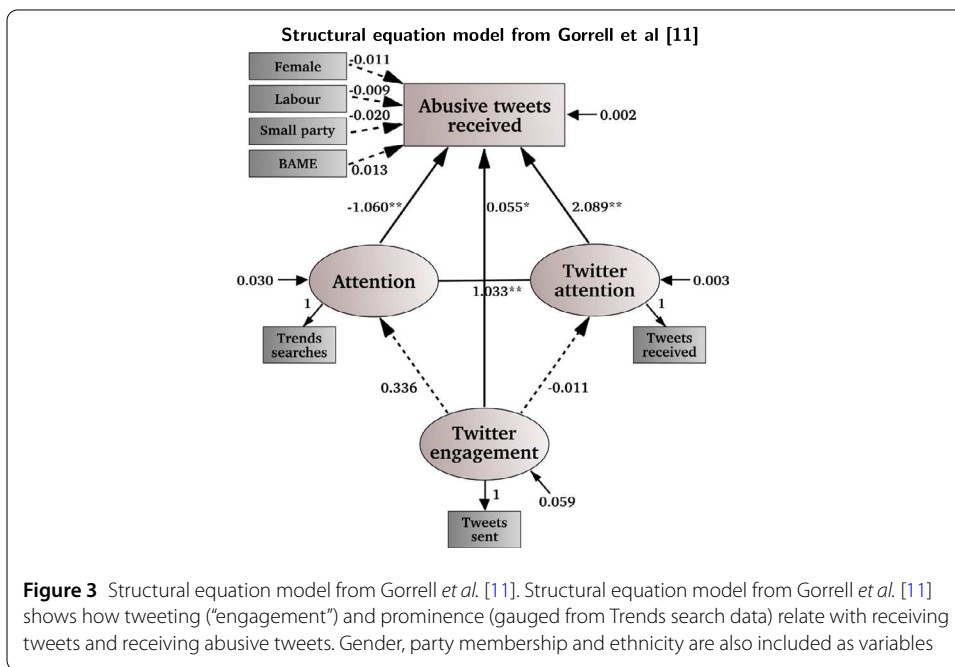
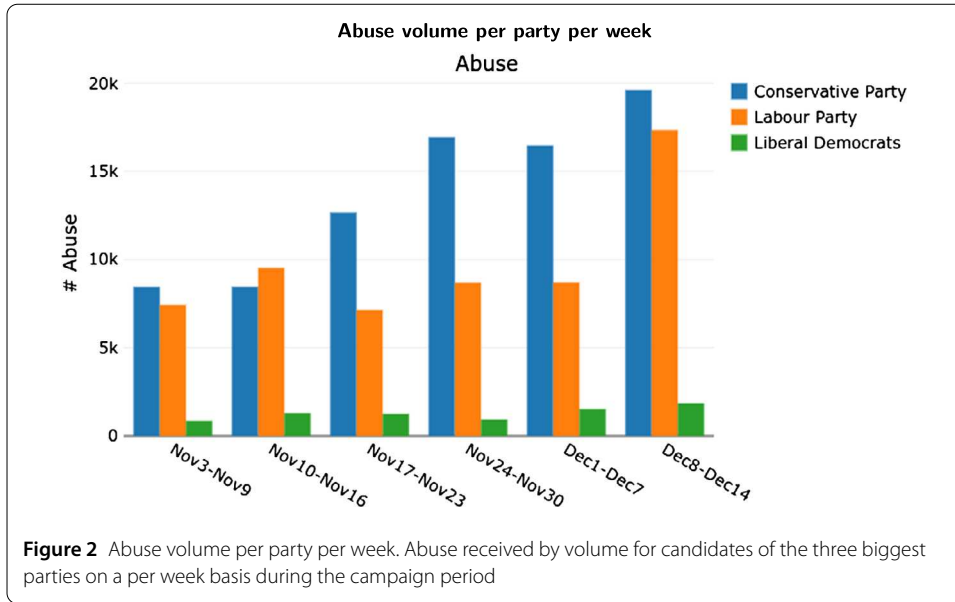
to note that this rose *in response to the campaign*, suggesting an event-driven component. The usual background up to that point had been more abuse by volume toward Labour leader Jeremy Corbyn, as shown in Gorrell *et al.* [6], despite the Twitter bias. Politically motivated abuse is discussed further below. In summary, there is compelling evidence that online abuse toward politicians continues to increase both in volume and as a percentage of replies.

#### 4.2 Exploring the four factors

We organise our understanding of the abuse an individual receives around four factors as introduced at the beginning of the paper; prominence, event surge, Twitter engagement and personal identity factors. Earlier work by Gorrell *et al.* [11] uses structural equation modelling to demonstrate that attracting attention on Twitter follows from being in the public eye, but abuse received exaggerates Twitter attention, and that engagement on Twitter relates positively with abuse received. The graph is repeated in Fig. 3 for convenience. Solid arrows indicate statistically significant relationships.

That earlier work doesn't explore the relative impact of prominence vs. events, both being subsumed under "attention", and only began to explore the effect of gender and ethnicity. Gorrell *et al.* [6] and Agarwal *et al.* [38] present investigations into the bursty, event-focused nature of online attention/abuse. Findings here and in Gorrell *et al.* [6] begin to





build up a picture of the way identity groups are treated differently. We proceed taking each factor in turn.

4.2.1 Factor 1: more prominent individuals receive disproportionately more abuse

Table 4 shows the ten most abused candidates across the period studied (November 3rd up to and including December 15th). It is evident that abuse (and indeed online attention generally) focuses itself predominantly on a handful of high profile individuals, with the remainder diminishing rapidly into a long tail of those receiving little abuse. This conforms to a Zipfian distribution [39], which exaggerates prominence.

**Table 4** Ten most abused candidates by volume

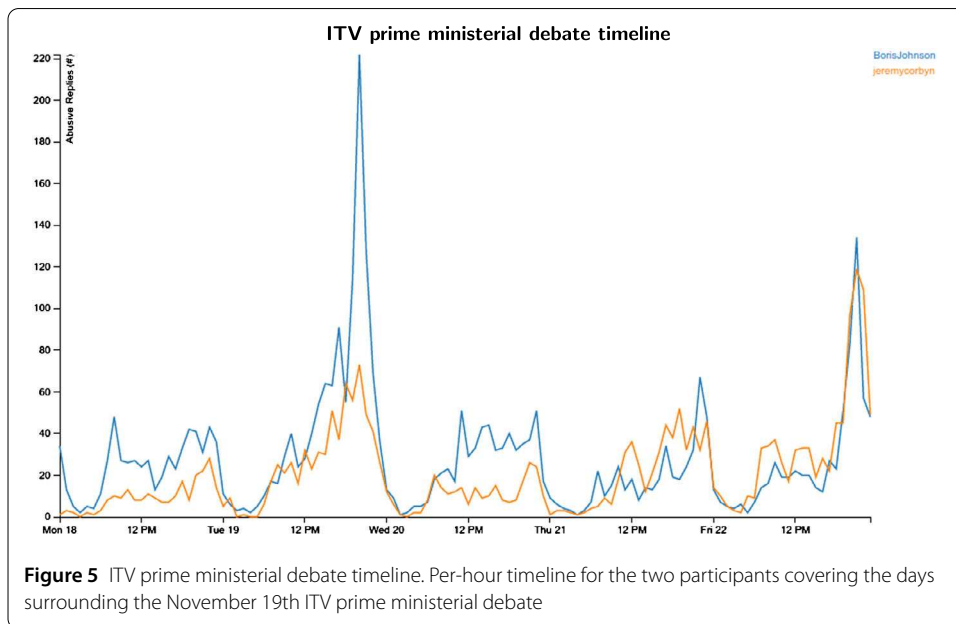
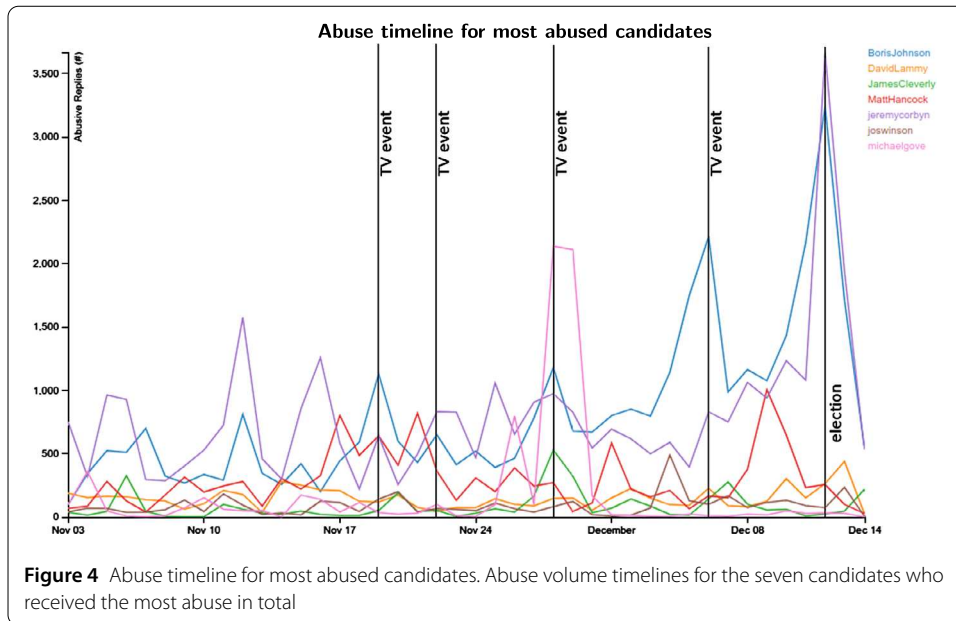
Name	Abusive replies	All replies	% abusive	% of total
Boris Johnson	34,256	565,396	6.06	21.70
Jeremy Corbyn	33,782	636,630	5.31	21.40
Matthew Hancock	12,156	186,543	6.52	7.70
Michael Gove	7255	82,240	8.82	4.60
David Lammy	6261	106,594	5.87	3.97
Jo Swinson	3819	110,533	3.46	2.42
James Cleverly	3571	58,856	6.07	2.26
Jacob Rees-Mogg	3342	48,311	6.92	2.12
Sajid Javid	3082	57,712	5.34	1.95
Diane Abbott	2262	52,279	4.33	1.43

Furthermore, Gorrell *et al.* [6] suggest that abuse may be more exaggeratedly disproportionate than the already highly concentrated distribution of online attention (as evidenced by, for example, replies received). The Pearson correlation coefficient comparing percentage of abusive replies received against total replies received in our data shows that prominence (as gauged by replies received) correlates positively and highly significantly with percentage of abuse ( $p < 0.001$ ) with a correlation coefficient of 0.10. In other words, a prominent politician can expect a much greater proportion of their Twitter replies to be abusive, compared with a less well-known politician, who even after factoring out the difference in volume of replies, is receiving a much more supportive response online.

#### 4.2.2 Factor 2: events lead to abuse surges

In Fig. 4 we see the timeline up to and including December 14th for the seven candidates who received the most abuse by volume. The two main party leaders, Jeremy Corbyn and Boris Johnson, received the most abuse by far, as shown above. There is somewhat of an increase across the period in abuse toward Mr Johnson. Furthermore, prominent Conservatives also receive significant levels of abuse. Michael Gove, a cabinet member previously associated with the “Brexit” transition process, receives a prominent spike around the time of the climate debate. Mr Johnson receives the highest spike before the election at the time of the BBC Prime Ministerial Debate on December 6th, echoing a pattern discussed below where television appearances lead to a spike in Twitter abuse toward Mr Johnson but less so toward Mr Corbyn. It is clear that the general pattern is for abuse to arrive in “spikes” [6, 38], and that events such as television appearances influence these spikes. Other peaks arise from engagement on the part of candidates. On November 12th, both Mr Johnson and Mr Corbyn made a number of opinionated tweets emphasising their priorities and the dangers of the opposition. Mr Johnson also shared the Conservative Party’s first broadcast. Health Secretary Matthew Hancock engaged in critical dialogue with Mr Corbyn on Twitter in mid-November on the subject of the national health service. Mr Hancock’s December 9th peak arises from a tweet in which he accused Labour activists of “aggressive intimidation”. Abusive responses questioned his credibility, in a way that seemed less likely to occur had a Labour politician made the same complaint. The subject of how remarks are coloured by the politics of the person who made them is taken up below.

Figure 5 gives an hour-by-hour timeline of the two party leaders for the days surrounding the first television debate of the campaign, and shows a common pattern with later campaign events, with a spike at the time of the actual event, particularly for Mr Johnson.

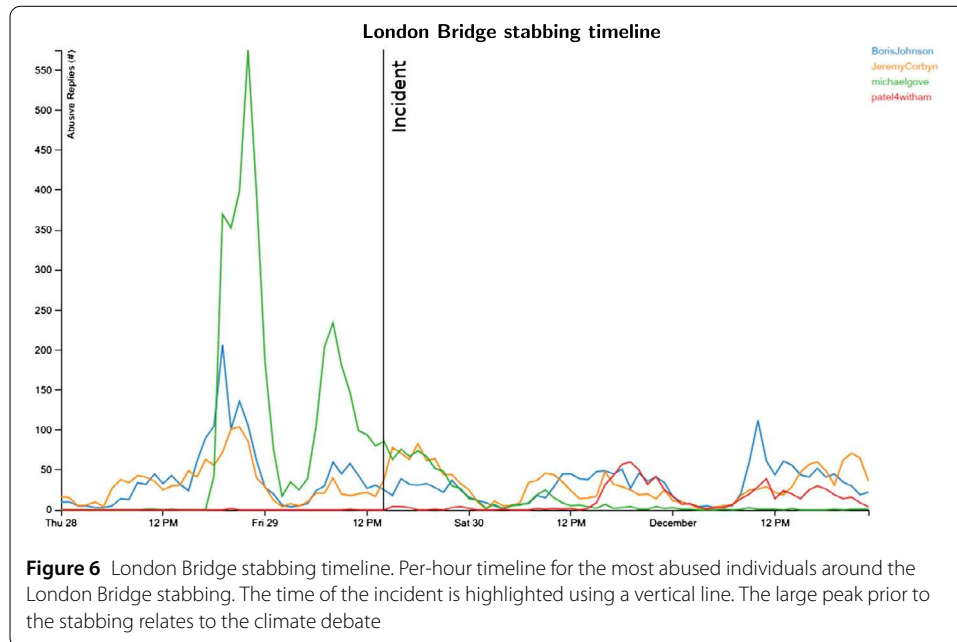


(Further event timelines can be found online as discussed in the “Availability of data and materials” section below.)

#### 4.2.3 Factor 3: *opinionated tweets precipitate abuse*

We have seen above that high profile events in conjunction with a precipitating tweet produce the highest surges of abuse. We now focus more specifically on precipitating tweets, particularly from individuals that aren’t normally the focus for such high levels of attention.

The period of November 28th and 29th was eventful, featuring two election-related television events (the November 28th Channel 4 Climate Debate and the November 29th BBC



Election Debate) and also encompassing an Islamist attack in which two people died. The main event drawing fire on Twitter was Michael Gove’s attempt to participate in Boris Johnson’s place on Thursday, and his subsequent attitude, expressed on Twitter, about being refused, resulting in the fourth highest abuse peak of the campaign, and by far the highest peak for someone who isn’t either Mr Johnson or Mr Corbyn. Both Mr Gove and Mr Johnson drew more abuse on Twitter on Thursday night than any of the actual participants, who did not particularly come under fire.

The timeline in Fig. 6 shows that the large peak in abuse toward Michael Gove around the time of the Thursday night climate debate, and the next-morning “echo”, dwarf any response to the London Bridge stabbing the following afternoon. Qualitative investigation of the data shows a great deal of personal abuse toward Mr Gove. The surge the following morning demonstrates that the peak is not just caused by people watching events on television and taking to Twitter to comment, but arises from the ensuing Twitter activity (people responding to what they see in their timelines the next day), which relates to how Twitter displays material. The tweet that Mr Gove wrote that received the most abusive replies said “Tonight I went to Channel 4 to talk about climate change but Jeremy Corbyn and Nicola Sturgeon refused to debate a Conservative #climatedebate”. The controversial factual spin may have contributed to virality.<sup>c</sup>

Although appearing small in comparison to Mr Gove’s peak, the surge in abuse toward Home Secretary Priti Patel is unusual. Two tweets drew more abuse to her than she usually receives. In the tweets she blames Labour government legislation for the release of the attacker. The abuse she received is predominantly of a general nature, but political (usually “tory \_\_\_”) and sexist (around half of which is “witch”) types appear prominently. Terrorism is a controversial subject. A tweet by Labour’s shadow home secretary Diane Abbott in this context on the theme of penal moderation is typical of tweets that lead to an abusive response for her as a highly visible left-wing representative. The theme of tolerance towards those who have committed a crime is often a divisive issue between the left and right of the political spectrum.

Whilst opinionated tweets on inflammatory subjects draw attention, especially for prominent figures, at times this doesn't explain the level of abuse received. The most abused tweet of the period as a percentage (counting only those with more than 100 abusive replies in total) was by a new candidate simply announcing his candidacy in relatively typical terms. Steve Barclay drew an extraordinary number of abusive responses with a tweet relating Brexit to football, not usually an inflammatory subject. Social factors can be a part of creating an unfavourable environment for someone's input. Where this takes the form of systematic adversity for an entire demographic group, democracy is compromised. We explore this in the next section.

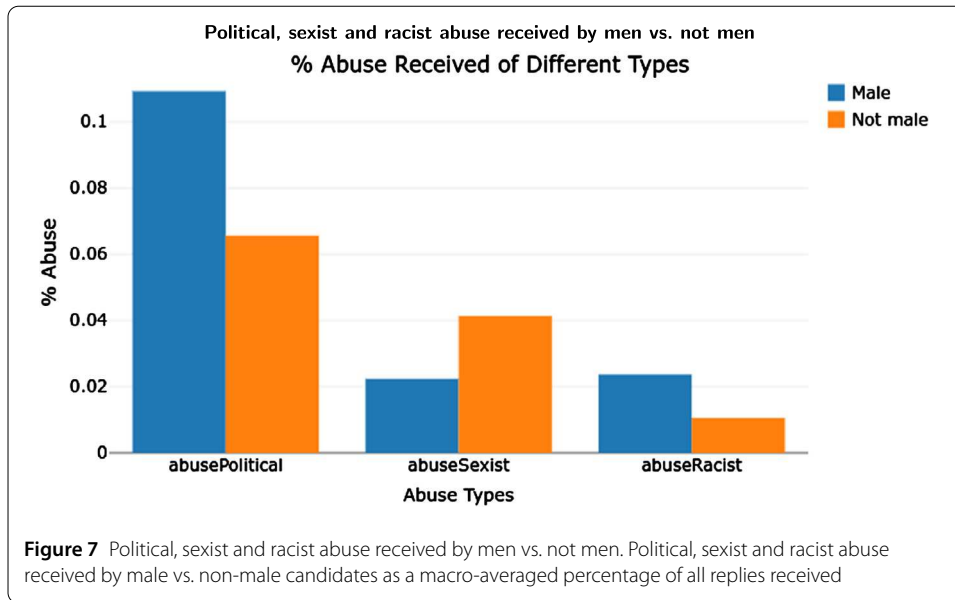
#### 4.2.4 Factor 4: *personal characteristics affect abuse*

We continue an established line of research here regarding the way identity groups are treated differently. Our work serves to highlight the ways in which discrimination was alive and well in the 2019 UK general election, albeit the short duration of the study period limits findings to only the strongest effects. Ethnic and religious minorities are underrepresented and therefore it is not possible to acquire reliable statistics for so short a time period comparing their experience to white candidates, but it is possible to see how women/non-gender-conforming candidates' experiences differ from men's, these being a larger (combined) sample. We also discuss discrimination on the grounds of politics in this section.

Although we don't present findings about the experience of ethnic minorities here, but instead refer the reader to earlier work [6], we do make use of the ability of our approach to identify racist abuse. To give some impressions of this, the racist abuse we detect targets a wide range of races/nationalities, with the British, the English and the white receiving a substantial proportion. Abuse towards the majority tends to be experienced as less offensive, making it more socially acceptable and therefore more frequent. Furthermore some of the racist abuse we detect forms part of a dialogue about race, and is often used to make a point. However a minority is explicit, unpleasant racism, with ethnic minorities and Muslims being the particular targets.

*Sexism* Whilst prominent individuals may receive consistently high abuse levels amounting to as much as 6 or 7% of their Twitter replies, on average male candidates received 1.28% abuse, and not male, 0.96%. This difference is not significant due to the short time period studied, but in keeping with significant results reported in Gorrell *et al.* [6] and Greenwood *et al.* [13]. Men received almost twice as much abuse focused on their politics, as Fig. 7 shows; on average they received 0.11% vs 0.07% for non-male candidates. Again, the result was not significant but in keeping with Gorrell *et al.* [6]. Men received half as much sexist abuse (0.02% vs 0.04%;  $p < 0.001$ , Mann–Whitney). Men received more racist abuse (0.02% vs 0.01%) but volumes are low, the result was not significant and Gorrell *et al.* didn't find this.

We take the opportunity to illustrate in practical terms the abuse that different groups are exposed to. Table 5 shows the most frequent abusive terms found across all types, followed by the most frequent politically abusive terms. Finally we see the most frequent sexist abusive terms.<sup>d</sup> Phrases that trip off the tongue may get to the top of these tables, whereas the long tail may contain more diverse ways of expressing a sentiment cluster that is harder for people to unite around words for. Religious and homophobic abuse are



**Table 5** Most frequently found abusive terms for all abuse types and for political and sexist abuse, alongside count for the whole corpus

All abuse terms	Count	Political abuse terms	Count	Sexist abuse terms	Count
fuck off	8342	tory scum	284	witch	224
idiot	6766	remoaner	276	stupid woman	149
twat	4337	fuck off commie	113	you stupid woman	148
coward	2649	tory bullshit	108	stupid man	108
idiots	2632	tory twat	79	you stupid man	104
scum	2324	tory cunt	60	you silly man	101
cunt	2068	tory bastards	57	silly man	88
moron	1979	bloody tories	54	you silly woman	78
piss off	1866	tory shit	50	stupid boy	69
wanker	1390	fucking tory	49	silly boy	65

too rare in the short time frame to produce interesting results (and confounded by much discussion of Boris Johnson’s quote “tank-topped bum boys”). Racism is in evidence but being rare, is better discussed in the context of a larger data sample, as in our previous work [6], where ethnic minority politicians are found to receive more racist and sexist abuse.

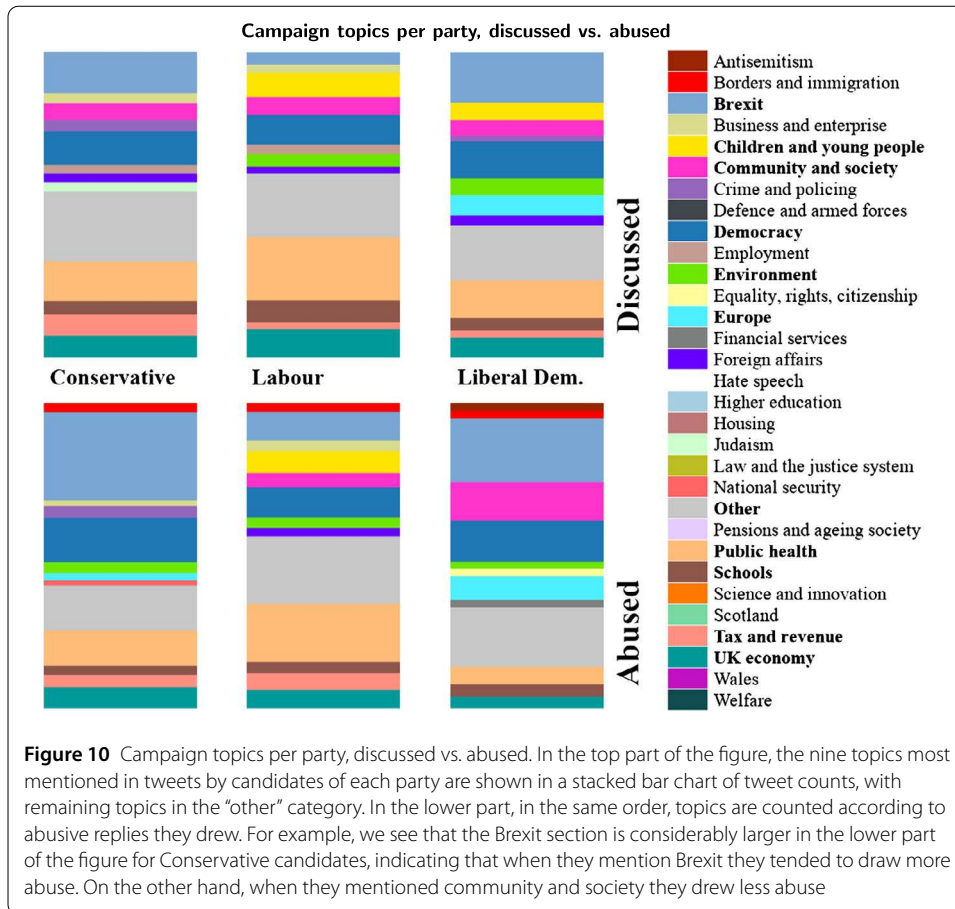
Word clouds allow us to show more terms than just the top ten.<sup>e</sup> We can see from the word cloud in Fig. 8 that sexist abuse toward men was counted and did occur in the corpus, though specifically misandristic terms are not readily available for men (equivalents for “witch”, “bint” etc., such as might be used against men by women, as opposed to between men) Excluding those with fewer than 0.2% sexist replies and fewer than 50 sexist replies overall, the candidates receiving the highest percentage of sexist abuse are given in Table 6, and are all women. Jo Swinson received the most sexist abuse of any candidate in this period, although by volume, Boris Johnson was not far behind with 351 items; for him, however, that only constituted 0.08% of all replies received.

In summary, findings support previous work suggesting that women receive more sexist abuse and men receive more general and political abuse. In other words, as a female candidate you are likely to be treated somewhat more politely online, but when you do receive





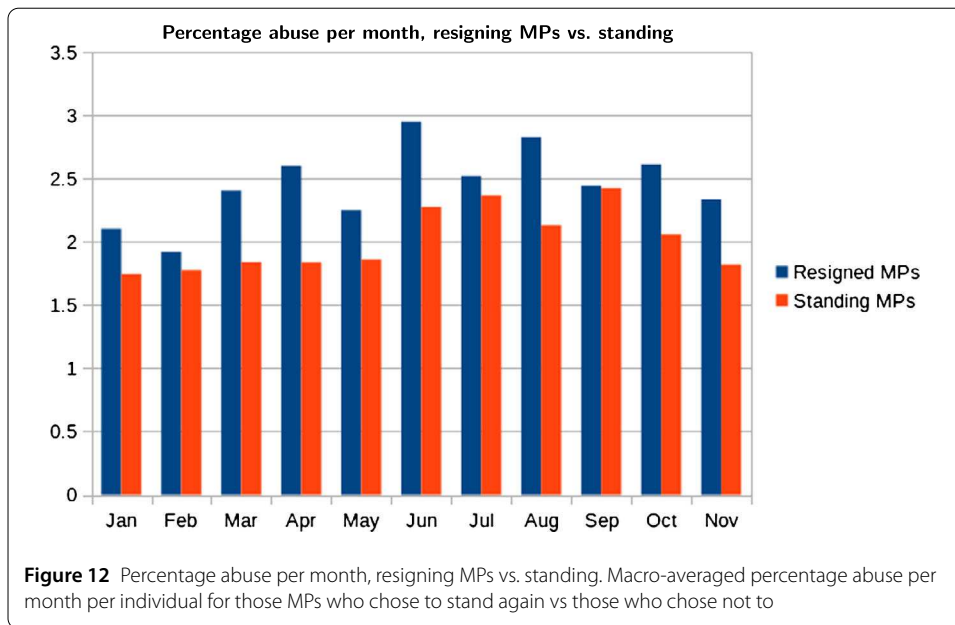
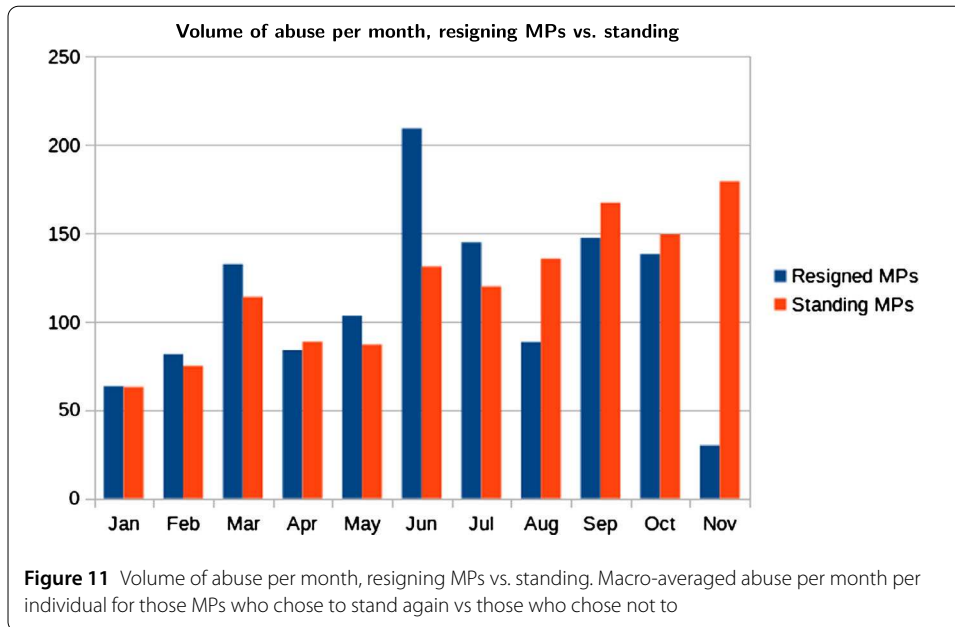




MPs cited concerns about the climate or leadership of the Labour Party. Additional to this, further MPs standing down, such as Louise Ellman, have had rocky relationships with their party, which affected their decision to stand again [40]. Their rebel status might be a contributing factor to the abuse they received. Of 76 MPs that chose not to stand again, 27 had some form of interrupted relationship with their original party. Of 21 Conservative MPs suspended from the party in September 2019, 12 chose not to stand again. Eight Labour Party MPs left to join Change UK earlier in the year, along with three Conservative MPs; of these, four chose not to stand again. Other MPs had interrupted relationships with their party for a variety of reasons, including resignations and suspension for personal conduct. 26 of 53 MPs with interruptions in their party relationship, excluding both incoming and outgoing speakers, chose not to stand again; a very much elevated proportion compared with 76 of a total 650 overall.

However several have also explicitly referred to abuse as the reason or one of their reasons for standing down, for example Nicky Morgan [41], Caroline Spelman [42], Teresa Pearce [43], Heidi Allen [44] and Mark Lancaster [45].

In Fig. 11 we directly compare average abuse per month received by MPs who chose not to stand again against those who did choose to stand again. We see that in all bar one of the earlier months of the year those individuals on average<sup>f</sup> received more abuse by volume, and particularly in June (following from new party Change UK’s lack of success in the European Parliament election). When considered as a percentage of replies received, the MPs that stood down had on average more abuse than the ones that are standing again



in every single month of the year, as shown in Fig. 12. The difference in percentage abuse is significant in a *t*-test at  $p < 0.01$ .

### 5 Conclusion

Between November 3rd and December 15th, we found 157,844 abusive replies to candidates' tweets (4.44% of all replies received)—a low estimate of probably around half of the actual abusive tweets. Overall, abuse levels climbed week on week in November and early December, as the election campaign progressed, from 17,854 in the first week to 41,421 in the week of the December 12th election. Abuse also increased month on month to a total increase of over 1% over the course of 2019. Taken alongside the 3.27% abuse found

in replies to candidates for the matching period over the 2017 UK general election, there is evidence that online abuse toward politicians is on the rise.

Twitter attention focuses disproportionately on the handful of most *prominent* politicians. Furthermore, the distribution of abuse is even more disproportionate, with the most prominent individuals receiving in excess of 6% abuse in their replies, compared with a more typical 1% for the average candidate. The “bursty”, *event-driven* nature of abuse is demonstrated here through being centred on the events of the campaign. However, within the big picture of abuse being received by prominent politicians in conjunction with prominent events, there is ample evidence of abuse varying in response to *particular tweets* (“virality”, which may constitute bullying, and reaction to opinionated tweets) and *particular types of people*. In the 2019 general election campaign, in keeping with previous research [6], men received more general and political abuse, and women received more sexist abuse.<sup>8</sup> Conservatives received more general and political abuse, as well as somewhat more sexist abuse. The personal nature of sexist abuse makes it a particular cause for concern.

We also found that MPs that chose to stand down had consistently received more abuse than MPs that chose to run again over the course of the previous year. Reasons why a person chooses to stand or not stand for election are many-factored, and causality is unclear, but the fact that we have found a positive, statistically significant relationship between being subjected to an abusive tone on Twitter and choosing to stand down as an MP should be a cause for concern. Taken together, these findings raise significant concerns regarding the increasingly unpleasant climate surrounding politics and the effect that is likely to have on political representation.

Social aspects of sexist and racist abuse make it complex to interpret, and our approach to categorising this has been cautious in defining both broadly, to include abuse towards the majority as well as minority. Despite this we have found significant grounds for concern about the discrimination politicians are subjected to. However details of the social context make a difference to the harm caused by hate speech, and an empirical study exploring how terms are used within and across groups, for example between men, or by women to men, may allow for greater specificity in future work.

#### **Funding**

This work was supported by the ESRC under grant number ES/T012714/1, “Responsible AI for Inclusive, Democratic Societies: A cross-disciplinary approach to detecting and countering abusive language online.”

#### **Availability of data and materials**

Individual tweets cannot be shared due to the sensitive nature of identifying those who send abusive tweets to public figures. However the system used to annotate the data is publicly available at <https://cloud.gate.ac.uk/shopfront/displayItem/gate-hate>. Additional information about the system can be found at <https://gate-socmedia.group.shef.ac.uk/election-analysis-and-hate-speech/ge2019-supp-mat/>. Aggregate data are available at [https://figshare.shef.ac.uk/articles/Which\\_Politicians\\_Receive\\_Abuse\\_/12340994](https://figshare.shef.ac.uk/articles/Which_Politicians_Receive_Abuse_/12340994) with a DOI of <https://doi.org/10.15131/shef.data.12340994>.

#### **Ethics approval and consent to participate**

Ethics approval was granted to collect the data through application 25371 at the University of Sheffield. All data used are in the public domain, and only public figures are identified by name in this work. Due to the sensitive nature of the data, it cannot be made public except in aggregate. Experimenter exposure to disturbing material is managed through short sessions. Readers are warned at the start of this work that they may find the language they encounter distressing.

#### **Competing interests**

The authors declare that they have no competing interests.

#### **Authors' contributions**

GG designed and performed the analysis, developed the abuse classifier and wrote the text. MEB wrote the software to extract summary data from indexed tweets. IR managed the indexing of the raw tweets into our search engine. MAG

helped to produce some of the figures. KB provided direction and guidance. All authors read and approved the final manuscript.

#### Endnotes

- <sup>a</sup> The chart of abusive tweet count per month since January naturally doesn't contain all candidates, since these were only announced in November. It is based on our previous data collection, and contains only candidates that also stood in 2017. However, considering most abuse/replies are received by the most prominent individuals, the overlap is very high. 99% of abuse in the new dataset also appears in the old one (and 95% of replies). For the calendar month of November, in the new dataset 94,566 abusive replies to candidates were found whereas the graph shows 93,516.
- <sup>b</sup> Party disparity not due to greater engagement on Twitter by Conservative candidates. Tweets authored: Wk 1, Cons 11,311, Lab 19,965; Wk 2, Cons 10,921, Lab 19,648; Wk 3, Cons 15,404, Lab 25,973; Wk 4, Cons 19,065, Lab 33,468.
- <sup>c</sup> The high response level, e.g. vs. the terror attack the following day, is possibly a form of "bikeshedding", i.e. responding to the accessible rather than the important. Online culture has also explored the compelling power of "someone on the Internet who is wrong" (e.g. see <https://www.xkcd.com/386/>).
- <sup>d</sup> Inclusion of terms is heuristic; various sources have been combined, and further terms added through observation as the system has matured over several years. Yet there may be some terms we have overlooked despite our best efforts.
- <sup>e</sup> If an abuse term of another type appears frequently with for example political abuse, it may appear in the word cloud.
- <sup>f</sup> Macro-average: the figure is calculated per individual and then averaged, to avoid prominent individuals dominating the overall result
- <sup>g</sup> The number of non-gender-conforming candidates was too small to draw conclusions from.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 18 February 2020 Accepted: 18 June 2020 Published online: 02 July 2020

#### References

1. Oppenheim M General election: women MPs standing down over 'horrific abuse', campaigners warn. <https://www.independent.co.uk/news/uk/politics/general-election-woman-mps-step-down-abuse-harassment-a9179906.html>. Accessed 2020-02-17
2. Gorrell G, Bakir ME, Roberts I, Greenwood MA, Iavarone B, Bontcheva K (2019) Partisanship, propaganda and post-truth politics: quantifying impact in online debate. *J Web Sci.* <https://doi.org/10.34962/jws-84>
3. General election 2019: conservatives 'see highest rise in Twitter abuse'. [www.bbc.co.uk/news/election-2019-50687425](http://www.bbc.co.uk/news/election-2019-50687425). Accessed 2020-02-17
4. Online Harms White Paper. <https://www.gov.uk/government/consultations/online-harms-white-paper>. Accessed 2020-02-17
5. Serving healthy conversation. [https://blog.twitter.com/en\\_us/topics/product/2018/Serving\\_Healthy\\_Conversation.html](https://blog.twitter.com/en_us/topics/product/2018/Serving_Healthy_Conversation.html). Accessed 2020-02-17
6. Gorrell G, Bakir ME, Greenwood MA, Roberts I, Bontcheva K (2019) Race and religion in online abuse towards UK politicians. Preprint. [arXiv:1910.00920](https://arxiv.org/abs/1910.00920)
7. Heilman ME, Wallen AS (2010) Wimpy and undeserving of respect: penalties for men's gender-inconsistent success. *J Exp Soc Psychol* 46(4):664–667
8. Rheault L, Rayment E, Musulan A (2019) Politicians in the line of fire: incivility and the treatment of women on social media. *Res Polit.* <https://doi.org/10.1177/2053168018816228>
9. Theocharis Y, Barberá P, Fazekas Z, Popa SA, Parnet O (2016) A bad workman blames his tweets: the consequences of citizens' uncivil Twitter use when interacting with party candidates. *J Commun* 66(6):1007–1031
10. Binns A, Bateman M (2018) And they thought papers were rude. *Br Journal Rev* 29(4):39–44
11. Gorrell G, Greenwood MA, Roberts I, Maynard D, Bontcheva K (2018) Twits, twats and twaddle: trends in online abuse towards UK politicians. In: Twelfth international AAAI conference on web and social media
12. Ward SJ, McLoughlin L et al (2017) Turds, traitors and tossers: the abuse of UK MPs via Twitter. Paper presented at the ECPR Joint Sessions 2017
13. Greenwood MA, Bakir ME, Gorrell G, Song X, Roberts I, Bontcheva K (2019) Online abuse of UK MPs from 2015 to 2019. Preprint. [arXiv:1904.11230](https://arxiv.org/abs/1904.11230)
14. Akhtar S, Morrison CM (2019) The prevalence and impact of online trolling of UK members of Parliament. *Comput Hum Behav* 99:322–327
15. Southern R, Harmer E (2019) Twitter, incivility and "everyday" gendered othering: an analysis of tweets sent to UK members of Parliament. *Soc Sci Comput Rev.* <https://doi.org/10.1177/0894439319865519>
16. Stambolieva E (2017) Methodology: detecting online abuse against women MPs on Twitter. Technical report, Amnesty International. <http://apo.org.au/node/107651>
17. Delisle L, Kalaitzis A, Majewski K, de Berker A, Marin M, Cornebise J (2019) A large-scale crowdsourced analysis of abuse against women journalists and politicians on Twitter. Preprint. [arXiv:1902.03093](https://arxiv.org/abs/1902.03093)
18. Pew (2017) Online Harassment 2017. <https://www.pewinternet.org/2017/07/11/online-harassment-2017/>. Accessed 2019-08-20
19. Gorrell G, Roberts I, Greenwood MA, Bakir ME, Iavarone B, Bontcheva K (2018) Quantifying media influence and partisan attention on Twitter during the UK EU referendum. In: International conference on social informatics, pp 274–290. Springer, Berlin

20. Vidgen B, Yasseri T, Margetts H (2019) Trajectories of Islamophobic hate amongst far right actors on Twitter. Preprint. [arXiv:1910.05794](https://arxiv.org/abs/1910.05794)
21. Chaturvedi S (2016) I am a troll: inside the secret world of the BJP's digital army. Juggernaut Books, New Delhi
22. Ibrahim NA (2018) Everyday authoritarianism: a political anthropology of Singapore. *Crit Asian Stud* 50(2):219–231
23. Saka E (2018) Social media in Turkey as a space for political battles: AKTrolls and other politically motivated trolling. *Middle East Crit* 27(2):161–177
24. Schmidt A, Wiegand M (2017) A survey on hate speech detection using natural language processing. In: Proceedings of the fifth international workshop on natural language processing for social media, pp 1–10. Association for Computational Linguistics, Stroudsburg
25. Fortuna P, Nunes S (2018) A survey on automatic detection of hate speech in text. *ACM Comput Surv* 51(4):85
26. Park JH, Shin J, Fung P (2018) Reducing gender bias in abusive language detection. Preprint. [arXiv:1808.07231](https://arxiv.org/abs/1808.07231)
27. Wiegand M, Ruppenhofer J, Kleinbauer T (2019) Detection of abusive language: the problem of biased datasets. In: Proceedings of the 2019 conference of the North American Chapter of the Association for Computational Linguistics: human language technologies, volume 1 (long and short papers), pp 602–608
28. Welcome to GOV.UK. <https://www.gov.uk/government/policies>. Accessed 2020-02-17
29. TermRaider. <https://gate.ac.uk/projects/arcodem/TermRaider.html>. Accessed 2020-02-17
30. Farrell T, Fernandez M, Novotny J, Alani H (2019) Exploring misogyny across the manosphere in Reddit
31. Detecting insults in social commentary. <https://www.kaggle.com/c/detecting-insults-in-social-commentary/data>. Accessed 2020-02-17
32. Rijsbergen CJ (1979) Information retrieval, 2nd edn
33. Hardt M, Price E, Srebro N et al (2016) Equality of opportunity in supervised learning. In: Advances in neural information processing systems, pp 3315–3323
34. Bolukbasi T, Chang K-W, Zou JY, Saligrama V, Kalai AT (2016) Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In: Advances in neural information processing systems, pp 4349–4357
35. Caliskan A, Bryson JJ, Narayanan A (2017) Semantics derived automatically from language corpora contain human-like biases. *Science* 356(6334):183–186
36. Democracy Club. <https://democracyclub.org.uk>. Accessed 2020-02-17
37. Gorrell G, Bakir ME, Temple L, Maynard D, Cifariello P, Harrison J, Kanai JM, Bontcheva K (2019) Local media and geo-situated responses to Brexit: a quantitative analysis of Twitter, news and survey data. Preprint. [arXiv:1902.06521](https://arxiv.org/abs/1902.06521)
38. Agarwal P, Sastry N, Wood E (2019) Tweeting MPs: digital engagement between citizens and members of Parliament in the UK. In: Proceedings of the thirteenth international AAAI conference on web and social media (ICWSM 2019), pp 26–37
39. Zipf's law. [https://en.wikipedia.org/wiki/Zipf%27s\\_law](https://en.wikipedia.org/wiki/Zipf%27s_law). Accessed 2020-02-17
40. Belam M, Walker P, Levitt D Which MPs are standing down at the 2019 general election? <https://www.theguardian.com/politics/2019/oct/31/which-mps-are-standing-down-at-the-2019-general-election>. Accessed 2020-05-21
41. Morgan N For the first time in 18 years. <https://twitter.com/NickyMorgan01/status/1189625485124354048>. Accessed 2020-05-21
42. Dame Caroline Spelman standing down over 'Brexit abuse'. <https://www.bbc.co.uk/news/uk-england-birmingham-49593559>. Accessed 2020-05-21
43. Crockett M "I fear this has broken me": how worried should we be about abuse pushing women out of politics? <https://www.stylist.co.uk/long-reads/women-mps-standing-down-uk-online-abuse-election/325744>. Accessed 2020-05-21
44. Proctor K Lib Dem MP Heidi Allen to stand down at next general election. <https://www.theguardian.com/politics/2019/oct/29/lib-dem-mp-heidi-allen-stand-down-next-general-election>. Accessed 2020-05-21
45. Mark Lancaster to stand down as Milton Keynes North MP. <https://www.bbc.co.uk/news/uk-england-beds-bucks-herts-50275204>. Accessed 2020-05-21

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---