Turds, Traitors and Tossers: The Abuse of UK MPs via Twitter

Liam McLoughlin and Stephen Ward
Culture, Communication and Media Research Centre
School of Arts and Media
University of Salford

Draft paper please do not quote without authors permission

Paper to be presented at the European Consortium of Political Research Joint Sessions, University of Nottingham, Nottingham, April 25-29th, 2017.

I. INTRODUCTION

The murder of the MP Jo Cox in June 2016 drew attention to the abuse, threats and violence directed towards political representatives in democracies. Though there was no evidence that Cox's killer had directly engaged in online harassment, it drew to attention to the experiences of abuse towards MPs via social media platforms. Twitter threats, impersonations and trolling were seen as an increasing matter of concern, as a number of MPs highlighted the almost daily barrage of abusive and threatening messages. Besides the obvious threat to MPs themselves, social media sites have been accused of facilitating a coarsening of democratic debate and allowing for the normalization of abuse, particularly against female representatives. At one level, social media abuse could be seen as the latest chapter in a long history of attacks on politicians – some have even suggested that abuse is simply a reflection of political anger and that any attempts to control social media platforms are a means of protecting politicians from the strength of public opinion (O'Neill, 2015). Alternatively, online abuse has been seen as symptom of apparently increasing levels of political polarisation in western democracies. The Deputy Speaker of the House of Commons, Lindsay Hoyle, has argued recently that it, [online abuse], undermines democracy by potentially restricting debate on emotive issues for fear of abuse by participants (Home Affairs Select Committee, 2017). At the same Parliamentary hearing, The Head of Parliamentary Security at Westminster stated:

> I do not think we are at the tail end at all. I think it is getting more publicity now and people are more aware of it... I think it is at a high level. In my assessment, it will go up from that (Home Affairs Select Committee, 21 March 2017)

Despite the obvious political concerns and media attention, there has been relatively little specific research on the phenomenon and few attempts to actively quantify the extent of the problem. Anecdotal media reporting suggests high and increasing levels of abuse on Twitter in particular. This research, therefore, attempts to provide a benchmark through quantifying the extent, scale and nature of the abuse of MPs via Twitter, as well as exploring some of the causes. Additionally, we have sought to gather evidence on those engaged in abuse – are they the stereotypical spotty-youth loners in their bedrooms? Or, are there common patterns, networks and profiles to abusers. In order to assess these questions, we gathered a dataset of over 270,000 tweets sent directly to the 573 UK MPs with Twitter accounts over a two and half month period (November 2016-January 2017). These were then sifted to identify around 7,000 tweets that contained abusive messages or hate speech (see definitions discussed below) that was further analysed to create an overview of the phenomena and also to track patterns in both abuse and hate speech.

II. MPS, ABUSE AND SOCIAL MEDIA

Given that social media abuse is a relatively new phenomenon it is perhaps not surprising that there is a relatively limited pool of previous academic research and literature. Although tracing media reports of MP abuse does provide some background evidence and context as we shall see. In analysing the rather disparate academic literature, four explanatory themes emerge: long standing concerns with mental illness; the nature of social media technologies and their associated communication styles; politically based explanations which see abuse stimulated by a growth of political polarization or extremism around certain emotive issues; and, finally, explanations which view abuse in terms of broader social problems around identity issues (race, religion and gender).

Psychological explanations: Mental illness as a long-standing factor

Whilst there has been increased recent attention on social media abuse of MPs it is worth remembering that politicians have always been on the receiving end of abuse, threats and violence; with research dating back from over one-hundred years can be shown to report on the issue (James *et al*, 2016). Although the lack of research pre-social media and the more private nature of communications make it somewhat difficult to accurately compare eras. Nevertheless, one long-standing explanation of threats and abuse towards MPs is the connection to mental illness and psychological disorders as explanations for abuse of public figures generally (James *et*

al, 2007; Dietz and Martell, 1989). The majority of these types of studies originate from psychiatry or social psychology research fields. In the main, they involve self-reporting from MPs and focus on threatening behaviour, physical attacks, and stalking rather than abuse per se. Yet, research going back to the 1980s indicates the prevalence of links between some forms of mental illness and the targeting of politicians. MPs attract abuse due to their increased public profile, along with associated perception as people with power, attracts members of the public with 'idiosyncratic personal causes or quests for justice in a manner that has attracted the term 'fixation'. (James *et al*, 2016:2).

More recent studies in a range of democracies, (UK, US, Canada, Australia, Norway, New Zealand), appear to highlight high levels of abuse and threatening behaviour including physical attacks, stalking, harassment, threats and inappropriate communications and also common links to mental disorders amongst abusers (Adams *et al*, 2009; James *et al*, 2016; Pathe *et al*, 2014; Schoeneman-Morris *et al*, 2007). However, one of these studies, (from New Zealand), notes that whilst MPs are reporting high levels of social media abuse/threats, compared to face-to-face meetings it was more difficult to determine whether mental illness was a factor. Furthermore, MPs tended to believe that abuse via social media was prompted by political disenchantment and less likely to lead to violence (Every-Palmer *et al*, 2015). Other studies have suggested a more mundane factor in online abuse - boredom and a desire to attract attention afforded by anonymity rather than a political or serious threat to endanger (Buckels, *et al*, 2014; Shachaf & Hara, 2010). This suggests that while online abuse is unpleasant and threats are often made, these are an aspect of online abuse, or trolling, to a well-known figure, rather than a necessary indication or threat to injure.

Technological Explanations: Anonymity, Personalisation and Informalisation

A second area of studies focuses on the supposed nature of the technological platforms themselves, especially social media, and resulting changes in communication style. The Internet has often been viewed as intensifying, accelerating and even routinising the problem of abuse (Streck, 1998). This is somewhat ironic given the high, normative, hopes in the e-democracy literature (Barber 1998; Coleman, 1999, Shane 2004; Dahlberg, 2009). Initially, such technologies were seen as having the potential to foster a more continuous, inclusive, conversational and mutually understanding relationship between the represented and their representatives (Coleman, 1999, 2005; Jackson, 2003, 2008; Lilleker and Jackson, 2014; Williamson, 2009). Subsequent empirical research, has led to regular criticism that MPs are failing to use the full potential of interactive technologies, ignoring dialogue and focusing on broadcasting or political marketing thus further distancing themselves from the public (Jackson and Lilleker, 2007, 2010; Williamson 2010; Francoli and Ward, 2008). However, politicians are often cautious about nature of Internet communication and their online conversations. Virtual communication is sometimes seen as overly time-consuming against the backdrop of busy schedules (Lusoli and Ward, 2005). There is also risk of losing control of your message to opponents (Stromer-Galley, 2000). Thirdly, hostile responses and the generally low quality nature of discussion/interaction means that there is often seen to be limited incentives to engage in interaction with the public (Stromer-Galley & Wichoski, 2011; Streck, 1998).

Social media communication, especially Twitter, is seen as impacting on the quality of engagement for a number of reasons: Firstly, the relative low cost of communication, arguably means that easier to access representatives than ever before (Shulman, 2009). The downside to this is that because communication is easy and immediate, it can stimulate more emotional and less thought through engagement (Papacharissi, 2004; Rowe, 2015). Secondly, on some Internet platforms, the anonymity provided contributes to what have been referred to as disinhibition effects (Suler, 2004; Joinson, 2007) - where citizens are freed from social cues and constraints to express views publicly in ways they that they wouldn't in everyday conversation. The anonymity factor is also seemingly further supported by a sense that social media spheres are somehow not subject to same legal restraints as traditional media or face-to-face communication and that perpetrators of abuse are unlikely to face sanction. Indeed, senior UK police officers themselves have admitted that they are just "in the foothills of tackling online abuse" (*Guardian*, 21 February

2017). Thirdly, the nature of communication online, it has been argued, has become more informal and more personalized. At one level, this reflects longer-term shifts in social attitudes towards authority figures that have become less deferential and hierarchical (Norris, 1999). Social media, such as Twitter, with its relative restriction in terms of message length, lends itself towards shorter, punchier forms of communication, arguably bringing with it the risk of more direct and impolite statements. A number of studies have noted a growth in the informalisation and personalization of political communication online (Serfaty, 2010) and also a blurring of the boundaries between private and public spheres. Politicians are now often advised to seek authenticity in their online personas and also to humanize themselves through increasing levels of personalisation (Hermans and Vergeer, 2013; Karvonen, 2010). Arguably, such strategies and language also invite more personalized (negative) comments in return. For instance, Theocharis *et al*'s (2016) study of 2015 EU election candidates in a range of countries found that those engaging in dialogue and interaction actually tended to generate more negativity in return, although they couldn't determine causality.

Political explanations: the Polarisation of Debate?

A more politically focused and increasingly popular explanation of social media abuse is the notion of increasing levels of polarization and populism amongst electorates in many western democracies. This links abuse and threats to a more extreme and divided climate of political debate, which, whilst not new, has been accelerated by the information and communication environment online (Prior, 2013; Leikes *et al*, 2015; Colleoni *et al* 2015). The internet's role in polarization has been disputed (Barbara, 2015) but its apparent facilitation of divided and abusive communication sphere rests on a number of factors: Firstly, the internet has been seen as increasing the amount of partisan sources in circulation which are not subject to the same professional journalistic standards as traditional media. Secondly, increasingly voters consume such partisan content on the basis of selective exposure i.e. they are drawn to material that backs up pre-existing beliefs or coincides with their pre-existing interests. The increasing levels of media choice mean that people can more easily access news and current affairs but also screen out information deemed as uninteresting, irrelevant or disagreeable (Prior 2007). Thirdly, selective exposure is then interlinked with, people's online networks and filters. In part, this reflects automated filter bubbles which select or organize what we see online on the basis of previous browsing history. Additionally, though the argument behind polarization suggests that our online networks are largely homogenous, with like-minded people communicating with one another (McPherson, Smith-Lovin and Cook 2001). Hence, Twitter, in particular, is often seen as an echo-chamber where people of similar political outlooks spread or replicate each other's messages but are rarely challenged by alternatives viewpoints or voices. The suggested longer-term impact of increased exposure to like-minded views is the adoption of more extreme positions (Mutz and Martin 2001). This filtering is then heightened by the aforementioned anonymity reducing social and psychological inhibitions and thus stimulating some individuals to express more extreme views and/or indulge in abuse of opponents in ways that they wouldn't in the offline world (Joinson 2007).

In the UK, mainstream media has linked upsurges in general levels of abuse against MPs to the increasingly divided and oppositional nature of politics heightened by several recent and ongoing debates. The finger of blame has increasingly been pointed at the binary (yes/no) nature of two recent referenda in the UK (the 2014 Scottish Independence and the 2016 EU referenda) both of which have been viewed as significantly heightening the polarized nature of UK political debate. In the wake of Brexit vote, there were a number of high profile accounts of direct threats and hate speech mainly over social media. For example, David Lammy, (London Labour MP), reported a welter of racist abuse on social media to the Police following his call to block Brexit (*Guardian*, 4 July 2016). More recently, Anna Soubry a leading Conservative remain supporter also reported death threats (*Independent*, 2 December, 2016). From the leave perspective, Rebecca Harris (Conservative MP, Castle Point,) has claimed in relation to online abuse:

> Another shocking thing in the wake of Brexit is that nice, normally liberal-minded people—people who would profess to be progressives—also think it is reasonable to abuse 17 million

> of their fellow countrymen… In our culture, people—seriously liberal, intelligent and educated people—think they can say those things online. They turn into keyboard warriors and say things that they would never dream of saying face to face to an individual (Hansard, 7 July 2016).

Similarly, during the earlier 2014 Scottish referendum campaign concern was expressed about the virulent nature of online trolling and abuse from so called "cybernats", (online Scottish Nationalists), towards their opponents. Subsequently, similar claims were made against no campaigners (so called cyber-brits – online unionists) (*Sunday Herald*, 17 February, 2017).

Whilst polarization is often seen as heightening divisions between conservative and liberal or left and right, intra party divisions can also generate highly polarized debates. In the UK, the increasingly poisonous arguments within the Labour Party between supporters of Jeremy Corbyn and his opponents within the Parliamentary Labour Party (PLP) have led to continual claims of abuse from both sides. The parliamentary vote of bombing of Syria (December 2015) sparked several reports of offensive messages including death threats to 25 Labour MPs who supported bombing (in opposition to Corbyn's position). Challenges to Corbyn's leadership also seemed to have provoked significant abuse. For example, MP Angela Eagle's (one of Corbyn's original challengers), Facebook tribute to Jo Cox was targeted by apparently Corbyn-supporting trolls whilst Corbyn opponent Tom Blenkinsopp MP reported one abuser to the police who was subsequently revealed to be a fellow party member (*Gazette*, 1 July 2016). Meanwhile, several of Corbyn's shadow cabinet team have also been targeted by offensive impersonation sites on social media allegedly by anti-corbynites (LabourAbuse).

Overall, the House of Commons Deputy Speaker has confirmed that abuse spikes when emotive issues are discussed and when individual MPs speak out on such issues they become targets (Home Affairs Select Committee, 2017) which suggests further support for polarization effect but also a predominance of a rather reactive model of social media abuse.

Identity factors: gender and race

The more technological determinist explanations of online abuse might lead us to expect generalized and randomised abuse towards MPs. Yet, both anecdotal and research-led reports indicate that certain types of MPs are targeted for social media abuse. As the Inter-Parliamentary Union (2016: 6) briefing puts it:

> social media have become the number one place in which psychological violence – particularly in the form of sexist and misogynistic remarks, humiliating images, mobbing, intimidation and threats – is perpetrated against women parliamentarians.

Moreover, this is not simply just a question of the volume of abuse but also the nature of that abuse. In particular, much media coverage has focused on abuse and threats directed at female representatives, (especially, younger women MPs), and those from ethnic minorities. Recent evidence given to the Home Affairs Select Committee indicated that Muslim and Jewish woman were indeed the No.1. targets of abuse (Home Affairs, Select Committee, 2017).

Female MPs themselves have repeatedly reported widespread and alarming levels of threats of sexual violence and repeated harassment, as well as more general misogynistic comments (Hansard, 2016). In the UK, two men have been jailed for online threats made against MPs (Stella Creasy 2014 and Luciana Berger, 2016), whilst Jess Phillips MP revealed that she had received over 600 rape threats in one evening via Twitter (*Daily Telegraph*, 31 May 2016). A recent BBC Radio 5 survey of female MPs (from all parties) indicated that the overwhelming majority (nine out of ten) reported receiving online and verbal abuse from the public whilst a third had considered quitting as a result (*BBC News Online*, 25 January 2017).

Interestingly, and perhaps revealingly, this does not seem to be simply a UK problem. The Inter-Parliamentary Union research briefing identified global issues of abuse and harassment suffered by female representatives in wide range of countries and political systems. The research further

noted that the general abuse of female representatives tended to be heightened by three factors: age; ethnicity and length of service. Younger representatives of minority ethnic backgrounds suffered more abuse especially when first elected (IPU, 2016). Similarly, media reports from Canada, Ireland and Italy in the past year all indicate similar patterns of threats and abuse via social media platforms (*CBC News*, 2016, 2017; *the Journal.ie*, 27 June 2016; *BuzzFeed News*, 8 February 2017). Laura Boldrini, the President of the Italian Chamber of Deputies, after revealing some of the regular violent abuse she receives on social media, has stated:

> It is unacceptable that women, after numerous battles to defend our rights and gain respect, now find ourselves constantly insulted and abused online, often having to face a choice: accept this kind of humiliation or stay away from the internet (*BuzzFeed News*, 8 February 2017)

In part, the abuse of female politicians has been linked to the general high level of misogyny online (Demos, 2014, 2016). Arguably, this is then exacerbated in political context where research suggests that politics and political online discussion in a range of countries has consistently shown to be dominated by men (Stromer-Galley, 2002; Harp and Tremayne, 2006; Trammell & Keshelashvili, 2005; Albrecht, 2006; Hagemann, 2002; Jankowski and van Selm, 2000; Jensen, 2003). Some studies have argued, therefore, that directed threats against woman MPs relate to attempt to delegitimise woman politicians, restrict their rights to communicate and inhibit them from taking active part in the political arena but also sense that abusers feel threatened by high profile woman politicians speaking out (IPU, 2016).

All these approaches have pointed towards a significant increase in the abuse of MPs with social media being a key driver. Although overall abuse is a considered to have grown, certain types of threatening and violent message are also seen as becoming particularly prevalent and targeted at certain demographic groups of MPs. Yet, whilst this may, indeed, be the case, reviewing existing studies highlights several related remaining problems. Firstly, there is a limited amount of consistent and precise empirical evidence. Secondly, comparing longitudinally over time between pre-internet and Internet eras remains problematic, as prior to the emergence of social media platforms much of the communication between representatives and represented was essentially private. Thirdly, methodologically, there are issues both around definitions of abuse and the fact that recent studies tend to rely on self-reporting by MPs which are of course potentially fraught with inconsistency. Finally, there is also a need for theory-building around catalysts and drivers for social media abuse.

III. RESEARCH QUESTIONS

In light of the above discussion and given the dearth of solid data, the main objective of our study was to provide some quantifiable evidence of Twitter abuse and provide a benchmark for future empirical studies. In particular, therefore, we sought to answer a number of broad questions:

- *What is scale and extent of direct abuse of MPs via Twitter?* Whilst the assumption is that abuse via social media is widespread and growing – there is very little quantitative evidence. James *et al*'s (2016) UK study, using 2010 data, indicated 10% of MPs reporting abuse via social media but this, of course, was at a relatively early stage of social media development. Consequently, therefore, we wanted quantify both the *overall level of abuse* (volume) the *spread amongst MPs* i.e. how many MPs received abusive tweets.

- *Are there particular patterns and targets amongst MPs for abuse?* In line with expectations that some individuals and groups of MPs are targeted more than others, we sought to identify: (a) which individual MPs were targeted; (b) whether particular groups of MPs were targeted (either by party gender and ethnicity etc.). Given the level of reported

abuse aimed at woman representatives – we were particularly interested to see whether our data confirmed this.

- *What types of abuse are prevalent and are certain types of abuse directed at particular MPs?* Whilst there are standard definitions for hate speech and identifying threatening communication is arguably easier (see discussion below), we were interested in whether could effectively categorise abuse. Additionally, we wanted to how far abuse could be linked to distinct political debates (such as Brexit) and whether particular groups of MPs received more threatening types of communication.

- *Are there any patterns or profiles of twitter-politician abusers?* Whilst the stereotype of Internet trolls/abusers is of social inadequate young males with addictions to technology. Thus we sought to find out whether there were serial abusers of politicians generally, whether there were networks of Twitter abusers and again whether there is any political or demographic dimension to this..

IV. DATA & METHODS

In order to answer our research questions there were several methodological hurdles: definitional issues around what constitutes abuse; how to apply such definitions in social media environment; how to detect abuse and how to determine the boundaries of MP abuse (i.e. direct or indirect).

Defining Abusive Tweets and Hate-speech

There is considerable and complex debate surrounding the definitions of abuse and hate-speech. Each definition given for either has at least some level of subjectivity alongside a mix of cultural factors. For example, Sellers, displays two extremes in relation to the Armenian Massacre in Turkey; whereby writers in Turkey have been prosecuted for hate speech when calling the action genocide, while in Switzerland, a politician was prosecuted for denying the Armenians were victims of genocide (2016:10). This can demonstrate the geo-political nature of the debate, and that the consequences of the debate, for some, is a matter of freedom. Previous approaches to defining abuse and hate-speech have used purposely-broad definitions. Burnap & Williams (2013) allowed human coders to define what hate-speech was and, therefore, it was necessary for definitions to be fairly broad. Alternatively, some researchers allow victims to define what they would consider abusive (Farris *et al,* 2016). However, the methodological approach in this paper, required a much more robust understanding before the data set could be coded.

Definition of Abuse

Definitions of harassment or abuse generally cover a wide range of content that has the intention to insult, disturb, or insult an intended person. However, there are two ontological approaches that divide most definitions. Many definitions have their foundations either in defining abuse as an action/intention, or as an impact on a victim. For example, Lenhart *et al*'s (2016) definition is that harassment is "unwanted contact that is used to create an intimidating, annoying, frightening, or even hostile environment *for the victim*" [emphasis added]. The focus here is on abuse/harassment being felt by the victim and not the action itself. In contrast, other definitions base their conception of abuse on the *act* of abuse, whereby the action, or intention, of the message itself is the basis for abuse. This can be seen in Bartlett *et al* whereby messages on Twitter were used to find abuse, and therefore could not test if victims felt abused (2014). This highlights the issue with victim-based definitions in research is that in each case of abuse, one must know that the victim feels abused or harassed. In comparison, action-based definitions can be taken from the message itself, if the context is also considered.

Computer Science approaches provide a further method for defining abuse. Definitions from

this discipline, rely on libraries of words which can be used to identify abusive content instead of traditional definitions.[1] However, this has its own issues which make it unsuitable for this type of research. Applying a keyword based methodology to Twitter messages has proven highly unreliable (see Nobata *et al,* 2016:146). Furthermore, libraries of words do not in themselves provide context similarly to the issues of sentiment analysis (see below). So, while the computer science approach might be useful in helping to identify abusive messages, machine algorithms are not yet at the stage of automatically detecting the highly complex and subjective nature of abuse to acceptable levels of accuracy.

For the reasons, above, we used Bartlett *et al's* definition of abuse on Twitter, where abuse is defined as slur words that are tweeted directly at a specific person with the intent to cause harm or distress. This can include casual use of slurs of derogatory stereotypes, so long as it is a specific and is personal attack at "you". (2014:24). This definition while open to some interpretation, allows itself to be coded directly into the dataset. Thus, tweets which are found to contain particular offensive or profane words directed to an MP can be argued to be abusive. A greater understanding of how this can be operationalised is shown in table 1.

Definition of Hate-speech

Hate speech can be defined as an expression of hatred towards a particular group, based on protected characteristics (such as ethic background, religious identity, sexuality, or gender) especially so towards groups who do, or have, faced structural and societal disadvantage (Waldron, 2014; Ferris *et al,* 2016:5). However, as with all definitions concerned with a subjective area, there has been a significant debate about what can be considered hate-speech.

Sellers conducted a comprehensive review of the definitions of hate-speech across common vernacular, academic, and legal. In this it was found that most definitions of hate-speech contain eight common themes:

1) Insults or undesirable communication to a specific or easily identifiable group
2) Content that expresses hatred towards groups
3) Speech that causes harm
4) When the speaker intends harm or bad activity
5) The speech incites bad actions beyond the speech itself
6) The speech is either public or directed at a member of the public
7) The context makes violent response possible
8) The speech has no redeeming purpose

(Sellers, 2016: 25-30).

What is clear from this framework is that while most forms of hate-speech can be considered harassment or abuse; not all abuse is hate-speech. Furthermore, it gives a working framework for deciphering what is hate-speech without the input of this papers authors, mitigating our own personal subjectivity while managing a definition.

Using these definitions, we created a classification system for the identification of abuse on Twitter towards MPs. This simple classification put tweets towards MPs into one of four categories: Non-abusive, Not-directed, Abusive, or Hate-speech. The application of these definitions can be found in table 1.

[1] See also Luis Von Ahn's list of offensive/profane words: https://www.cs.cmu.edu/~biglou/resources/

TABLE 1: EXAMPLE OF ABUSE/HATE-SPEECH CLASSIFICATION

Classification	Reasoning	Examples
Non-abusive	Tweet contains no profanity or derogatory language	• @<MPsHandle> Can you tell me how you intend to vote in tonight's debate • I hope I get to see @<MPsHandle> tonight!
Not-directed	Tweet contains abusive slurs or hate-speech, but not directed towards the MP	• Fucking Pot holes again. @<MPsHandle> Sort it out. • That group are a bunch of whores. I hope @<MPsHandle> kicks them out.
Abusive	Abusive, profane language directed towards an MP	• @<MPsHandle> You're a Wanker • I hope that @<MPsHandle> knows he's a complete fucker.
Hate-speech	Profane or derogatory language which relates to a characteristic the MP belongs; language which implicitly or explicitly implies or encourages threats towards an MP	• @<MPsHandle> is a stupid kike. • Well that proves how much of a slut @<MPsH andle> is. • I hope @<MPsHandle> gets what's coming to him: A baseball bat.

Data Collection

The findings of this paper were generated by a dataset of **270,717** tweets sent to MPs by Twitter users. Tweets were collected over a two-and-a-half-month period from November 14, 2016 to January 28, 2017. The initial return captured over 764,523 tweets and associated meta-data. However, processing to filter out spam, empty tweets, or repeated tweets significantly reduced the number contained within the dataset. Perhaps a marginal finding from this data-cleansing is that UK MPs attract an increased amount of spam compared to the fourteen per cent reported elsewhere (Yardi *et al,* 2010). This is on par with spammer tactics on the service that target more influential accounts. Overall, we are confident that the collected dataset is representative of the tweets sent to MPs on Twitter, and when analysed provides an understanding of the scope and scale of abuse received by MPs.

For the data collection, we utilized the Twitter collection software *TAGs*[2]. This data collection tool automates the data collection process by routinely accessing Twitter's search API with a query and returning with tweets which fits the search parameters imputed alongside associated meta-data. As *TAG*s is hosted on Google's cloud based services, it presented a reliable research option with automatic notifications if any issues in the data collection arise. A potential issue with this method is that while Twitter does have tools to deal with abusive messages on the service, details about its implementation are a closely guarded secret (Ho, 2017). Therefore, it is possible that Twitter may have removed abusive tweets before they were collected. However, as displayed below, the level of abusive messages collected is ample for analysis.

[2] TAGS (or Twitter Archive Google Sheets) and is available at https://tags.hawksey.info/

The identification of tweets sent to MPs was simplified by the structure of Twitter. Users of the social media platform often self-code tweets through either hashtags, or, by directing their posts to specific intended audiences. In the case of hashtags, users can precede a word, category, or phrase by a hash symbol (#<word>) to denote that the tweet is on a particular subject, or give the tweet an envisioned purpose. To specify an immediate audience, users can include the mention function (@<*recipient handle*>) that can take the form of direct messages to the recipient, replies to a recipient's tweet, or simply mentioning them in a standard tweet. Alternatively, users can choose not to code the messages if their intended immediate audience is to be limited to others who follow them directly. For this research, the self-coding of tweets avoids some of the methodological issues present in social media research (c.f Murthy, 2017). It can be assumed if a user is directing or notifying another they are trying to communicate with them, and this can be distinguished from all other tweets through the following search term:

> To:@<MPsTwitterHandle1> OR @<MPsTwitterHandle2> OR @<MPsTwitterHandle3>…

This search term was expanded to include all MPs Twitter handles and then imported to TAGs for data collection. The MPs Twitter handles used in the search were taken from a pre-collected database of all verified Twitter accounts which included an MP's Twitter handle, age, gender, party affiliation, constituency, and account details which included age of the account, activity, and biographic information. This database of MPs account handles (n=573) was updated on a weekly basis to check for MPs who had left or joined the service throughout the data collection period. Only four MPs did not receive any Tweets during the data collection period.

Detecting Abusive Tweets and Hate-speech

Using the definition of abuse described above, we created a semi-automated process for the identification of abusive tweets and those that contained hate-speech. Fully automated classification systems have been used to classify tweets through methods such as sentiment analysis, distant supervision, or machine learning (c.f Ratkiewicz *et al*, 2011; Chen, 2012; Dinakar *et al*, 2012; Burnap & Williams, 2013). These auto-classifiers still present wide margins of error and despite becoming more accurate they are not yet perfect. Human coding of the entire dataset would be a preferable alternative but issues of human error and subjectivity could arise (Sloane, 2017:168). An optimal method would be to use human intelligence tasks (HITs) to code the data, however, this does come with significant resource requirements (Buhrmester *et al*, 2011; Sloane *et al*, 2015). To this end, a multi-layered approach was taken which used sentiment analysis, keyword identification, and manual verification to identify abusive and hateful tweets within the dataset. This approach used aspects of machine classifiers where they are accurate, and retained human coding for others.

Firstly, sentiment analysis software (*SentiStrength*) was applied to the dataset (Thelwall *et al*, 2012). This software uses a lexical approach to detect both positive and negative sentiment strength. Each tweet was ranked from most negative sentiment to least negative sentiment using two scores. These two scores were taken from the *SentiStrength* analysis: a positive sentiment score, a figure is given from 1 (no positive sentiment) to 5 (very strong positive sentiment); and for negative sentiment a figure is given from -1 (no negative sentiment) to -5 (very strong negative sentiment). Sentiment analysis systems are good at understanding the overall sentiment of a tweet, however, the system cannot detect the direction of the sentiment. For example, a tweet to an MP may be insulting, triggering a high negative sentiment, but the analysis tool cannot detect if this negativity is directed towards the MP. The system was useful for ranking all the tweets by overall sentiment, which significantly increased the speed of the second stage of analysis.

The second step was to code the top 3,000 negative tweets manually. The definition classifications of abuse and hate speech mentioned above was used to detect if the tweet was

abuse or hateful, and then defined if this tweet was directed towards the MP. These codes were then analysed through a summative content and thematic analysis; which identified, and reporting patterns of use of particular keywords (Hsieh & Shannon, 2005; Braun & Clarke, 2006). A total of 78 keywords which were abusive or hateful in content were extracted.

The third step was to filter the dataset for tweets that contained any one of the identified keywords. Two tests were the conducted on each tweet. Firstly, each tweet was then manually verified to determine if the tweet was abusive or could be considered hate-speech. If tested positive for the first test, a second test verifying if the MP was the intended recipient of the abuse or hate speech was conducted. This third step was necessary, as shown by. Bartlett *et al* who found that a majority of tweets containing offensive slurs did so in a non-abusive, non-offensive manner (2014:7).

Limitations

There are several limitations to this type of social media research. Firstly, there are limitations on the of Twitter's public search API. The data accessible through this API has restrictions on the number of calls and a cap on the amount of data at any given time (Voss, Lvov, & Thompson, 2017:244). This was mitigated using *TAGs* whereby the collection of tweets was scheduled to avoid the limits set by the Twitter API.

Secondly, it is expected that the data collection and detection methodologies will underreport on the levels of tweets that contain abuse and hate-speech. Twitter has recognised that the service does enable a significant amount of abuse and hate speech, and have taken some action to reduce the amount of these tweets on the service (Ho, 2017). The exact method of how Twitter detects and removes these tweets, either through user reports or automated methods is a secret to reduce the ability of abusers to develop a work-a-round. This could mean that some abusive messages sent to MPs may have been deleted before our collection methods could add them to our dataset. Additionally, our method for detecting abusive messages focuses on the detection of abuse and hate-speech using keywords with a high frequency of abuse and hate-speech. This could allow some abusive terms that are less prevalent to be excluded. In addition to this, users may indeed intentionally obscure offensive words (Nobata *et al,* 2016:146). For example, replacing some letters with non-alphanumeric characters such as the use of dollar signs instead of an *s*. The dataset was used with the assumption that it underreports abuse and hate-speech.

A third limitation of the research method is the reliance on textual communications, as most messages on Twitter are textual. However, previous research has indicated that a large amount of communication is through both image and video mediums (Laird, 2012). However, until services are available to research which can understand and categorise visual content through machine learning tools reliably, this analysis would instead need to be human coded, which would be extremely resource extensive.

Finally, political events inevitably lead to individual MPs receiving more hate-speech than they would otherwise attract. For example, Phillip Davies MP, Anna Soubry MP, and Tim Farron MP received significant amount of online communication over specific events or actions. It is also worth considering that the data collection period included festive holidays, which may have resulted in lower levels of abuse or hate speech towards MPs.

V. FINDINGS

Overall Scale of Abuse and Hate-speech

The first area of enquiry within this research was to detect the overall scale of abuse towards MPs on Twitter. Notable journalistic claims have been made surrounding the nature of abuse to MPs

online. So far there has not been any clear indication of the overall levels of abuse/hate speech through the Twitter service or online abuse in general (Buckels, *et al,* 2014:97). From the total coded dataset of tweets to MPs (N=270,717) it was found that the proportion tweets which were classified as 'abusive' made up 2.57% of the dataset (n=6,952). In other terms, twenty-five messages in every thousand tweets contain some type of direct abuse towards an MP.

Despite overall levels of abuse, it was expected that individual MPs would attract different levels of abuse which is reflected in the dataset. The average MP received 17.99 abusive messages, with a standard deviation of 48.9, and a range of 0 to 677 abusive messages. Suggesting that the level of abuse is dependent on other factors other than the single independent; which is they are an MP.

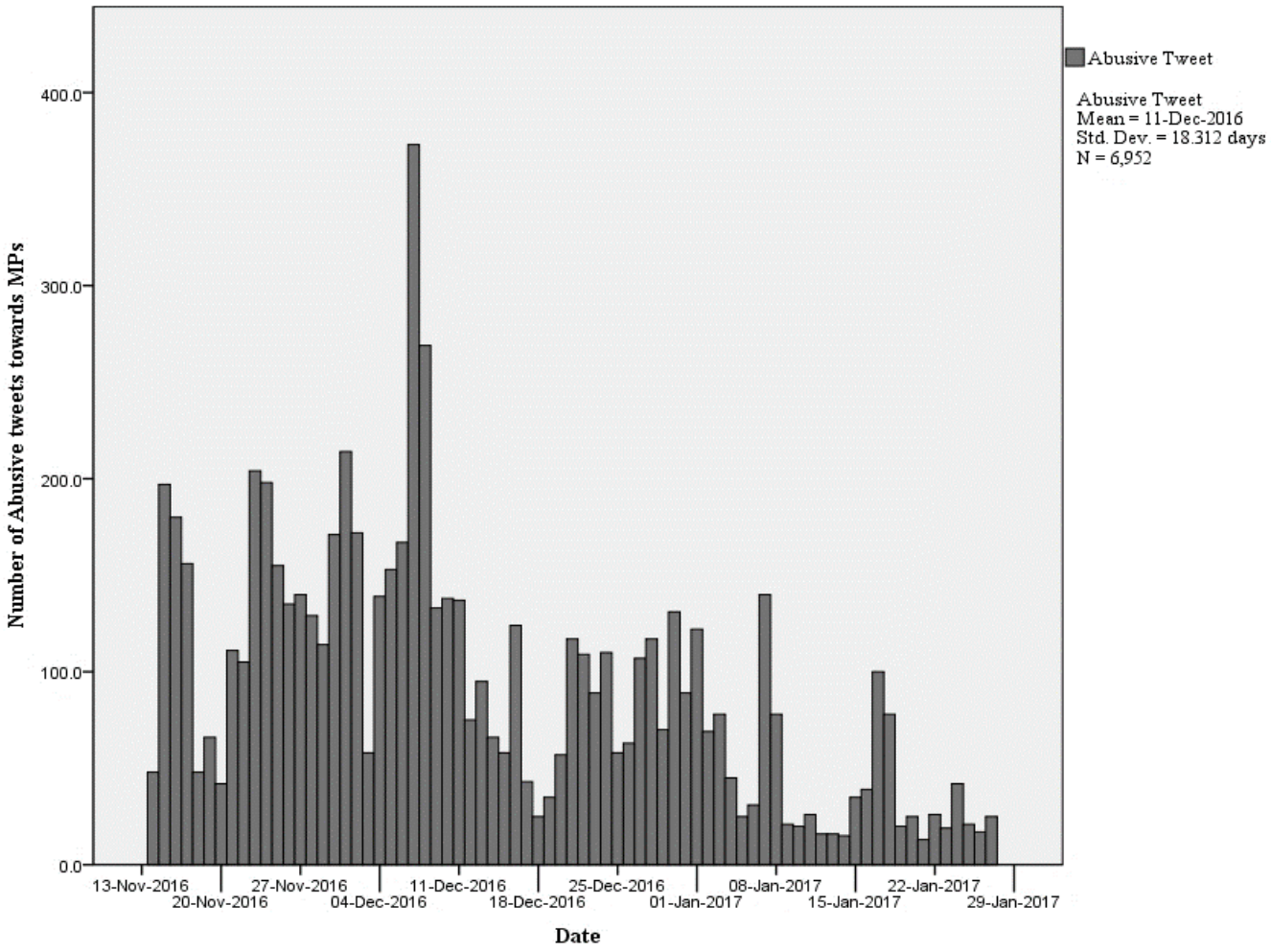
In total, 62% of the MPs had at least one abuse tweet sent to them during the two months (N=569; *n*=370). This is significantly more than the 10.1% of MPs who had reported receiving 'inappropriate social media contact' in James *et al*'s 2010 survey of UK MPs (2016:10). This is an indication of one of three options: levels of abuse on social media has increased; abuse has remained consistent for each MP but more MPs are now on the service to be abused; or simply that each MPs perception of abuse is different and was therefore under-reported in 2010. However, in reference to the James *et al* survey of MPs, the number of MPs abused on Twitter is more similar to the amount of abuse received by MPs by letters, fax, or emails with 63.9% of MPs reporting abuse/harassment through these mediums (2016:10). This could infer that while the scale of abuse may be different, the number of MPs being abused remains comparable to previous predominantly textual mediums.

Posts which included hate speech towards MPs occurred at a significantly reduced frequency to those categorised as abusive. Only 0.42% of tweets contained some type of hate speech (n=125) out of the total. Furthermore, the scope of MPs sent hate-speech tweets was also considerably reduced; with 6.6% of MPs receiving some form of hate speech (n=38). The only comparable research is from Ethiopia which found that 0.7% of all online political communication contained hate-speech (Gagliardone *et al,* 2016). In both this research, evidence from Ethiopia, anecdotal evidence and news presence suggested that political hate-speech is more widespread than what this data suggests.

The timeline of when the abusive tweets occurred displays that abuse was not consistent across the data collection period. Indeed, there is evidence of significant groupings of abuse on particular days. Most notably there was a significant spike of abuse on December 7th 2016. As an explanation it can be considered that abuse is not a day-to-day occurrence, but instead fuelled by outside factors, political events, or current affairs stories. For example, of the December 7th, this was also the day that the House of Commons voted on the Government's Brexit timetable. While common user traits of social media are visible, for example, less abuse on the weekends aligns with when people use Twitter. There are significant deviations which indicates that abuse is reactive to outside factors.

Overall, the evidence suggests that most MPs encounter abuse to some extent on Twitter with the level of abuse differing substantially from one MP to another. Therefore, we tested the levels of abuse across three characteristics: recognition (measured by total tweets received), gender, and political party. The characteristics were tested with the overall dataset, and then again with an interquartile dataset to test the 'average' experience of MPs by treating the experience of the top and bottom twenty-five per cent of MPs in terms of recognition as outliers. This provided a different perspective of abuse towards MPs.

GRAPH 1: Abusive Tweets by Date



Relationship between recognition against abuse (Overall)

Measuring the recognition and popularity of an MP in the public eye is fraught with difficulty due to the subjective nature of the terms. However, there are some statistics provided by this database which can be used to determine a value for an MPs individual name recognition and popularity. In research presented by Cha *et al* (2010) it was found that the number of followers a user has does not necessarily relate to the user's name recognition but does to their overall popularity. There was a relationship between mentions and name recognition (ibid, 2010). Retweets symbolised interest in the content of the tweet rather than the poster. Therefore, by using the measures of followers (for popularity of MP) and @mentions (for recognition) we could measure the impact of these two fields on abuse with linear regression.

The first finding was that increased name recognition has a positive relationship with levels of abuse (r=0.893 r2= 0.797 p=<0.01). This indicates that the more an MP is mentioned on Twitter, the more abuse they receive.

The second finding is that popularity has a positive relationship with abuse, however, this relationship is weaker than with name recognition (r=0774. r2= 0.599 p=<0.01). On the assumption that a more popular MP will be liked more, and therefore abused less, this finding should not be surprising (see Chen *et al*, 2010).

TABLE 2: MPS WITH MORE THAN FIFTY ABUSIVE TWEETS

MPs Handle	No. Abusive Tweets	Total No. Tweets	% of Tweets Abusive
Jeremycorbyn	677	16,944	4.00
Timfarron	438	5,133	8.53
BorisJohnson	409	9,452	4.33
michaelgove	329	4,234	7.77
theresa_may	273	11,779	2.32
Anna_Soubry	264	7,582	3.48
ChukaUmunna	243	7,355	3.30
nick_clegg	223	3,157	7.06
George_Osborne	198	1,505	13.16
Jeremy_Hunt	180	2,618	6.88
Ed_Miliband	161	2,376	6.78
PhilipDaviesMP	126	966	13.04
BenPBradshaw	120	1,500	8.00
AlexSalmond	114	2,521	4.52
HackneyAbbott	110	2,688	4.09
DavidTCDavies	109	1,561	6.98
OwenSmith_MP	104	2,005	5.19
DavidLammy	101	3,507	2.88
RhonddaBryant	78	3,521	2.22
DouglasCarswell	75	2,559	2.93
SarahChampionMP	72	2,388	3.02
jessphillips	72	4,164	1.73
RichardBurgon	71	3,816	1.86
andyburnhammp	57	2,933	1.94
labourlewis	55	3,589	1.53
PHammondMP	51	2,592	1.97
PeteWishart	51	4,353	1.17
Total	4,761	116,798	4.07

Impact of Gender on the levels of abuse received

Gender has some impact on the level of abuse received by MPs. However, despite media reporting of the targeting of female MPs on the service, the data shows that male MPs attract more abuse. Male MPs had 3 per cent of all tweets sent to them being deemed abusive. This is in comparison to the 1.7 per cent to female MPs. Alternatively, male MPs receive 30 abusive messages in every thousand, compared to the 17 in every thousand received by females. This result was tested with a Point Biserial correlation, which found a positive relationship between being a male MP and receiving abuse (r= 0.39 p= <.01). However, the R^2 (0.02) suggests this relationship is only accounts for two per cent of the variation of abusive tweets. Therefore, suggesting while there is a relationship between being a male MP and the levels of abuse on

Twitter, this relationship is not as significant as other factors.

While male MPs attracted more abuse, it seems opposite was true for hate Speech. Female MPs received significantly higher proportion of hate speech compared to male MPs. Out of all tweets that contained hate speech, 86% was directed at females. (N=125, *n*=108). This result is partly due to our classification where offensive gendered slurs counted as hate-speech, but also due to the increased levels of hate-speech attracted by women. This results does suggest that while women may not receive more unwanted contact than male MPs, the abuse women receive is gendered in its content.

Impact of political party of an MP on the levels of abuse received

The initial analysis of abuse towards MPs segmented by political party suggests that being a member of a national party with a large vote share in the 2015 general election attracted a larger amount of abuse as a percentage compared to that of parties based in devolved nations of the UK with smaller overall vote shares. This aligns with the findings of the popularity section – which suggests the most recognised MPs will attract more abuse and as these MPs will belong to the more recognised political parties. Table 2 displays this relationship. It is worth noting that the Liberal Democrats and UKIP are heavily skewed due to individual members. The Liberal Democrats abuse was mostly due to abuse directed to Leader Tim Farron MP whereas UKIP's abuse was due to Douglas Carswell, the party's only MP.

The interquartile dataset

During the analysis, it became clear that the overall dataset was affected by high profile MPs, or by MPs who had an extremely minimal use of the service. Therefore, to test the *average* MPs experience, we further filtered the data to the interquartile in terms of overall recognition. This filtered database contained 284 MPs.

The differences found between the overall and the interquartile show a difference in experience between the most popular and average MPs. Firstly, the interquartile MPs are abused significantly less. Only 0.95% of tweets to the interquartile contained any abuse, compared to 2.57% of the overall dataset. There were only two messages considered hate-speech in the interquartile dataset which did not allow for analysis in this area other than to suggest than to confirm that recognition is related to hate-speech (*n=2;* N=51,132).

The interquartile set showed no difference in terms of the gender relationship with abuse. As with the overall data, males continued to attract a significant amount of abuse, 1.04% of all tweets abusive for males MPs, compared to 0.7% of all tweets to female MPs.

VI. PERPETRATORS OF ABUSE

While there is much to be said about the abuse itself, an overview of the perpetrators of abuse is harder to quantify. While cultural stereotypes of online trolls or abusers depicts most online abusers as young white males who abuse to alleviate boredom (Hardaker, 2013; Fichman & Sanfilppo, 2016:143). However, the quantification and description of online trolls is masked on Twitter, in part due to the optional anonymous environment afforded by Twitter, but also because the service does not collect gender or age information from users directly. For these reasons, research into *who* trolls is problematic. Thus, while some user traits can be analysed from the data, some key identifiers of abusers could not be found from this research.

In addition to perceptions of online abusers being young white males, there is also the belief that perpetrators are often nocturnal. Speaking to the Home Select Committee Lindsay Hoyle MP called the abusers of MPs keyboard warriors who are active at the middle of the night (Home Affairs Select Committee, 2017). The cultural perceptions created an obvious test: to measure if the abuse of MPs is nocturnal in nature. Using our data, we tested if there were any patterns of abuse in terms of time of day. It was found that the difference between when abusers attacked MPs and overall tweets to MPs was negligible; with the mean abuse occurring at 2:59pm (Std. Dev. = 0.249 days) compared to the overall mean of 2:37pm. (Std. Dev. = 0.286 days). This goes some way to dispel the myth that all online abusers of MPs are surfing the web late at night, or that the time of day can be used to predict when abuse will happen.

One finding is that the perpetrators of abuse towards MPs may do so as a reaction to content by MPs within their social awareness streams. In our dataset, tweets were coded to determine if a tweet was made directly to the MP, or as a reply to tweet an MP had made. In a test, we found a positive relationship between abuse and the tweet being in the form of a reply ($r=0.22$ $p=<0.1$).

TABLE 3: ABUSE BY POLITICAL PARTY

Political Party	No. Abusive Tweets	Total No. Tweets	% of Tweets Abusive	No. of MPs in sample	Mean Abuse per Party MP
Lib Dems	678	10,470	6.48	9	75.4
UKIP	75	2,559	2.93	1	75
Conservative	2,695	99,087	2.72	276	9.8
Labour	3,020	128,915	2.34	207	14.6
Independent	36	1,710	2.11	3	12
DUP	36	1,757	2.05	7	5.1
SNP	379	22,650	1.67	53	7.1
Plaid Cymru	6	514	1.17	3	2
Green	25	2,370	1.05	1	25
SDLP	2	284	0.70	3	.67
Sinn Fein	0	342	N/A	4	.0
UUP	0	59	N/A	2	.0
Total	6,952	270,717	2.57	569	12.2

This goes against arguments that perpetrators purposefully go online to abuse, but instead suggests that the phenomenon of abuse is more often an immediate reaction triggered by what an abuser sees.

Following on from this, it was found from the list of users who had sent abusive tweets to MPs, overall, are not serial transgressors. The 6,952 abusive messages sent to MPs was made by 4,775 Twitter users, which gives a mean of 1.4 abusive tweets per user in this list. This is somewhat low when contrasted with the whole dataset, whereby 270,717 tweets were sent by 83,648 different Twitter users (3.23 messages per user). Many abusers sent only one abusive message (n=3,799), while only 28 accounts had sent over 10 abusive messages to MPs. This suggests that abuse is distributed across a vast number of different Twitter handles, and overall, is not perpetrated by a small group of highly abusive members. Accounts who had tweeted more than one abusive message tended to distribute their abusiveness across several MPs. Further inquiry is needed to understand this relationship.

TABLE 4:

Number of Abusive Messages sent	Frequency of Accounts
1 - 10	4,750
11- 20	17
21-30	10
101-110	1
Total	4,775

*No accounts sent abusive messages from a range between 29 to 107

VI. DISCUSSION AND CONCLUSIONS

The level of abuse on first sight appears relatively low. Indeed, given the apparent collapse in trust in MPs from the public and general levels of contempt for politicians amongst the electorate, we might have expected a greater volume of abuse. However, it is worth remembering that we took a fairly strict definition of abuse i.e. direct to MP on one platform. We were not measuring abuse in general or abuse directed at family members or staff for example. Moreover, we have not assessed the totality of abuse across different technological platforms. It is possible that some abusive tweets were blocked or removed before they could be collected into the dataset. Since there is little or no previous comparable data, it is difficult to tell how and whether abuse of MP's on social media is a growing problem. James et al study of self reported abuse from 2010 is the only reference point we could find. If we used this as a base-line, then abuse on social media has certainly expanded significantly with the majority of MPs facing abuse of some sort on Twitter and a smaller number receiving more regular abuse. Again, how far this represents an overall growth in abuse is difficult to interpret since we don't know whether there is a substitution effect here i.e. whether, for example, abusers have moved from email or letter to Twitter.

Overall levels of abuse are only one part of the picture. It is clear that certain individual politicians are targeted for abuse. This, perhaps not surprisingly, relates to their recognition more generally and how active they are on Twitter with gender and party seeming to have less of an impact. However, our results do suggest that female MPs attract more hate speech and threatening behaviour. In part, this represents definitional issues around abuse and hate-speech but also reflects the different tone of abuse directed to some woman MPs. A positive phenomenon following abusive tweets to MPs (for example, recent death threats made towards Anne Soubry MP) is an overwhelming counter-response by other Twitter users in support of the MP. Even if the level of abuse is relatively low, it is still also important to consider the perception and fear of abuse and whether this potentially restricts debate and the use of social media for connecting with the public. It is clear that some MPs have removed themselves from Twitter as a result of abuse and some self-censor before communicating on social media.

In terms of the pattern of abuse, our results confirm a more reactive model e.g. that Twitter abusers are often responding to something they have seen about the MPs online or to the discussion of emotive political issues in which MPs have taken part. Abuse clearly relates to the nature and profile of political discussion, for example, abuse peaked in our dataset with the parliamentary discussion of Brexit. Similarly, on a more regular basis we can also connect small spikes of abuse to the high profile political event of the week, PMQs. In general, there are some hints here of polarized nature of UK political debate on some issues, with several MPs being targeted especially for pro-remain views. Interestingly, there is also the appearance of terms of abuse (e.g. snowflake) spilling over from the recent highly polarized US presidential debate.

Although more work and research are required to identify abuser profiles and patterns of abuse, our initial findings do not fit the media stereotypes. The majority of abusers do not appear to be serial trolls tweeting late at night. Nor did we detect any networks of abusers although there was some minor evidence of bot activity (a potentially new development in online abuse). The character of most of the abuse was one-off comments in relation to political debates or events. Arguably, these are types of comments which would once have been privately shouted at a television screen – however, social media has given people more direct access to publicly shout such comments at public figures. This potentially means that such abuse has more power and impact than the past previously private communications, both specifically on the targets of abuse and on a more general perception of MPs. The apparently irregular nature of abusers also underlines a theoretical problem here. Whilst considerable attention has focused on trolling or cyber-bullying in both media discussion and social science literature, these are not necessarily helpful frames for understanding the majority of political abuse online. Most our abuse was not from repeat abusers nor was it necessarily designed to amuse other tweeters or provoke response from the target as suggested by most definitions of trolling.

In sum, this study has attempted to provide an empirical benchmark for assessing abuse of politicians and also a methodology for doing so. We would argue that analysis of actual abuse online provides a more accurate and broader understanding than traditional self-reporting studies and, indeed, media representations of the problem. Nevertheless, there is further work to be done in a number of areas: firstly, to compare different technological platforms to see whether the nature of platforms and the rules and norms around them can influence the civility of discussion. Secondly, we need larger, repeat studies to monitor whether social media abuse is growing and, if so, why. Thirdly, UK politics is often regarded as one of the more adversarial political systems that potentially encourages more aggressive forms of contact between politicians and citizens. It would therefore be useful to have comparative studies to help understand the influence of systemic political factors (e.g. the nature of representative system, the party system, the linkage to constituency) and to see whether patterns of abuse are replicated cross-nationally.

REFERENCES

Adams, S.J., Hazelwood, T.E., Pitre, N.L., and Bedard, T.E. (2009). Harassment of Members of Parliament and the Legislative Assemblies in Canada by Individuals Believed to be Mentally Disordered. *Journal of Forensic Psychiatry and Psychology,* 20: 801-814.

Albrecht, S. (2006). Whose voice is heard in online deliberation? A study of participation and representation in political debates on the Internet. *Information, Communication & Society*, 9, 62–82.

Avery, J. (2010). 'Gender Bender Brand Hijacks and Consumer Revolt: The Porsche Cayenne Story.' In J. Avery, S. Beatty, M.B Hollbrook, R.V. Kozinets, & B. Mittal (ed.), *Consumer Behavior: Human Pursuit of Happyness in the World of Goods.* (pp. 645-649). Cincinnati, OH: Open Mentis.

Barber, B. R. (1998). "Three Scenarios for the Future of Technology and Strong Democracy". *Political Science Quarterly*, 113: 573–589

Barbera, P. (2015). "How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the U.S." Paper prepared for APSA Conference.

Bartlett, J., Reffin, J., Rumball, N., and Williamson, S. (2014). *Anti-Social Media.* DEMOS: London. Available at: https://www.demos.co.uk/files/DEMOS_Anti-social_Media.pdf

BBC News Online, (2017) 'Mistreatment of Women MPs Revealed', 25 January. http://www.bbc.co.uk/news/uk-politics-38736729. (Last accessed 4 April 2017)

Bloom, D. (2015, 17 September). Labour MP John Woodcock Quits Twitter over 'abuse' from Jeremy Corbyn's Supporters. *The Mirror.* Available at: http://www.mirror.co.uk/news/uk-news/labour-mp-john-woodcock-quits-6459827

Braun, V., and Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in Psychology,* 3(2): 77-101. DOI:10.1191/1478088706qp063oa

Buckels, E.E., Trapnell, P.D., and Paulhus, D. (2014). Trolls just want to have fun. *Personality and Individual Differences,* 67: 97-102. DOI: 10.1016/j.paid.2014.01.016

Buhrmester, M., Kwang, T., and Gosling, S.D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science,* 6(1):3-5.

Burnap, P. and Williams, M.L. (2013). Cyber Hate Speech on Twitter: An Application of Machine Classification and Statistical Modelling for Policy and Decision Making. *Policy & Internet,* 7(2):223-242. DOI: 10.1002/poi3.85

BuzzFeed News (2017) 'The President Of Italy's Chamber Of Deputies Is Launching A Campaign Against Fake News', 8 February. https://www.buzzfeed.com/albertonardelli/the-president-of-italys-lower-house-says-fake-news-is-fuelli?utm_term=.qt3n6genl#.jue51oq5K. (Last accessed 3 April 2017).

Carter, A., and Sneesby, J. (2017, 25th January). Mistreatment of Women MPs revealed. *BBC News.* Available at: http://www.bbc.co.uk/news/uk-politics-38736729

CBC News, (2016) 'Sandra Jansen latest in a long string of female politicians to face abuse'. 23 November.http://www.cbc.ca/news/canada/calgary/sandra-jansen-alberta-mla-misogyny-1.3865047. (Last accessed 4 April 2017).

CBC News, (2017) 'Premier Kathleen Wynne bombarded with homophobic, sexist abuse'. 25 January. http://www.cbc.ca/news/canada/toronto/kathleen-wynne-twitter-abuse-1.3949657. (Last accessed 4 April 2017).

Cha, M., Haddadi, H., Benevenuto, F., and Gummadi, K. (2010) Measuring User Influence in Twitter: The Million Follower Fallacy. *Proceedings of ICWSM'10.* Available at: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1538/1826

Chen, W., Yuan, Y., and Zhang, L. (2010). Scalable influence maximization in Social Media Networks under the linear threshold model. Paper presented to *The 2010 IEEE International Conference on Data Mining:* December 13-17. University of Technology Sydney, Australia. (pp. 88-97). Available at: http://ieeexplore.ieee.org/document/5693962/

Chen, Y., Zhou, Y., Zhu, S., and Xu, H. (2012). Detecting offensive language in social media to protect adolescent online safety. Paper Presented to *2012 International Conference on Privacy, Security , Risk and Trust and International Conference on Social Computing (socialcom):* September 3-5. (pp.71-80). DOI: 10.1109/SocialCom-PASSAT.2012.55

Chikashi, N., Tetreault, J., Thomas, A., Mehdad, Y., and Chang, Y. (2016). Abusive Language Detection in Online Content. Paper presented to *WWW 2016:* April 11-15. Montreal, Canada. DOI: 10.1145/2872427.2883062

Coleman, S. (1999). "Can the New Media Invigorate Democracy?" *The Political Quarterly*, 70: 16–22.

Coleman, S. (2005). "New mediation and direct representation: reconceptualizing representation in the digital age", *New Media & Society*. **7**.2: 177-198.

Colleoni, E., Rozza, A and Arvidsson, A. (2014). "Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data." *Journal of Communication,* 64 (2): 317–332.

Dahlgren, P. (2009). *Media and Political Engagement: Citizens, Communication and Democracy*. Cambridge: CUP.

Daily Telegraph, (2016) 'Labour MP Jess Phillips receives more than 600 rape threats in one night'. 31 May. http://www.telegraph.co.uk/news/2016/05/31/labour-mp-receives-more-than-600-rape-threats-in-one-night/. (Last accessed 4 April 2017).

Demos, (2016) *The Use of Mysogynistic terms on Twitter.* https://www.demos.co.uk/wp-content/uploads/2016/05/Misogyny-online.pdf. (Last accessed 5 April 2017).

Dietz P, Martell DA (1989) Mentally Disordered Offenders in Pursuit of Celebrities and Politicians. Washington, DC: National Institute of Justice

Dinakar, J., Jones, B., Havasi, C., Leiberman, H. and Pickard, R. (2012). Common Sense Reasoning for Detection, Prevention and Mitigation of Cyberbullying. *ACM Transactions on Interactive Intelligent Systems,* 2(3). DOI: 10.1145/2362394.2362400

Every-Palmer, S., Barry-Walsh, J., & Pathé, M. (2015). Harassment, stalking, threats and attacks targeting New Zealand politicians: A mental health issue. *Australian and New Zealand Journal of Psychiatry*, *49*, 634–641. DOI: 10.1177/0004867415583700.

Farris, R., Ashar, A., Gasser, U., and Joo, D. (2016). *Understanding Harmful Speech Online: Networked Policy Series.* Cambridge, MA: Berkman Klein. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2882824

Francoli, M. and Ward, S., (2008). '21st century soapboxes? MPs and their blogs'. *Information Polity*, 13(1-2): 21-39.

Gagliardone, I., Pohjonen, M., Beyene, Z., Zerai, A., Aynekulu, G., Bekalu, M., Bright, J., Moges, M.A., Seifu, M., Stremlau, N., Taflan, P., Gebrewolde, T.M., and Teferra, Z. (2016). Mechachal: Online Debates and Elections in Ethiopia – From Hate Speech to Engagement in Social Media. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2831369

Gazette, (2016) 'Corbyn row Twitter troll who threatened Labour MP is exposed - as a party member'. 1 July. http://www.gazettelive.co.uk/news/teesside-news/corbyn-row-twitter-troll-who-11550844. (Last accessed 2 April 2017).

Guardian, (2016) 'David Lammy receives death threats after EU referendum result'. 4 July. https://www.theguardian.com/politics/2016/jul/04/david-lammy-receives-death-threat-after-eu-referendum-result. (Last accessed 3 April 2017).

Guardian, (2017).'Police are only in the foothills of tackling online abuse, MPs are told'. 21 February. https://www.theguardian.com/politics/2017/feb/21/police-are-only-in-the-foothills-of-tackling-online-abuse-mps-told. (Last accessed 3 April 2017).

Hagemann, C. (2002). Participation in and contents of two Dutch political party discussion lists on the internet. *Javnost/The Public*, 9, 61–76.

Hansard (2016). H.C. Vol 612, 7 July.

Harp, D., & Tremayne, M. (2006). The gendered blogosphere: Examining inequality using network and feminist theory. *Journalism & Mass Communication Quarterly*, 83, 247–64.

Hermans, L. and Vergeer, M., (2013). 'Personalization in e-campaigning: A cross-national comparison of personalization strategies used on candidate websites of 17 countries in EP elections 2009'. *New media & society*, 15 (1): 72-92.

Ho, E. (2017). An Update on Safety. Retrieved 15 March 2017 from: https://blog.twitter.com/2017/an-update-on-safety

Home Affairs Select Committee, (2017) *Hate Crime and its violent consequences* (HC609) http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/home-affairs-committee/hate-crime-and-its-violent-consequences/oral/49182.html. (Last accessed 5 April 2017).

Hsieh, H., and Shannon, S.E. (2015). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research,* 15(9). DOI: 10.1177/1049732305276687

Hurst, M. (2016, 3 May). MP Quits Twitter Over Troll Abuse. *Grimsby Telegraph.* Available at: http://www.grimsbytelegraph.co.uk/8203-mp-quits-twitter-troll-abuse/story-29217838-detail/story.html

Independent, (2016) 'Police investigation death threats targeting Remain-supporting Conservative MP Anna Soubry', 2 December. http://www.independent.co.uk/news/uk/politics/anna-soubry-jo-cox-murder-threat-brexit-eu-referendum-a7452546.html. (Last accessed 3 April 2017).

Inter-Parliamentary Union (2016) *Sexism, Harassment and Violence against Women Parliamentarians*, Geneva: IPU

Jackson N. (2008) 'Representation in the Blogosphere: MPs and Their New Constituents'. *Parliamentary Affairs,* 61 (4): 642-660

Jackson, N. (2003). MPs and web technologies: an untapped opportunity?. *Journal of Public Affairs*, 3 (2): 124-137.

Jackson, N. and Lilleker, D. (2011). Microblogging, constituency service and impression management: UK MPs and the use of Twitter. *The Journal of Legislative Studies*, 17 (1): 86-105.

Jackson, N.A. and Lilleker, D.G. (2009). MPs and E-representation: Me, MySpace and I. *British Politics*, 4(2): .236-264.

Jackson, N.A., and D.G. Lilleker (2007). 'Seeking unmediated political information in a mediated environment: The uses and gratifications of political parties' e-newsletters.' *Information, Community and Society* 10 (2): 242-264.

James, D.V., Farnham, F.R., Sukhwal, S., Jones, K., Carlise, J., and Henley, S. (2016). Aggressive/Intrusive Behaviours, Harassment and Stalking of Members of the United Kingdom Parliament: A Prevalence Study and Cross-National Comparison. *The Journal of Forensic Psychiatry & Psychology,* 27(2): 117-197. DOI: 10.1080/14789949.2015.1124908

James, D.V., Mullen, P., Meloy, J.R., Pathé, M. Farnham, F., Preston, I., and Darnley, B. (2007). The Role of Mental Disorder in Attacks on European Politicians, 1990-2004. *Acta Psychiatrica Scandinavica,* 116(5): 334-344. DOI: 10.1111/j.1600-0447.2007.01077.x

Jankowski, N., & van Selm, M. (2000). The promise and practice of public debate in cyberspace. In K. T. Hacker & J. van Dijk (ed.), *Digital Democracy: Issues of Theory and Practice* (pp. 149–65). Thousand Oaks, CA: Sage.

Jensen, J. L. (2003). Virtual democratic dialogue? Bringing together citizens and politicians. *Information Polity*, 8, 29–47.

Joinson, A. (ed) (2007). *Oxford Handbook of Internet Psychology*. Oxford: OUP.

Journal.ie, (2016) We will kill you": Irish politicians speak out against online abuse. 19 June. Available at http://www.thejournal.ie/irish-politicians-abuse-online-2833042-Jun2016/ (last accessed 3 April 2017)

Karvonen, L., (2010). *The personalisation of politics: a study of parliamentary democracies*. ECPR Press.

Laird, S. (2012, 29 October). Instagram Users Share 10 Hurricane Sandy Photos Per Second. *Mashable*. Available at: http://mashable.com/2012/10/29/instagram-hurricane-sandy/

Lelkes, Y., Sood, G. and Iyengar, S. (2015), The Hostile Audience: The Effect of Access to Broadband Internet on Partisan Affect. *American Journal of Political Science*.

Lenhart, A., Ybarra, M., Zickuhr, K., and Prince-Freeney, M. (2016). *Online Harassment, Digital Abuse, and Cyberstalking in America*. New York, USA: Data & Society. Available at: https://www.datasociety.net/pubs/oh/Online_Harassment_2016.pdf

Lilleker, D. and Jackson, N. (eds.) (2011). *Political Campaigning, Elections and the Internet: Comparing the US, UK, France and Germany*. London: Routledge.

Lilleker, D. and Jackson, N., (2014). 'Interacting and representing: can Web 2.0 enhance the roles of an MP?' *In: ECPR Joint Sessions of Workshops, 14--19 March 2009, Lisbon*. (Unpublished)

Lilleker, D. G., Koc-Michalska, K., Schweitzer, E. J., Jacunski, M., Jackson, N. and Vedel, T. (2011). 'Informing, Engaging, Mobilizing or Interacting: Searching for a European Model of Web Campaigning'. *European Journal of Communication* 26 (3), 195-213

Lumsden, K., and Morgan, H.M. (2017). Media Framing of Trolling and Online Abuse: Silencing Strategies, Symbolic Violence and Victim Blaming. *Feminist Media Studies,* 17(6), ISSN:1471-5902.

McPherson, M., Smith-Lovin, L., and Cook, J.M. (2001). "Birds of a feather: Homophily in social networks". *Annual Review of Sociology*, 28: 415–444.

McVeigh, T. (2015, 18 April). Jack Monroe Quits Twitter Over Homophobic Abuse. *The Guardian*. Available at: https://www.theguardian.com/technology/2015/apr/18/jack-monroe-quits-twitter-over-homophobic-abuse

Murthy, D. (2017). The Ontology of Tweets: Mixed-Method Approaches to the Study of Twitter. In L. Sloan, and A. Quan-Haase. (eds). *The Sage Handbook of Social Media Research Methods*. London: Sage.

Mutz, D.C. and Martin, P.S. (2001). "Facilitating communication across lines of political difference: The role of mass media". *American Political Science Review*, 95 (1): 97–114.

Nobata, C., Tetreault, J., Thomas, A., Mehdad, Y., and Chang, Y. (2016). Abuse Language Detection in Online User Content. Proceedings of *WWW'16 25th International Conference on World Wide Web*. (pp.145-153). DOI: 10.1145/2872427.2883062

Norris, P. ed. (1999). *Critical citizens: Global support for democratic government*. OUP Oxford.

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

Pathé, M., Phillips, J., Perdacher, E., & Heffernan, E. (2014). The harassment of Queensland members of parliament: A mental health concern. *Psychiatry, Psychology and Law, 21*, 577–584.

Prior, M. (2007). *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge: Cambridge University Press

Prior, M. (2013). "Media and Political Polarisation". *Annual Review of Political Science*, 16: 101–27.

Ratkiewicz, J., Conover, M.D., Meiss, M., Goncalves, B., Flammini, A., and Menczer, F. (2011). Detecting and Tracking Political Abuse in Social Media. Paper Presented at *The fifth AAAI Conference on Weblogs and Social Media.* Available at: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2850

Rowe, I., (2015). 'Civility 2.0: a comparative analysis of incivility in online political discussion'. *Information, Communication & Society*, 18(2): 121-138.

Schoeman-Morris, K.A., Scalora, M.J., Chang, G.H., Zimmerman, W.J., and Garner, Y. (2007). A Comparison of Email Versus Letter Threat Contacts towards Members of the United States Congress. Journal of Forensic Sciences, 52(5):1142-1147. DOI: 10.1111/j.1556-4029.2007.00538.x

Sellers, A.F. (2016). Defining Hate Speech. Cambridge, MA: Berkman Klein. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2882244

Serfaty, V., (2010). 'Web campaigns: Popular culture and politics in the US and French presidential elections'. *Culture, Language, Representations,* 8*:* 115-129.

Shachaf, P., & Hara, N. (2010). Beyond Vandalism: Wikipedia Trolls. Journal of Information Science, 36(3):357-370.

Shane, P.M. (2004). *Democracy Online: The Prospects For Political Renewal Through The Internet*. Routledge: London.

Sloan , L. , Morgan , J. , Housley , W. , Williams , M. , Edwards , A. , Burnap , P. and Rana , O. (2013). Knowing the Tweeters: Deriving sociologically relevant demographics from Twitter. *Sociological Research Online*, 18(3). DOI:10.5153/sro.3001

Sloan, L. (2017). Social Science 'Lite'? Deriving Demographic Proxies from Twitter. In L. Sloan, and A. Quan-Haase (eds). *The Sage Handbook of Social Media Research Methods*. London: Sage.

Streck, J.(1998) 'Pulling the Plug on Electronic Town Meeting: Participatory Democracy and the Reality of the Usenet'. In Toulouse, C and T. Luke (eds). *The Politics of Cyberspace*, Routledge: New York, 18-47.

Stromer-Galley, J. (2000). 'Online interaction and why candidates avoid it', *Journal of Communication*, 50 (4): 111–132.

Stromer-Galley, J. (2002). Diversity and political conversations on the Internet: Users' perspectives. *Journal of Computer-Mediated Communication*, 8. http:/jcmc.indiana.edu/ vol8/issue3/stromergalley.html.

Stromer-Galley, J. and Wichowski, A., (2011). 'Political Discussion Online'. In Ess, C and M. Consalvo (eds.), *The Handbook of Internet Studies*, Wiley: Blackwell: Oxford, 168-187.

Suler, J. (2004). 'The online disinhibition effect.' *Cyberpsychology & behavior*, 7 (3): 321-326.

Sunday Herald, (2017) 'The truth about Scotland and online abuse: cybernats and cyberbrits are just as bad as one another', 18 February. http://www.heraldscotland.com/news/15102321.The_truth_about_Scotland_and_online_abuse____39_cybernats__39__and___39_cyberbrits__39__are_just_as_bad_as_each_other/. (Last accessed 4 April 2017).

Thelwall, M., Buckley, K., and Paltoglou, G. (2012). Sentiment Strength detection for the social web. *Journal of the American Society for Information Science and Technology,* 63(1):163-173.

Theocharis, Y., Barberá, P., Fazekas, Z., Popa, S.A. and Parnet, O., (2016). 'A Bad Workman Blames His Tweets: The Consequences of Citizens' Uncivil Twitter Use When Interacting With Party Candidates'. *Journal of Communication*, 66 (6): 1007-1031.

Trammell, K. D., & Keshelashvili, A. (2005). Examining the new influencers: A self- presentation study of A-list blogs. *Journalism & Mass Communication Quarterly*, 82, 968 – 82.

Voss, A., Lvov, I., and Thompson, S.D. (2017). Data Storage, Curation and Preservation. In L. Sloan, and A. Quan-Haase (eds). *The Sage Handbook of Social Media Research Methods.* London: Sage.
Waldron, J. (2014). *The Harm in Hate Speech.* Cambridge, MA: Harvard University Press.
Ward, S. and Lusoli, W. (2005). 'From weird to wired': MPs, the Internet and representative politics in the UK'. *Journal of Legislative Studies*, 11(1): 57-81.
Whitehead, H. (2016, 3 May). Brigg and Goole MP Andrew Percy Quits Twitter Over 'Increasing Levels of Personal Abuse'. *Scunthorpe Telegraph.* Available at: http://www.scunthorpetelegraph.co.uk/brigg-goole-mp-andrew-percy-quits-twitter/story-29217306-detail/story.html
Whittaker, E., and Kowalski, R.M. (2014). Cyberbullying Via Social Media. *Journal of School Violence,* 14(1). DOI: 10.1080/15388220.2014.949377
Williamson, A. (2009). 'The effect of digital media on MPs' communication with constituents'. *Parliamentary Affairs*, 62 (3): 514-527.
Williamson, A. (2010). *Digital citizens and democratic participation*. London: Hansard Society
Yardi, S., Romero, D., Schoenbeck, G., and Boyd, D. (2010). Detecting Spam in a Twitter Network. *First Monday,* 15(1-4). Available at: http://firstmonday.org/ojs/index.php/fm/article/view/2793/2431?utm_source=twitterfeed&utm_medium=twitter

APPENDIX 1: IDENTIFIERS OF ABUSIVE/HATE-SPEECH TWEETS TO MPS

These are the keywords used as described within the methodology to identify abusive or hate filled messages towards MPs. Deviations of these keywords were also used. For Example, "fuck off" would also identify tweets which included "fuckoff" and "fuck-off". Plurals of words were also collected.

'and die'	Faggot	Muzzies	Snowflake
Arsehole	Fascist	Nazi	Swamp
Bastard	Fool	Nonce	Thug
Bigot	Fuck off	Nutjob	Tosser
Bitch	Fuck You	Pedo/paedophile	Traitor
Buffoon	Fuckin/fucking	Pervert	Troll
Bullshit	Golliwog	Pillock	Turd
Cock-nosed	Hypocritical	Piss off	Twat
Cow	Ideologue	Ponce	Undemocratic
Crony	Idiot	Prick	Vile
Cunt	Imbecile	Radical	Wanker
Dick head	Kike	Rag Head	Wankstain
"Die in"	Kuffar	Rape	Weasle
Dirty	Libtard	Redtory	WhiteGenocide
Disgrace	Londonistan	Retard	Whore
Dishonourable	Loon	Scum	Wretch
Dumb	Loser	Shill	"You Racist"
Dyke	Lunatic	Shit	Zealot
Elitist	Moron	Slag	
Extremist	Murat/Muzrat	Sleaze	