Politics & Gender (2025), 1–28
doi:10.1017/S1743923X25100111

RESEARCH ARTICLE

When Does Fame Not Matter? Examining Gender Differences in Politicians' Social Media Experiences

Maarja Lühiste[1], Stiene Praet[2], Sebastian Adrian Popa[1], Yannis Theocharis[3], Pablo Barberá[4], Zoltán Fazekas[5] and Joshua A. Tucker[6]

[1]School of Geography, Politics and Sociology, Newcastle University, Newcastle-upon-Tyne, UK; [2]Department of Engineering Management, University of Antwerp, Antwerp, Belgium; [3]School of Social Sciences and Technology, Technical University of Munich, Munich, Germany; [4]Department of Political Science and International Relations, University of Southern California, Los Angeles, CA, USA; [5]Department of International Economics, Government and Business, Copenhagen Business School, Copenhagen, Denmark and [6]Department of Politics, New York University, New York, NY, USA
Corresponding author: Maarja Lühiste; Email: maarja.luhiste@ncl.ac.uk

(Received 22 March 2024; revised 22 May 2025; accepted 02 June 2025)

Abstract

Past research alerts to the increasingly unpleasant climate surrounding public debate on social media. Female politicians, in particular, are reporting serious attacks targeted at them. Yet, research offers inconclusive insights regarding the gender gap in online incivility. This paper aims to address this gap by comparing politicians with varying levels of prominence and public status in different institutional contexts. Using a machine learning approach for analyzing over 23 million tweets addressed to politicians in Germany, Spain, the United Kingdom, and the United States, we find little consistent evidence of a gender gap in the proportion of incivility. However, more prominent politicians are considerably and consistently more likely than others to receive uncivil attacks. While prominence influences US male and female politicians' probability to receive uncivil tweets the same way, women in our European sample receive incivility regardless of their status. Most importantly, the incivility varies in quality and across contexts, with women, especially in more plurality contexts, receiving more identity-based attacks than other politicians.

Keywords: Gender and political representation; campaigns; social media; incivility; machine learning; elections; elite-voter interaction

Introduction

The 2019 British general election witnessed 50 members of the House of Commons not running for re-election. While it is not unusual for some MPs to stand down, the reasons stated for doing so were novel. Abuse, threats and the normalization of a culture of intimidation, especially on social media, were explicitly mentioned by mostly female MPs (Gorrell et al., 2020). This highlights yet another barrier aspirants seeking public office face, and that the height of this barrier may depend on candidate characteristics.

Existing studies offer inconclusive insights to the gender gap in online incivility, with some suggesting female politicians to receive more uncivil content (Collignon and Rüdig 2020; Gorrell et al. 2020), while others reporting the opposite (Erikson, Håkansson, and Josefsson 2021; Greenwood et al. 2019; Southern and Harmer 2021; Ward and McLoughlin 2020). This may partly be due to sample selection effects, the difference between using elite surveys versus reliance on observed incivility on social media, and many studies selecting on the dependent variable (i.e., uncivil comment). The lack of cross-national data and the omitting of several individual level variables, such as politicians' online and offline prominence, may further contribute to the mixed record of evidence. This paper aims to address these gaps in the literature by examining the dynamics of gender-based online incivility in a cross-national setting and by comparing male and female politicians with varying levels of prominence. Most importantly, we take one of the first steps in exploring the gender differences in the *type* of uncivil language in a cross-national setting.

We theorize that the underlying cause of the gender gap in the proportion and type of incivility can be due to bias derived from one's gender (gender discrimination logic) as well as due to their status and recognizability (prominence logic), with prominence potentially affecting men's and women's experiences to a different extent. We use machine learning models to analyze more than 23 million tweets directed at electoral candidates and members of national parliaments from Germany, Spain, the United Kingdom, and the United States at two separate time points: 2014 and 2016–17. Although Twitter was an important campaign tool during these time periods, there were also significant variations in individual politicians' Twitter use. Our approach thus presents a unique opportunity to investigate what affects the proportion and type of incivility and how the effect of traditional predictors of incivility, such as prominence and candidate viability, may vary depending on politician's gender and broader context at a time when Twitter was a widely used social media campaigning tool.

While our data offers little consistent evidence of a direct gender effect, women and men at different career stages and at varying levels of Twitter-prominence receive different levels of incivility. By juxtaposing the Twitter experiences of highly prominent members of the legislature and politicians enjoying Twitter-celebrity status to their lesser-known colleagues in four advanced democracies, this study moves beyond existing literature, which largely focuses on highly established politicians in a single-country case study.

Understanding the extent to which widely-used social media platforms, such as Twitter at the time, democratize public discourse by reducing existing inequalities between politicians — or, on the contrary, demobilize certain social groups — has important democratic consequences. Uncivil behavior aimed at female candidates, regardless of their prominence, likely limits women's choice of possible communication more than men's. More dramatically, extensive and severe attacks that female politicians experience from the moment they enter politics and social media may not only result in a more cautious online-persona, but also damage women's willingness to participate in politics (Krook 2017). The implications of differential experiences on social media thus go beyond currently active politicians. Increasingly unpleasant climate on social media could affect differently young women's and men's eagerness to participate in politics. As such, our findings point to potential policy implications, where a further need to regulate social media content, more similarly to traditional media, may be necessary to ensure the equality of political representation.

Incivility on Social Media

Social media are widely used for information consumption and political communication. Political elites, among others, have widely adopted Twitter as a tool for personalization, mobilization and promotion (Barberá and Zeitzoff 2017; Enli and Skogerbø 2013), though strong cross-party and cross-national differences exist (van Vliet, Törnberg, Uitermark and 2020). By allowing for (semi)public profiles and asymmetric following of others, Twitter has provided an opportunity to directly communicate with "powerful" individuals who were previously beyond most users' reach. Furthermore, Twitter was one of the first platforms to offer an opportunity for politicians who traditionally struggled for equal and equitable coverage in the traditional media (Aaldering and Van Der Pas 2020; Greene and Lühiste 2018; Hayes and Lawless 2016; Lühiste and Banducci 2016) to bypass the gatekeepers and control their own message. But while Twitter enabled new ways to reach the public and the elites, there have also been concerns about the platform's misaffordances likely inhibiting its interactive and democratic potential. Indeed, abundance of anecdotal evidence suggest that Twitter has over time become an environment that is rife with incivility and where prominent users are often viciously attacked (Friedersdorf, 2015).

Despite the agreement of the existence of incivility in the online sphere, past research debates the very concept of incivility. Some see it as violation of norms of politeness (Mutz 2015); others see it as violation of democratic norms (Papacharissi 2004), with recent conceptual refinements noting that *intolerance* — rather than incivility — better describes the second perspective making it, at the same time, also more easily distinguishable due to the democratically poisonous effect of behaviors that violate political equality (Rossini 2022). These two approaches to incivility are rarely examined in the same studies, leading to some conceptual and empirical muddling few have tried to disentangle (for exceptions see Muddiman 2017).

The conceptual debate notwithstanding, researchers document an increase in online incivility and intolerance (Rossini 2022; Theocharis et al. 2020). Twitter (now X) in particular, which has undergone major structural changes and even saw the mass firing of its content moderation team, allows one not only to remain largely anonymous, but restricts interaction toward short and direct messages, facilitating blunter communication (Sydnor 2018, 6). Furthermore, while the communication can be easy and rapid, attacks and abuse are largely unpunished due to Twitter's lack of capacity (and, more recently, willingness) to deal with it in a timely manner. Hence, the causes of online incivility, independent of the gender being targeted, lie at least partially on anonymity and the online disinhibition effect (Suler 2004).

Gender Discrimination Logic and Prominence Logic

Not everyone suffers from online incivility to the same extent and in the same way. Women parliamentarians are reportedly the number one target of online psychological violence, including "sexist and misogynistic remarks, humiliating images, mobbing, intimidation and threats" (Inter-Parliamentary Union 2016, 6)— language that falls within various definitions of incivility and intolerance in the literature. Moreover, attacks on female representatives tend to be defined by the age, ethnicity, and length of service, with young and minority ethnic representatives suffering more incivility, especially when first elected (Ward and McLoughlin, 2020). Yet, few studies have empirically and systematically mapped gender differences in online incivility on a large cross-national scale.

As a result, current record of evidence is mixed, with reports focusing on female parliamentarians only emphasizing the high prevalence of attacks toward them. For example, the overwhelming majority of the UK women MPs have experienced aggressive behavior and received physical threats (Collignon and Rüdig 2021; Krook 2017), in particular those belonging to ethnic minorities (Demos 2016). Moreover, when asked about online incivility in candidate surveys, female respondents not only report higher levels of attacks, but are also more likely than men to express fear (Collignon and Rüdig 2020, 2021) or perceive limits to their room for maneuver (Erikson, Håkansson, and Josefsson 2021; Tenove et al. 2023) as a reaction to their experiences. Unsurprisingly, a number of female representatives have been forced to quit social media after harassment campaigns (Inter-Parliamentary Union 2016) or decided to leave parliamentary politics altogether (Gorrell et al. 2020).

Yet, most studies using elite surveys with varying response rates, where empirical evidence relies on self-reported instances of abuse (see for example Collignon and Rüdig 2020, 2021; Krook 2017), differ in their conclusions from recent research documenting online incivility by either hand-coding or classifying the tweets targeted at the politicians (Greenwood et al. 2019; Southern and Harmer 2021; Ward and McLoughlin 2020). Studies looking at a wider range of different types of politicians show that male politicians, on average, receive more uncivil tweets than their female contenders (Greenwood et al. 2019; Southern and Harmer 2021; Ward and McLoughlin 2020). Yet, none of these findings are uniformly one-directional, with women being more likely stereotyped by identity (men by party) and to be questioned in their position as an MP (Southern and Harmer 2021). Moreover, Ward and McLoughlin (2020) point to differences in the intensity of incivility: while men, overall, receive a larger proportion of uncivil tweets, women and minority politicians are more likely to experience hate speech.

To complicate matters further, Rheault, Rayment, and Musulan (2019), using machine learning models to predict incivility in about 2.2 million messages addressed to Canadian and US politicians on Twitter, find evidence that women

https://doi.org/10.1017/S1743923X25100111 Published online by Cambridge University Press

— not men —are more heavily targeted by uncivil messages. But these findings are conditional too, with gender gap in receiving online incivility being moderated by candidate visibility. Their study, thus, provides a first indication that prominence might have an important role to play in who gets attacked online — a key consideration we discuss below.

While past empirical evidence offers mixed results, theories of gender role expectations and gender-based discrimination (Eagly and Karau 2002) suggest that female politicians may be susceptible to more online incivility than their male counterparts. While participation of women in politics is more common than ever before, stereotypes about traditional gender roles persist. As such, female candidates can still be seen as overstepping the private sphere when getting involved in politics, and thus subjected to more questioning of their presence in (online) public debates, which often comes with abuse and harassment (Eagly and Karau 2002; Krook and Restrepo Sanín 2016; Rheault, Rayment, and Musulan 2019). This may particularly be the case in the context where traditional gender roles are more prevalent.

While past research offers a mixed record of evidence, our theorizing building on gender role expectations and gender stereotyping lead us to hypothesize:

H1 Female politicians receive a higher proportion of uncivil tweets in comparison to their male contenders, all else being equal.

Yet not all female and male politicians have the same likelihood of experiencing incivility on social media (Rheault, Rayment, and Musulan 2019), with *prominence* being described as the primary driver of online attacks (Theocharis et al. 2016). For example, most party leaders and cabinet ministers enjoy greater name recognition than the average candidate running for a first — not to mention, second-order election — which likely translates to higher Twitter visibility and, potentially, to more attacks. But this type of status-led prominence is not the only driver of visibility. Certain politicians' capacity to strategically leverage Twitters' affordances and reach "celebrity politician" status therefore underlines a recent phenomenon that can neither be ignored,[1] nor conflated with traditional notions of prominence. We therefore propose that both status-led prominence and Twitter-celebrity status (measured crudely by number of followers) likely affects politicians' Twitter experiences.

H2a More prominent politicians receive a higher proportion of uncivil tweets in comparison to less prominent and less known politicians.[2]

However, does prominence affect male and female politicians the same way or does an implicit gender bias manifest through prominence affecting men's and women's Twitter experiences to a different extent? Building on gender role theory and emerging evidence from most recent research, we theorize that prominence likely affects women's experiences of abuse more so than men's. Recent scholarship examining political violence against politicians more broadly — not just online — suggests that the most pronounced gender gap in violence is observed among politicians high in the political hierarchy (Håkansson 2021; Herrick et al. 2021). Both of these studies rely on elite survey data from local level elections, thus demonstrating how female mayoral candidates in Sweden (Håkansson 2021) and female mayors in the US (Herrick et al. 2021) are the most likely to experience meaningful levels of violence and psychological abuse. Furthermore, Collignon and Rüdig's (2021) candidate survey data from the 2019 UK general election suggest that candidate gender remains a significant predictor of abuse, even when controlling for candidate viability. These findings, however, all rely on self-reported instances of harassment and abuse, which men and women may not always be equally inclined to do. As such, substantiating past survey findings with data from other observational sources is needed. And while there is some evidence from the US and Canadian context suggesting women with higher follower-count to receive more uncivil messages on Twitter than their comparable male counterparts (Rheault, Rayment, and Musulan 2019), the potentially gendered effect of status-led prominence on online incivility is yet to be established outside the North American context.

Gender role theory offers some reasons why women and men of similar prominence may experience different levels of online incivility. Due to the ever-persistent traditional gender role expectations, female politicians are often seen as violating their perceived status (Eagly and Karau 2002). By taking this argument further, Håkansson (2021) suggests that positions higher in the political hierarchy not only demand more assertiveness and confidence but are also more associated with power-seeking behavior than lower-level positions. Hence, if attacks toward women in politics are driven by a dislike of female politicians, women at higher positions of power are more visible and more likely to be perceived as violating the gender role expectations, leading us thus to a more nuanced hypothesis:

H2b Prominence has a larger differential effect on the proportion of online incivility received by female politicians than by male politicians.

Does Uncivil Content Differ by Gender?

While past social media research on the topic has relied on counting tweets that fall into operational definitions of incivility or intolerance, most women politicians who have publicly reported online incivility note that attacks toward them are not merely about "how much," but more about the nature of incivility (Erikson, Håkansson, and Josefsson 2021).

Some past social media research has differentiated less severe attacks from more severe types of incivility. "Milder" versions include name-calling, mockery, character assassination, and belittling or insulting others (Borah 2012; Sobieraj and Berry 2011), while the use of homophobic, racist, and sexist language exemplify "heavier" attacks (Munger 2016; Papacharissi 2004). Incivility, thus, is understood as a continuum in which civil language lies on one end, impoliteness or mildly uncivil language, such as sarcasm and insults, lie somewhere in the middle, and strongly uncivil language, such as racial slurs and obscenity, at the other end of the scale (Sydnor 2018). Specifically, Papacharissi (2002, 267) argues that incivility differs from mere impoliteness in that it demonstrates offensive

https://doi.org/10.1017/S1743923X25100111 Published online by Cambridge University Press

behavior toward particular social groups (such as women), thereby disrespecting collective democratic traditions. While such definitions do not explicitly distinguish identity-based attacks from other examples of incivility, they tend to consider racist or sexist abuse at the more severe end of uncivil communication. Others have argued that the consideration of uncivil behavior alone is insufficient for understanding how uncivil discourse might be threatening democratic values. Rossini (2022) suggests a conceptual distinction instead, where incivility — often occurring alongside meaningful discursive engagement — should be distinguished from intolerance that tends to occur in specific and more homogeneous discussions around minorities and civil society.

Literature on gender-based political violence offers further considerations for operationalizing such attacks. Most definitions in this literature consider (1) aggressive acts aimed largely or solely at women in politics, (2) because they are women, often using gendered means of attack, and (3) with the goal of deterring their participation to preserve traditional gender roles and undermine democratic institutions as types of violence against women in politics (Krook 2017). While it refrains from classifying some forms of abuse as more severe than others, existing scholarship nevertheless distinguishes between hostile and benevolent sexism (Chen et al. 2020; Glick and Fiske 2018). Benevolent sexism is defined as a behavior that perpetuates stereotypical attitudes toward women ("women should stay at home"), thus directly linking to Papacharissi's (2002) definition of incivility, but also demonstrates a lack of tolerance — as per Rossini's (2022) account — for specific social groups (women) who challenge the stereotypically male-dominated political arena (Krook 2017).

Yet not all attacks that occur against women in politics need to be necessarily gendered as "women can be victims of happenstance or when generalized political violence harms men and women in roughly equal proportions and in the same way" (Bardall, Bjarnegård, and Piscopo 2019, 932). Whether a specific act of incivility online qualifies as gendered or not would depend on whether: (1) the perpetrators have *gendered motives* and use violence to preserve men's control of politics; (2) *gendered forms* result in men and women experiencing the abuse in different ways; or (3) the subjective meaning-making processes of the recipients of abuse lead to *gendered impacts* (Bardall, Bjarnegård, and Piscopo 2019).

Against this background, we expect female politicians to receive different and/or additional types of incivility than those directed at men. While past research does not offer consistent evidence that men receive less severe attacks than women, we expect the latter to be more susceptible to *gendered* means of incivility. Examples of gendered attacks may also include, but are not limited to, references to sexual identity or morality (i.e., accusations of being a bad wife or a bad mother, name-calling, belittling, etc.). And while these may not include words that are often classified as offensive (making them difficult to capture empirically — for examples see Siegel 2020), they may nevertheless be harmful and have serious repercussions for one's career. Considering this, we expect gendered means of incivility to be less relevant for men, as men's participation in politics is considered more "normal" (Eagly and Karau 2002; Meeks 2019).

https://doi.org/10.1017/S1743923X25100111 Published online by Cambridge University Press

However, this does not mean that male politicians would never be susceptible to gendered incivility. Moreover, besides gendered attacks, some men (and women) may experience other types of identity-based incivility, either due to their racial or ethnic identity, age, disability, or sexual orientation.

We thus generate the following hypothesis:

H3 Uncivil language received by female politicians is more gendered than that received by male politicians.

Data & Methods

Data Collection & Case Selection

We collect a diverse set of replies to and mentions of politicians on Twitter from four countries, resulting in a unique, multi-layered and extremely rich multilingual and cross-national dataset. Our cross-national approach represents a novelty in social media research, which has primarily focused on single-country case studies (most commonly on the US), often including only high-profile politicians.

The first layer includes data from the 2014 European Parliament election in Germany, Spain, and the UK, where we collected Twitter account details for each identified candidate before the elections (January–April 2014).[3] Kantar Public[4] used the list of candidates and accounts to collect all Twitter communication centered around the candidates during and after the election campaign (May 5, 2014–June 1, 2014) using the Twitter firehose. This includes a total of 341 MEP candidates in the UK, 266 in Spain, and 163 in Germany.

As EP candidates are often regarded as second-tier politicians, who are less known and thus potentially less exposed to vitriolic comments, we supplement these data with incumbent MPs in these three countries and the members of the US Congress. For each of the countries, we compiled a list of sitting legislators and manually identified whether they had an active Twitter account as of February 2016 (September 2016 for the US).[5] This includes a total of $N = 487$ accounts in the US, $N = 740$ in the UK, $N = 344$ in Spain, and $N = 375$ in Germany.[6] For these politicians, we collected the data directly via the Twitter Streaming API: in the US, October 17, 2016–October 15, 2017;[7] and in the European countries, December 19, 2016–October 15, 2017. The data include all tweets that are either direct replies to a politician's post or that mention any politicians' Twitter accounts,[8] a total of 15,456,186 tweets in the US, 4,017,468 tweets in the UK, 3,117,873 tweets in Spain, and 779,443 tweets in Germany.

Besides including a wide variety of politicians from different countries, our data also spans a long time period, including a series of less or more politicized events, like the EP elections (May 2014), the 2016 US elections, the debate following the Brexit referendum in the UK (June 2016), the government formation process in Spain (October 2016), and a series of state-level (May 2017: Schleswig-Holstein and North Rhine-Westphalia) and federal elections in Germany (September 2017). The lengthy time span allows us to explore incivility in both periods of heated debate and during politically quieter times. Furthermore, the selection of countries in our sample does not only vary with regards to levels of incivility on Twitter, but also in terms of electoral and party systems. We include countries with strong multiparty systems that use proportional representation (Germany and Spain), where focus on individual candidates and personalities is less pronounced, in comparison to the more majoritarian/plurality systems used in the UK and the US. This allows us to present first descriptive circumstantial evidence of potential cross-system differences in politicians' experiences on social media and allow for a more nuanced examination of female and male politicians' Twitter experiences. While our case selection includes countries with more and less prevalent traditional gender norms, our sample is missing a country with very high levels of gender equality (i.e., Sweden or Finland). As such, our findings are likely generalizable to most industrialized Western democracies apart from Northern European countries.

Measuring Incivility

To classify tweets as uncivil, we manually labeled nearly 30,000 random sample of tweets,[9] which we then used to train a machine learning classifier. The manual coding scheme was developed by the authors for a larger project, containing various categories related to tweet content, including incivility. Further information on the manual labelling process, including coder training, is presented in Appendix A.

Introduction

- In this job, you will be presented with tweets about the 2014 European elections. You will need to classify each tweet into the following series of categories:

[…]

Civil *vs* Uncivil

- **Civil**: a tweet that adheres to politeness standards, i.e. it is written in a well-mannered and non-offensive way.
    - *@paulmasonews why doesnt #EU take a longer term view? Doesnt #Germany remember their 1940s bailout allowing recovery & growth? #Greece*
- **Uncivil**: an ill-mannered, disrespectful tweet that may contains offensive language. This includes:
    - threatening one's rights
    - assigning stereotypes or hate speech ("n***r", "faggot")
    - undermining or excluding a social group ("women should stay at home rather than do politics", "what do men know about childcare?")
    - threatening one's rights (freedom to speak, life preferences)
    - name-calling ("weirdo", "idiot")
    - aspersion ("liar", "traitor")
    - pejorative speak or vulgarity
    - sarcasm
    - ALL CAPS
    - incendiary, obscene, humiliating

Examples

- *@SLATUKIP – "@DavidCoburnUKip Oh shut up David. You're a bore. @marley68xx"*
- *@NicolaSturgeon You're embarrassing yourself and Scotland. Let the grown ups work it out.*
- *@SherryT: I'm intelligent enough to know who to discuss scientific matters with and who not to. You are too emotional, you'd be better off walking around carrying a metoo sign. I look forward to your reply, I enjoy a good laugh.*

We then used the 29,474 human-coded tweets[10] to train machine learning models (logistic regression) that predict incivility based on the words in a tweet. The text is preprocessed by replacing named entities, mentions, numbers, and URLs with placeholders (e.g., PERSON, MENTION, NUMBER, URL); removing punctuation and stopwords; and lemmatization of the words. Hence if these tweets contain any relevant textual information, they were annotated by humans based on the relevant text. If they do not contain any relevant textual information (i.e., they only include images or links) we exclude them from our analysis. Subsequently, unigrams are transformed into a numerical matrix using term frequency-inverse document frequency (tf-idf). Prediction performance is reported using 10-fold cross validation: we split the data in 10 subsets and use nine subsets for training and one for testing. This is repeated 10 times, so that each subset is used exactly once for testing. The model hyperparameters[11] were optimized with five-fold cross validation on the training data. The out-of-sample performance of the machine learning classifiers is summarized in Table 1. Area Under the ROC Curve (AUC) measures the performance of a classifier in predicting the *probability* of a tweet being uncivil, independent of the frequency of the classes.[12] The F1 score measures the accuracy of a classifier in assigning the *actual label* (uncivil or not) for a certain probability threshold.[13] The F1 scores for all countries are low,[14] indicating that the classifiers are not sufficiently accurate in assigning the correct label per tweet. However, high AUC (above 80%), indicates that the classifier *probabilities* are relatively trustworthy in all countries, except for Germany.[15] Considering the prediction uncertainty, we base further analysis on the predicted *probability* of a tweet being uncivil and exclude Germany from the main quantitative analysis.[16]

Another advantage of using the probability-based measure of incivility is that it provides a little bit more nuance to our otherwise binary classifier (civil/uncivil). Unfortunately, classifying multiple types of incivility would have

Table 1. Out-of-sample performance of machine learning classifiers to predict incivility

	AUC	F1	Recall	Precision
UK	0.82 (+/− 0.02)	0.52 (+/− 0.04)	0.53 (+/− 0.05)	0.51 (+/− 0.08)
Spain	0.80 (+/− 0.03)	0.54 (+/− 0.03)	0.56 (+/− 0.07)	0.53 (+/− 0.05)
Germany	0.67 (+/− 0.02)	0.44 (+/− 0.02)	0.60 (+/− 0.06)	0.35 (+/− 0.03)
US	0.83 (+/− 0.01)	0.65 (+/− 0.02)	0.63 (+/− 0.04)	0.67 (+/− 0.03)

Notes: For the US, we included synthetic labels (see details in Yannis Theocharis et al. 2020) for an additional set of 16,000 tweets.

AUC measures the probability that the model ranks a random positive example higher than a random negative example (Provost and Fawcett 2013). A perfect AUC is 1, while 0.5 indicates random performance.

F1 is the harmonic mean of precision and recall, with values ranging from 0 to 1. The F1 metric depends on the distribution of the classes. A baseline model always predicting "uncivil" would score 0.33 for the European sample (20% uncivil tweets) and 0.55 for the US sample (38% uncivil tweets). All F1 scores in the table exceed baseline performance and our US F1 score is comparable to that of others using the same Twitter dataset (Davidson, Sun, and Wojcieszak 2020).

Recall is the percentage of uncivil tweets (as determined by human annotators) that are correctly identified by the model as uncivil.

Precision is the percentage of tweets predicted by the model to be uncivil that are actually uncivil, according to human annotators.

required significantly more training data, and capturing more nuanced categories would likely have resulted in even lower classifier accuracy. However, to address the issue of differentiating between various forms of incivility, we supplement our main quantitative analysis with inductive text analysis of uncivil tweets. This additional analysis gives us both deeper insights into the different types of incivility as well as provide a crude concept validity check.

Analysis

Who Experiences More Incivility?

We first examine if incivility varies across gender and whether other factors moderate the gender effect.[17] We use machine learning output to compute the outcome variable, capturing the *probability* of a politician to receive uncivil comments, by averaging the probability of being uncivil across all tweets addressed at a given politician.[18] The dependent variable is thus the average probability of tweets addressed to a politician being uncivil.

Derived from our theorizing, the main explanatory variables are gender[19] and prominence. We operationalize prominence by distinguishing the status of the politician from their Twitter visibility. In both datasets, Twitter visibility is measured by the (logged) number of followers.[20] To capture politician's status-led prominence, we distinguish between Member of the House and Member of the Senate in the US. In Europe, we distinguish between a *top tier* (MPs who are government ministers), a *second tier* (MPs who do not hold positions in the government), a *third tier* (2014 EP candidates in safe party list positions), and a *fourth tier* (EP candidates who had little or no chance of being elected) politicians.[21]

We also control for ideology and ideological extremism. Ideology is measured with a party-level left-right score, captured by Chapel Hill Expert Survey's (CHES) (Bakker et al. 2020) in Europe and by the party label (Democrats and Republicans) in the US. In Europe, we capture ideological extremism with the absolute value of the difference between the left-right CHES placement and the theoretical midpoint of the scale (i.e., 5), while in the US it is measured as the absolute value of DW-NOMINATE (a metric of ideology estimated using roll-call votes in Congress; see Lewis et al. 2021). We further include controls for government party (only available for European data), age (only available for the US), and the percent of total tweets addressed to a politician that are replies. All the explanatory variables are re-scaled to take values between 0 and 1, allowing for substantive interpretation of the effects.

We employ linear models to account for the continuous nature of our outcome variable and for ease of interpretation.[22] As the set of predictors varies slightly from the US to Europe, we run separate models and due to the structure of the European data (i.e., politicians nested in parties and countries), we run a series of multilevel models with country fixed effects. In both cases, we first report models, where the only explanatory variable is gender (Model 1 in Tables 2 and 3). To test our main hypotheses about the effect of gender (H1) and status-led and Twitter-based prominence (H2a), Model 2 in Table 2 and Models 2 and 3 in Table 3 include variables capturing politician's gender, their status-led prominence, and their

Table 2. Predicting incivility, US

	Model 1	Model 2	Model 3
Variables	Baseline	Controls	Interaction
Female	−0.58 (0.62)	−0.07 (0.48)	−5.25 (5.12)
Log-followers (0–1)		18.57 (3.12)***	17.08 (3.38)***
Senate		−0.25 (0.63)	0.17 (0.69)
Independent		−7.75 (4.14)	
Republican		0.71 (0.42)	0.73 (0.42)
Extremism		6.86 (1.38)***	6.61 (1.39)***
Age		−1.01 (0.98)	−1.12 (0.98)
Replies		9.52 (0.59)***	9.51 (0.59)***
Female X Senate			−2.03 (1.36)
Female X Followers			8.57 (8.03)
Constant	15.78 (0.28) ***	−6.22 (1.98)**	−5.18 (2.18)*
R2	0.00	0.48	0.48
Adj. R2	−0.00	0.47	0.47
Num. obs.	490	487	486

Notes: Outcome variable denotes the probability of a tweet being uncivil averaged across tweets of a given politician and is measured on a 0 to 100 scale. Linear estimations with standard errors in parentheses.
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Twitter-visibility as well as controls for ideology and extremism (Model 2 in Tables 2 and 3).[23] To test our Hypothesis about whether prominence affects women and men's experiences on Twitter the same way or differently (H2b), we present Model 3 in Table 2 and Model 4 in Table 3 that include the relevant interaction effects.

Focusing on our central variable of interest — gender — we do not find consistent support for H1. In both the US (Table 2) and Europe (Table 3), gender is not a statistically significant predictor across models that include controls (except Model 4 in Table 3), tentatively suggesting that there is no consistent difference in the proportion of incivility received by female and male politicians. The only exceptions are Model 1 and 4 in Table 3. In a model with no controls (Model 1, Table 3), European men have a slightly higher likelihood of receiving proportionally more uncivil tweets than women. Yet, the effect size is small, and it disappears once controlling for prominence — the most important predictor of incivility, thus supporting H2a. However, when including interaction effects between gender and prominence (Model 4, Table 3), the sign of the main effect of gender flips and becomes significant, indicating that prominence affects women's and men's likelihood to receive online incivility in varying ways in our European sample. Notably, women in less prominent positions and with

https://doi.org/10.1017/S1743923X25100111 Published online by Cambridge University Press

Table 3. Predicting incivility, only Spain and UK

	Model 1	Model 2	Model 3	Model 4
Variables	Baseline	Controls	Controls	Interaction
Female	−0.51 (0.24)**	−0.19 (0.18)	−0.20 (0.18)	2.96 (0.84)***
Spain	6.58 (0.74)***	6.21 (0.39)***	5.32 (0.57)***	5.28 (0.59)***
Log-followers (0–1)		7.41 (0.81)***	5.71 (0.81)***	7.62 (0.94)***
MP		5.61 (0.27)***	1.39 (0.79)*	1.20 (0.81)
Safe EP seat		−0.36 (0.45)	−0.38 (0.45)	−0.40 (0.55)
Cabinet member		8.05 (0.79)***	3.93 (1.08)***	3.48 (1.25)***
Government party		−1.45 (0.63)**		−0.46 (1.10)
Left-right		4.74 (1.08)***		
Extremism		1.26 (0.92)		
Replies			4.63 (0.82)***	4.65 (0.82)***
Female X MP				0.74 (0.45)
Female X Safe EP				0.25 (0.88)
Female X Cabinet member				1.41 (1.51)
Female X Followers				−6.38 (1.65)***

(Continued)

Table 3. *Continued*

	Model 1	Model 2	Model 3	Model 4
Variables	Baseline	Controls	Controls	Interaction
Constant	10.58 (0.52)***	4.17 (0.51)***	5.72 (0.54)***	4.84 (0.60)***
AIC	5973.52	4703.80	5161.92	5147.81
BIC	5998.45	4762.35	5211.58	5222.30
Log likelihood	−2981.76	−2339.90	−2570.96	−2558.91
Obs. (candidates)	1080	972	1060	1060
Obs. (parties)	48	20	48	48
Var: Intercept	3.67	0.37	2.26	2.51
Var: Residual	13.97	7.18	7.15	7.06

Note: Outcome variable denotes the probability of a tweet being uncivil averaged across tweets of a given politician and is measured on a 0 to 100 scale. Linear estimations with standard errors in parentheses.
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

fewer Twitter followers are more likely to receive uncivil content than comparable male politicians.

In the US, the more objective prominence measure (i.e., the difference between being a Senator or a House Member) does not seem to matter, while Twitter visibility has a strong positive and statistically significant effect (Model 2, Table 2). Politicians with the highest number of followers are expected to receive, on average, 19 percentage points more uncivil content in comparison to a politician with the fewest followers. In Europe, Twitter visibility matters, too, though the difference is relatively smaller (only approximately 6 percentage points [Model 3, Table 3]). Yet, status-led prominence matters, too, with MPs and Cabinet ministers receiving, on average, respectively approximately 6 and 8 percentage points more uncivil content in comparison to MEP candidates (Model 2, Table 3). The broader range of politicians in our European sample may explain why status-led prominence only matters in Europe.

While the prominence does not moderate the effect of gender in the US, we find the opposite effect to our expectation in Europe (H2b) (Model 4, Table 3), where Twitter visibility substantially increases the proportion of incivility male politicians receive, without having any statistically significant impact on the incivility received by women (see Figure 1).[24] To put this into perspective, in our sample the median number of uncivil tweets that female politicians with few followers (i.e., less than 100) receive is seven compared to one for men with a similar number of followers. In contrast, a popular male politician (i.e., more than 10,000 followers) receives in median approximately 100 more uncivil tweets than a popular female politician (529 compared to 418). Our findings thus suggest that while male politicians need to worry more about Twitter incivility once they reach higher political visibility, their female colleagues are exposed to uncivil content from the moment they enter Twitter. Notably, in a situation of receiving relatively few tweets, even just a couple of uncivil comments are likely to be noticed, and thus potentially having a stronger impact on the receiver.

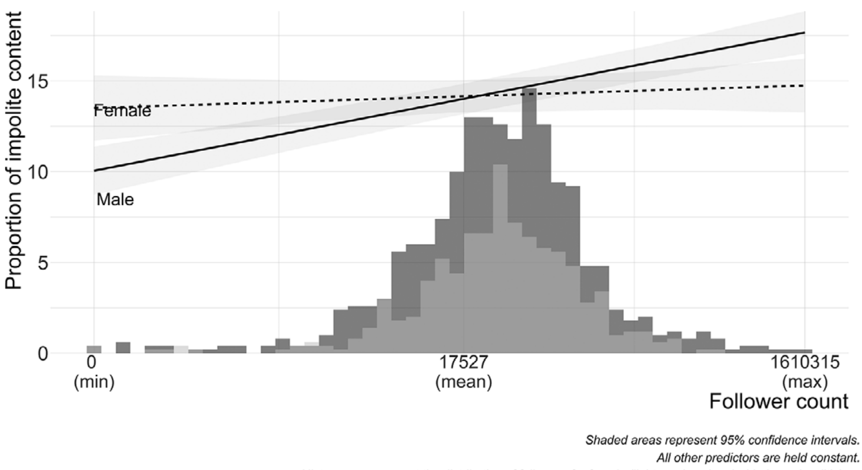

Figure 1. Conditional impact of gender on incivility depending on Twitter visibility, European countries.

https://doi.org/10.1017/S1743923X25100111 Published online by Cambridge University Press

Regarding other control variables, politicians representing right-leaning parties, regardless of their gender, are more likely to receive more uncivil content, and more ideologically extreme politicians in the US get substantially more incivility online. On both continents, direct replies in comparison to mentions attract more incivility.

Differences in the Type of Incivility

We use various text analysis methods for exploring any gender differences in the uncivil language addressed to politicians (H3). While in-depth qualitative analysis is beyond the scope of this paper, the inductive text-analysis tools used below allow us to explore if the specific words used in uncivil tweets that are targeted at women and men vary from one another. First, we use keyness measure[25] to identify words that are significantly related to one gender. Second, we perform additional analysis by training classification models that predict incivility split by gender. In both cases, we present a couple of example tweets where there are words with the highest keyness (most distinct from male versus female sample) or the words most likely to appear in tweets classified as uncivil.[26] Both analyses suggest that the language associated with uncivil content varies by gender. Put simply, the specific words in uncivil tweets aimed at male and female politicians are not the same, with words indicating more gendered attacks being more prevalent in uncivil tweets sent to women.[27]

To compute the keyness of a word *w*, we perform an association test using the χ^2 value to the frequency of *w* in the target corpus (uncivil tweets targeted to women) versus the frequency of *w* in the reference corpus (uncivil tweets received by men). We find that around 1% of the words in all uncivil tweets have a significant keyness, meaning they are significantly related to one gender rather than the other (see Table 4).

Figure 2 displays the words with highest keyness scores in uncivil tweets, some of which allude to identity-based language use, while not necessarily being intuitively predictive of incivility (i.e., "lady," "woman," "sir," and "mr" in the US or "frau" and "herr" in Germany). As a next step, we examine the individual tweets where the words with the highest keyness scores are used.

In the UK, the words with the highest keyness scores in uncivil tweets targeted at female politicians include, for example, "child," "leader," and "politician," while those received by men use words, such as "fuck," "cunt," and "old."[28] Examining

Table 4. Number of words with p-value below 0.05 for χ^2 value of words from target group (uncivil tweets to women) compared to reference group (uncivil tweets to men)

	Significant words	Percentage
UK	26	0.01
Spain	16	0.00
Germany	13	0.00
US	75	0.02

https://doi.org/10.1017/S1743923X25100111 Published online by Cambridge University Press

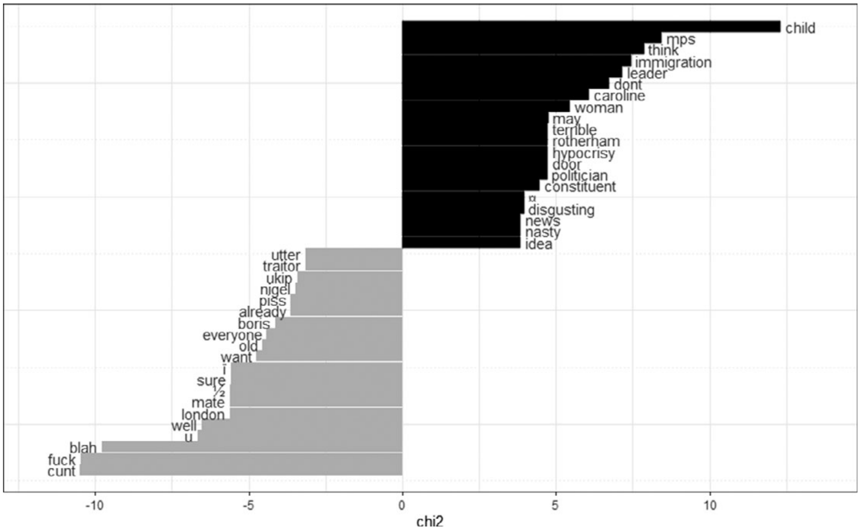

Figure 2a. Word keyness plot for uncivil tweets by gender. Black bars are associated with female and grey bars with male gender: UK.

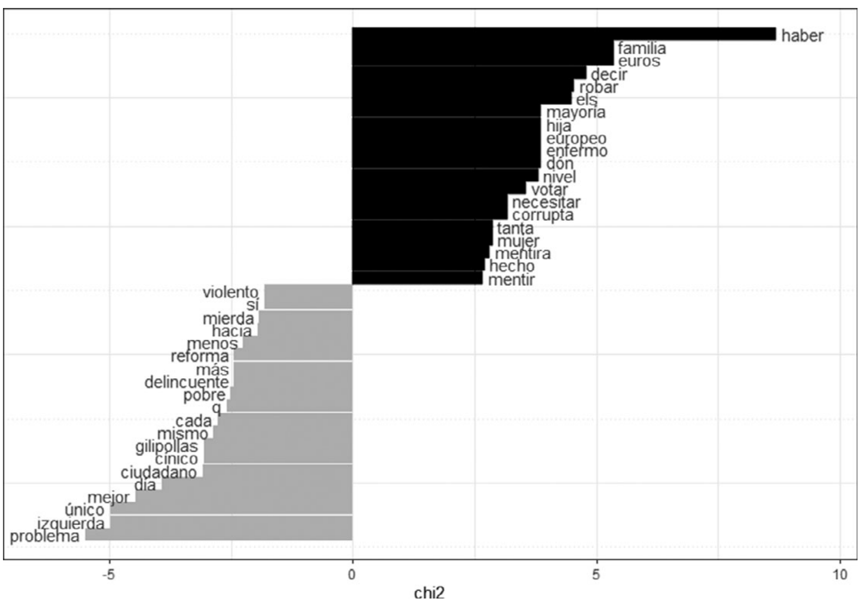

Figure 2b. Word keyness plot for uncivil tweets by gender. Black bars are associated with female and grey bars with male gender: Spain.

the tweets where these words appear suggests that the uncivil tweets sent to UK male politicians include many swear-words and name-calling:

> "@SadiqKhan Fuck off ya fake"

https://doi.org/10.1017/S1743923X25100111 Published online by Cambridge University Press

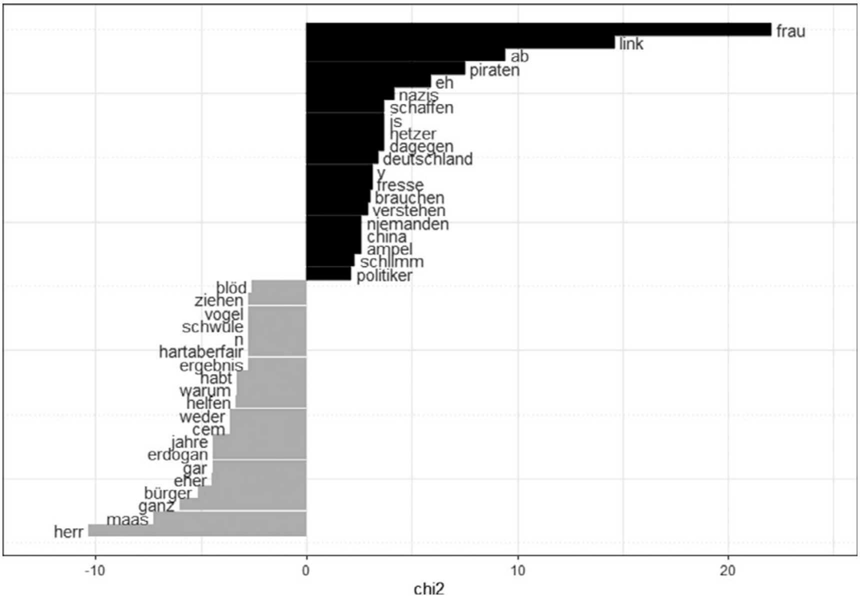

Figure 2c. Word keyness plot for uncivil tweets by gender. Black bars are associated with female and grey bars with male gender: Germany.

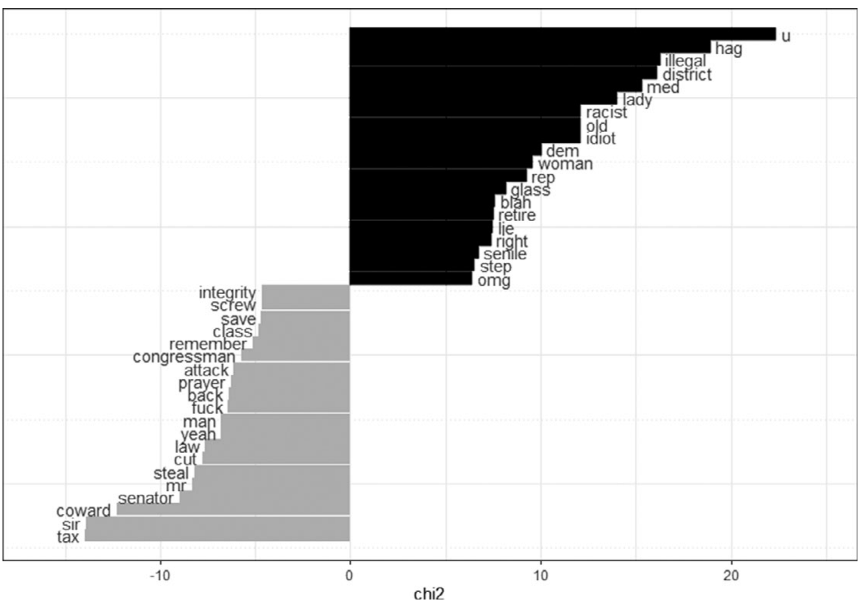

Figure 2d. Word keyness plot for uncivil tweets by gender. Black bars are associated with female and grey bars with male gender: US.

https://doi.org/10.1017/S1743923X25100111 Published online by Cambridge University Press

"@jeremycorbyn You're a fucking twat"
"@BorisJohnson You're a cunt. A sweaty, dipping, lying, absolute fucking cunt. Fuck off back to whichever swamp you somehow crawled out of"

At the same time, some of the uncivil tweets targeted at UK female politicians, where high-keyness words appear, tend to question their quality as a politician or their suitability for role models:

"@HackneyAbbott @SMCommision as a whole black children are given the wrong role models. Expectations are deliberately set low."
"@HarrietHarman and you're a danger to children!!"
"@YvetteCooperMP Now there's the truth of how incompetent & traitorous are our politicians, msm et al....."

We see some similar examples of gendered incivility targeted at female politicians in the US, too: highest scoring keyness words for women include, for example, "lady," "woman," and "hag," while those received by men contain words, such as "tax," "sir," and "coward." In both countries, some words (i.e., "child," "immigration," "tax") also allude to women and men being targeted in relation to specific issue areas. Unlike in the UK, many of the uncivil tweets targeted at the US female politicians are also heavy on profanity and name-calling — features that were more prevalent in offensive tweets sent to British men. Moreover, in the US, our sample of uncivil tweets includes references to female politicians' appearance and the questioning of some Republican female politicians' ability of and credentials in representing women. The tweets below exemplify some of the more extreme instances of gendered incivility targeted at US female politicians, with high-keyness words:

"@MaxineWaters you are a lying, crazy old hag who doesn't belong in out House of Reps. Get out"
"@NancyPelosi @HouseGOP Nancy, your profile image is more photo-shopped than mine, you nasty old anti-American hag!"
"@lisamurkowski You should be ashamed of yourself. You have capitulated to the male dominance governing [US flag] we need STRONG YOUNG women to run this country. Get yourself to a retirement home soon"

When examining uncivil tweets with high keyness score words sent to US male politicians, we observe some more policy-related tweets (i.e., about taxation) than mere personal attacks and name-calling. However, there are also several offensive tweets, including name-calling:

"@PeterRoskam NO TAX REFORM UNTIL WE SEE FAKE PRESIDENT TRUMP'S TAX RETURNS!"
"@SenJohnMcCain You should be proud. Too bad others will lose what you're getting from the taxpayers. Judas."

It is somewhat harder to identify any clear gendered incivility patterns in Germany and Spain. While the specific words with the highest keyness scores

https://doi.org/10.1017/S1743923X25100111 Published online by Cambridge University Press

vary across genders in Spain, both male and female politicians receive uncivil tweets related to corruption, questioning their dignity and suitability to partake in politics:[29]

> "@DolorsMM @cope_es And your family keeps stealing from the social security system. Public embarrassment"
> "@Rafa_Hernando @PPopular I think you are a problem for society. You're in the government and all corrupt. Either you explain why or you resign."
> "@GLlamazares What you are doing is unbelievable. Don't ever say that you are left-wing. You are a traitor."

Similarly to Spain, uncivil tweets targeted at German politicians appear to use fewer deliberately abusive words than in the US and the UK. Hence the impoliteness sounds comparatively polite. This may partly explain why our classifier performs poorly with the German sample:

> "@petertraube Thank you, Mr. Outrage-leader. But please also tell your stepmother and the courtier of the interior. They don't answer my calls..."
> "@MdB_Stroebele This woman is hard to top when we talk about double standards. An ice-cold fishing-rod!"

While it is difficult to observe coherent themes in the tweets with high keyness-scoring words targeted at German male politicians, some of the communication classified as uncivil toward women involves name-calling and questioning female politicians' ability to understand politics. The tweets below exemplify some of the more explicitly gendered uncivil tweets targeted at German female politicians:

> "@SteinbachErika Confused old women, who constantly send their nonsense via Twitter to alleviate their difficult-to-treat persecution mania."
> "@katjakipping Leftist dreamer didn't pay attention in history class and blanketly calls people Nazis. This must have consequences."

To further explore any gender differences in the type of incivility, we perform additional analysis by training classification models that predict incivility split by gender.[30] More specifically, we divide the data to tweets toward male politicians versus tweets toward female politicians, and train separate classification models to predict incivility for each group (see Appendix E, Table E1). The coefficients of these models provide further insights to the predictive terms for incivility (see Table 5).[31] While there is a core set of uncivil words — similarly to words with highest keyness scores — there are large gender differences in the top predictive terms. For example, out of the total 100 terms in the UK (50 for men and 50 for women), only 22 terms overlap, further suggesting that the type of incivility likely varies dependent on the recipient.

A cursory reading of various examples provides further circumstantial evidence of female candidates receiving more morality-focused words (vile, shame, shameful, disgraceful, hypocrite, deluded), including references to personal attacks (kill, destroy, rape), rather than simple swear words. However, words

https://doi.org/10.1017/S1743923X25100111 Published online by Cambridge University Press

Table 5. Words most likely to be associated with incivility for each gender

	United Kingdom
Male	fuck, stupid, cunt, disgrace, shit, idiot, lie, traitor, fucking, joke, liar, twat, fool, like, piss, **kill**, arse, clown, troll, utter, hate, anti, fuck, youre, *old*, racist, mouth, lol, there, stop, hes, weak, useless, *white*, gang, shame, youve, excuse, hell, tweet, oh, shut, URL, country, fake, nonsense, amp, terrorist, actually, **die**
Female	idiot, stupid, disgrace, politician, fascist, disgraceful, liar, racist, shit, rubbish, disgusting, vile, dont, lie, really, criminal, fucking, *white*, bunch, shame, silly, fuck, deluded, fool, immigration, terrible, **rape**, hate, *disabled*, awful, **destroy**, useless, mouth, joke, hypocrite, ever, like, shut, lose, nonsense, shameful, **kill**, may, choose, sick, poor, youre, yeah, clown
	Spain
Male	mierda, vergüenza, corrupto, puta, traidor, puto, basura, tonto, coño, cobarde, verguenza, ladrón, fascisto, político, cojones, sinvergüenza, cara, vergonya, cada, pena, miserable, comunisto, *español*, república, cárcel, españa, delincuente, madre, ridículo, vete, tan, ignorante, pedro, teneis, perro, ets, tras, mentira, odio, cada día, nazi, niño, nuevo, único, cinismo, meter, país, poca, verdad, ah
Female	vergüenza, mierda, corrupto, fascisto, puta, verguenza, robar, mentira, cojones, puigdemont, miserable, ladrón, tonto, sinvergüenza, teneis, traidor, coño, cinismo, **matar**, nivel, corrupción, odio, hecho, nazi, maldito, cobarde, salir, mafia, hijo, vergonya, criminal, mentir, asco, meter, gente, publicar, tan, corrup- cion, vote, país, usted, pegar, twitter, terroristo, dia, soraya, llamar, independentisto, culo, cada vez
	Germany
Male	ganz, dumm, peinlich, mal, grünen, habt, erdogan, *türken*, lieb, einfach, lügen, merkel, machen, bürger, nennen, jahre, maas, kinder, cem, eigentlich, blöd, lassen, hoffentlich, weder, grüne, immer, wissen, mußen, volk, mehr, sorry, spd, gar, ach, stasi, *kurz*, typ, na, mensch, welt, kommen, kanzlerin, wählen, teil, beispiel, *alt*, halt, größt, überhaupt, herr maas
Female	spd, dumm, mal, merkel, nazis, deutschland, *frau*, *alt*, volk, politiker, machen, *deutsch*, ahnung, eigen, immer, fresse, halten, ab, steinbach, nazi, eh, wollen, opfer, einfach, kanzlerin, echt, seid, grüne, lügen, lernen, bitten, vergessen, frau steinbach, müssen, sollen, endlich, glauben, sorry, *weeißen*, hetze, verhin- dern, wohl, bloß, cdu, tun, paar, linken, verdienen, türkei, halt
	US
Male	ass, stupid, traitor, fuck, shit, coward, hypocrite, lie, idiot, fucking, liar, shame, disgrace, hate, racist, pathetic, hell, asshole, corrupt, ignorant, disgusting, *old*, suck, crap, bullshit, dumb, mouth, clown, bitch, fake, screw, garbage, blood, dick, comment, shut, paul, fool, nothing, full, sick, moron, hypocrisy, yeah, stop, embarrassment, resign, ashamed, sexual, loser
Female	lie, idiot, stupid, liar, hypocrite, shit, racist, ass, *old*, shut, dumb, suck, shame, moron, med, ur, hell, evil, embarrassment, bitch, face, brain, worse, crazy, damn, hag, office, incompetent, delusional, fuck, job, crap, loser, scumbag, retire, actually, give, criminal, majority, asshole, useless, traitor, fraud, fucking, ashamed, person, illegal, elect, human, disgusting

indicating a reference to appearance or identity (i.e., old, white, disabled) are amongst the top predictive words for both men and women. The following tweets targeted at Labour MP Diane Abbott and US Representative for California's 43rd congressional district Maxine Waters exemplify some of the morality-focused incivility:

"@HackneyAbbott Your a disgrace women !!! Vile person"
"@MaxineWaters ur one stupid low life biotch"

At the same time, top five tweets with the highest incivility prediction scores (99%), mostly consisting of various insults and swear words, in the UK were all targeted at men.

"@jeremycorbyn fuck off you, you stupid cunt"
"@jeremycorbyn @realDonaldTrump fuck off you stupid cunt"
"@Nigel_Farage @UKIP FUCK FUCK SHIT CUNT DAAAAAMMN MOTHERFUCKER"
"@Nigel_Farage fuck the fuck off you fucking bigoted twat"
"@Jeremy_Hunt @NEAmbulance fuck off you stupid tory cunt"

Similarly, in Spain, the US and Germany, very few terms overlap across genders (with the least overlapping terms in Germany: 14). Moreover, the very word *politician* is amongst the top predictive terms for women both in Germany and the UK, but not for men, potentially indicating that female politicians are attacked with a reference to their *capacity* as politicians.

Discussion

Social media came with a big promise: a personal publicity channel, allowing traditionally disadvantaged candidates, like women, to bypass media gatekeepers and communicate their message in their own terms (Tucker et al. 2017). Focusing on theoretical ideas stemming from political communication and politics and gender research, we hypothesized that this prospect is potentially lost due to the gender differences in the quantity and type of incivility received by politicians. We tested these assumptions using a unique cross-national and multilingual dataset, capturing the Twitter experiences of politicians in different contexts and of varying levels of prestige. Our analysis relies on data from a period when Twitter was an important tool for political campaigning, but when different politicians in different countries used it to a varying extent. Also, Twitter was content moderated at a time, with some highly abusive content likely removed before the users saw it. As such, our data provide good variation for examining different types of politicians' social media experiences as well as perhaps somewhat more conservative test than more recent data from X (formerly Twitter).

Our quantitative analysis points to indirect, rather than direct, gender differences in the proportion of online incivility received, with prominence

https://doi.org/10.1017/S1743923X25100111 Published online by Cambridge University Press

moderating the effect of gender. More specifically, while Twitter-visibility matters on both sides of the Atlantic, status-led prominence only matters in Europe. This may partly be due to our US sample lacking in "low" level politicians, but more likely it reflects the more professionalized, personalized, and polarized nature of American politics. Particularities of the political systems are important here as European politicians (even cabinet members and the MPs, especially in PR systems) are considerably less visible on Twitter than their US counterparts, with European female politicians having both substantially fewer followers and receiving significantly fewer mentions and replies.[32]

These observed trends are instructive and the inclusion of cases from outside North America, allow us to present more nuanced findings. The fact that European female politicians do not experience proportionally more incivility online could point to other structural disparities. For example, female politicians in Europe could avoid building an extensive social media presence, especially because of their initial more negative experiences on the platform — as evidenced by our analysis. For example, a politician who receives only a small number of responses, might actually read and remember them. Hence, the impact of such messages could be more influential and affect politicians' future behavior more than in the case of someone receiving thousands of responses every day. Based on what we know from the US, an increased professionalization and personalization of European politics could result in a diminished gender gap in social media popularity also in Europe. However, the question of whether such changes will eradicate the gender gap in incivility across all groups or, contrarily, will only expose European female politicians to even more attacks, withstands. An even darker prediction would see the more professionalized and personalized campaigns in Europe resulting in an even greater gender gap in the willingness to run for high level office.

We also acknowledge that our findings regarding the interaction of gender and prominence go against some past research, which suggest that the most high-profile, rather than less prominent, women receive the most attacks both online and offline (Collignon and Rüdig 2021; Herrick et al. 2021; Rheault, Rayment, and Musulan 2019). However, most of this body of work relies on self-reported evidence from elite surveys, rather than content analysis of social media data. As such, due to the prevalence of gender roles and norms and its likely impact also on men, elite survey data may suffer from male politicians under reporting the incivility and attacks they receive. Additionally, male politicians enjoying high levels of prominence may be less likely to manage their own social media accounts, which could lead to further under reporting of incivility. More importantly, our results may reveal women's unwillingness to engage more actively on a platform where they receive incivility the moment they enter politics or Twitter. Furthermore, our findings are in line with some past research, that suggest young and minority representatives get the most abuse (Ward and McLoughlin 2020).

Our second major puzzle concerned the *nature* of online incivility in a cross-national setting. In line with findings from qualitative research (Erikson, Håkansson, and Josefsson 2021; Sobieraj 2020), revealing that online incivility targets women's identities, and consists of rape threats and attacks on their appearance,

https://doi.org/10.1017/S1743923X25100111 Published online by Cambridge University Press

we find that uncivil tweets directed to female politicians focused also on morality, personal attacks, appearance, and them being (un)suitable politicians. Hence, we found circumstantial evidence of gendered language in the tweets classified as uncivil targeted at women, while male politicians received incivility that was heavy on profanity and name-calling. The fact that women politicians are more likely to receive gendered and intolerant attacks, rather than random incivility and name-calling, points to potentially serious consequences. Communicative environments in which references to one's very identity are used as a potential silencing tool may reveal a broader political culture rife with intolerance. As such, our findings point to potential policy implications, where a further need to regulate social media content more similarly to traditional media may be necessary to ensure the equality of political representation.

Importantly, our study goes beyond anecdotal evidence, by providing valuable insights into quantitative and qualitative gender differences in uncivil language targeted at politicians in a variety of contexts. The application of machine learning methods enables us to analyze millions of tweets in four different countries but does not come without limitations. Our choice to work with logistic regression classifiers is motivated by the ease of interpretation — the words or topics with the largest coefficients are the most predictive of uncivil language — and higher or similar performance to other classifiers.[33] Even though our models are competitive with other available alternatives, their performance is still far from perfect, especially for Germany. As mentioned above, this might be due to more subtle incivility. Similarly, while the use of binary classification does not automatically allow us to easily distinguish problematic tweets from profanity, a more qualitative reading of the tweets classified as "uncivil" showed we nevertheless captured a range of different types of abuse. Yet, by solely focusing on word usage, the method may miss some of these forms of incivility. More advanced NLP methods, which consider the context for each occurrence of a given word would have limitations, too, as these do not yet understand all subtleties of natural language. This reaffirms what other political scientists studying hateful language have stressed: automated machine learning methods have difficulties in capturing the wide variety of more subtle but nevertheless harmful uncivil language (Siegel 2020), some of which may be more often used against women than men. This is not surprising, as — based on the ICR scores — even humans struggle with consistently identifying uncivil communication. Regardless of these limitations, our results show important cross-national differences in the amount and type of incivility targeted at female and male politicians.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S1743923X25100111.

Notes

1. The South Korean incumbent party, for example, developed a "Twitter Influence Index," used as a criterion for candidate selection (Lee 2013).

2. Prominence can be due to the politician's status as Senator or cabinet member or due to their Twitter-celebrity status.

3. In all EU countries, MEPs are elected using proportional electoral system.
4. Kantar Public, renamed as Verian in 2023, is a leading social research agency for UK and international policymakers, producing several flagship UK government surveys including continuous and longitudinal studies.
5. The US list was obtained from the United States GitHub account.
6. We note that some of the accounts are inactive, leading our final samples of politicians to be smaller than the total number of accounts identified.
7. We focus only on those politicians who served both in the 114th Congress and in the 115th Congress. Including those who were newly elected in 2017 to the 115th Congress leads to the same substantive conclusions.
8. If more than one account is addressed, we limit the analysis to the first account.
9. The US tweet sample is not fully random as we supplemented the initial random sample with tweets that had a higher probability of incivility.
10. The size of our training data for the US was increased by using synthetic labels for an additional set of 16,000 tweets using Google's Perspective API (see details in Nulty et al. 2016).
11. We used L2 regularization and optimized the regularization parameter C in the range [0.00001,0.0001,0.001,0.01,0.1,1,1 for AUC.
12. AUC is the probability that the model ranks a random positive example higher than a random negative example (Provost and Fawcett, 2013). A perfect model would achieve an AUC of 1, while an AUC of 0.5 indicates a random model.
13. F1 score is the harmonic mean of precision and recall: $F1 = 2x\frac{precision\,x\,recall}{precision + recall}$, and lies between 0 and 1. Recall is the percentage of tweets in a given category (according to human annotators) that are correctly classified. Precision is the percentage of tweets predicted to be in a given category that are correctly classified.
14. F1 scores depend heavily on the frequency of classes. A baseline model that always predicts the class of interest (i.e., uncivil) will achieve an F1 score of $\frac{2r}{r+1}$ with r the proportion of uncivil tweets. For the European sample (20% uncivil tweets), the baseline F1 score is 0.33 and for the US (38% uncivil tweets) it is 0.55. Thus all F1 scores are well above baseline performance and our US F1 score is comparable to that of others using the same Twitter dataset (Davidson, Sun, and Wojcieszak 2020).
15. As the most predictive words in Table 5 for Germany are more neutral than in other countries, incivility in Germany appears more subtle and harder to detect.
16. See Supplementary Appendix B for country-specific analysis, including Germany (see Table B2). Given that the patterns of results presented in Table B1 Models 3 to 6 (i.e., for UK and Spain) are similar to the one presented in Table 3, we chose to report the pooled analysis in the main body of the paper as this streamlines the analysis (i.e., by reducing the number of tables and models) and also strengthen the robustness of our findings by increasing the sample size. We note that all our conclusions hold, also when using tweet-level analysis (Appendix C).
17. As the performance of our machine learning process is far from optimal, we also ran the models using a random sample of human coded data (see Appendix D). The effects point in the same direction.
18. This correlates at 0.91 level with an alternative measure that reflects the proportion of uncivil tweets (defined as tweets that have a probability > 0.5 to be uncivil) addressed at the politician.
19. In the US, we use the Voteview data (Lewis et al. 2021), while in Europe our human coders assigned gender based on the Twitter profile and available biographical information.
20. We collected these data in April 2017 in the US and in March 2015 in Europe. As the number of followers can depend on the population of the country, Appendix B Table B3 presents results, using the number of followers weighted by the population size in millions, leading to almost identical results.
21. We use candidate list position at the EP elections, rather than their incumbency status, as the measure of prominence due to party-determined list position being the most important measure of likely campaign intensity and electoral success in a closed list system.
22. As the DV is bounded between 0 and 100, we also ran beta regression models, which lead to substantively identical results.

23. As we miss information on party ideology and extremism on some of our European sample (i.e., candidates from small parties that are not covered by CHES), we present an additional model including all candidates (Model 3 in Table 3). The results also hold with the reduced European sample, controlling for ideology and extremism (see Table B5 in Appendix B).

24. Even if these results are driven by the Spanish and UK data, we note that this effect is substantially the same in a pooled analysis that also includes the US (see Table B4 in Appendix B).

25. Keyness is a measure part of quanteda R package https://quanteda.io/reference/textstat_keyness.html.

26. The example tweets are selected based on the presence of words that are the most distinct across genders (keyness) and that are the most explicit examples of uncivil tweets.

27. We also include German tweets in this additional analysis to explore if the uncivil content appears different from other countries, which may potentially help explain the poorer performance of our classifier.

28. We would also like to note the differences in the specific words used in different English-speaking countries in our sample, with some swear-words that are traditionally used against women, more likely to be used as offensive terms when talking about male politicians in the UK.

29. See Appendix E for the original tweets in Spanish and German together with the English translation.

30. We also perform topic modeling to find themes/topics that are associated with gender (see Table F1 in Appendix F). The results, however, do not reveal sufficiently clear patterns for additional insight (see Table F2).

31. The most predictive bigrams can be found in Tables E2 and E3 in Appendix E.

32. On average, European female MPs have 39% less followers than male politicians (12,686 vs. 20,873), and they receive on average 51% less mentions and replies (2,745 vs. 5,556).

33. Other classifiers in our benchmark include (Multilayer) Perceptron, Lasso Regression, Linear Regression, Support Vector Machine, Naive Bayes, Decision Tree, and Random Forrest.

References

Aaldering, Loes, and Daphne Joanna Van Der Pas. 2020. "Political Leadership in the Media: Gender Bias in Leader Stereotypes during Campaign and Routine Times." *British Journal of Political Science* 50 (3): 911–31.

Bakker, Ryan, Liesbet Hooghe, Seth Jolly, Gary Marks, Jonathan Polk, Jan Rovny, Marco Steenbergen, and Milada Anna Vachudova. 2020. "1999-2019 Chapel Hill Expert Survey Trend File'Version 1.2." *Available on chesdata.eu.*

Barberá, Pablo, and Thomas Zeitzoff. 2017. "The New Public Address System: Why Do World Leaders Adopt Social Media?" *International Studies Quarterly* 62 (1): 121–30.

Bardall, Gabrielle, Elin Bjarnegård, and Jennifer M. Piscopo. 2019. "How Is Political Violence Gendered? Disentangling Motives, Forms, and Impacts." *Political Studies* 68 (4): 916–35.

Borah, Porismita. 2012. "Does It Matter Where You Read the News Story? Interaction of Incivility and News Frames in the Political Blogosphere." *Communication Research* 46 (6): 809–27.

Chen, Gina Masullo, Paromita Pain, Victoria Y. Chen, Madlin Mekelburg, Nina Springer, and Franziska Troger. 2020. "'You Really Have to Have a Thick Skin': A Cross-Cultural Perspective on How Online Harassment Influences Female Journalists." *Journalism* 21 (7): 877–95.

Collignon, Sofia, and Wolfgang Rüdig. 2020. "Harassment and Intimidation of Parliamentary Candidates in the United Kingdom." *The Political Quarterly* 91 (2): 422–9.

Collignon, Sofia, and Wolfgang Rüdig. 2021. "Increasing the Cost of Female Representation? The Gendered Effects of Harassment, Abuse and Intimidation towards Parliamentary Candidates in the UK." *Journal of Elections, Public Opinion and Parties* 31 (4): 429–49.

Davidson, Sam, Qiusi Sun, and Magdalena Wojcieszak. 2020. "Developing a New Classifier for Automated Identification of Incivility in Social Media." In *Proceedings of the Fourth Workshop on Online Abuse and Harms*, 95–101.

Demos. 2016. "The Use of Misogynistic Terms on Twitter." *Technical Report*. London: Demos.

https://doi.org/10.1017/S1743923X25100111 Published online by Cambridge University Press

Eagly, Alice H., and Steven J. Karau. 2002. "Role Congruity Theory of Prejudice Toward Female Leaders." *Psychological Review* 109 (3): 573.

Enli, Gunn Sara, and Eli Skogerbø. 2013. "Personalized Campaigns in Party-Centred Politics." *Information, Communication & Society* 16 (5): 757–74.

Erikson, Josefina, Sandra Håkansson, and Cecilia Josefsson. 2021. "Three Dimensions of Gendered Online Abuse: Analyzing Swedish MPs' Experiences of Social Media." *Perspectives on Politics* 21 (3): 896–912.

Friedersdorf, Conor. 2015. "In Defense of Civility on Twitter." *The Atlantic.* http://www.theatlantic.com/politics/archive/2015/11/in-defense-of-civility-ontwitter/413860/.

Glick, Peter, and Susan T. Fiske. 2018. "The Ambivalent Sexism Inventory: Differentiating Hostile and Benevolent Sexism." In *Social Cognition*, 116–60. Routledge, London, UK.

Gorrell, Genevieve, Mehmet E. Bakir, Ian Roberts, Mark A. Greenwood, and Kalina Bontcheva. 2020. "Which Politicians Receive Abuse? Four Factors Illuminated in the UK General Election 2019." *EPJ Data Science* 9 (1): 18.

Greene, Zachary, and Maarja Lühiste. 2018. "Symbols of Priority? How the Media Selectively Report on Parties' Election Campaigns." *European Journal of Political Research* 57 (3): 717–39.

Greenwood, Mark A., Mehmet E. Bakir, Genevieve Gorrell, Xingyi Song, Ian Roberts, and Kalina Bontcheva. 2019. "Online Abuse of UK MPs from 2015 to 2019." *arXiv preprint arXiv:1904.11230.*

Håkansson, Sandra. 2021. "Do Women Pay a Higher Price for Power? Gender Bias in Political Violence in Sweden." *The Journal of Politics* 83 (2): 515–31.

Hayes, Danny, and Jennifer L. Lawless. 2016. *Women on the Run: Gender, Media, and Political Campaigns in a Polarized Era*. Cambridge University Press, New York, US.

Herrick, Rebekah, Sue Thomas, Lori Franklin, Marcia L. Godwin, Eveline Gnabasik, and Jean Reith Schroedel. 2021. "Physical Violence and Psychological Abuse against Female and Male Mayors in the United States." *Politics, Groups, and Identities* 9 (4): 681–698.

Inter-Parliamentary Union. 2016. "Sexism, Harassment and Violence Against Women Parliamentarians." Technical Report, Inter-Parliamentary Union, Geneva.

Krook, Mona Lena. 2017. "Violence Against Women in Politics." *Journal of Democracy* 28 (1), 74–88.

Krook, Mona Lena and Juliana Restrepo Sanín. 2016. "Gender and Political Violence in Latin America. Concepts, Debates and Solutions." *Política y gobierno* 23 (1), 127–62.

Lee, Eun-Ju. 2013. "Effectiveness of Politicians' Soft Campaign on Twitter Versus TV: Cognitive and Experiential Routes." *Journal of Communication* 63 (5): 953–74.

Lewis, Jeffrey B., Keith Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin, and Luke Sonnet. 2021. "Voteview: Congressional Roll-Call Votes Database." Accessed June 2, 2021. https://voteview.com.

Lühiste, Maarja, and Susan Banducci. 2016. "Invisible Women? Comparing Candidates' News Coverage in Europe." *Politics & Gender* 12 (2): 223–53.

Meeks, Lindsey. 2019. "Owning Your Message: Congressional Candidates' Interactivity and Issue Ownership in Mixed-Gender Campaigns." *Journal of Information Technology & Politics* 16 (2): 187–202.

Muddiman, Ashley. 2017. "Personal and Public Levels of Political Incivility." *International Journal of Communication* 11: 21.

Munger, Kevin. 2016. "Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment." *Political Behavior* 39: 629–49.

Mutz, Diana C. 2015. *In-Your-Face Politics: The Consequences of Uncivil Media*. Princeton: Princeton University Press.

Nulty, Paul, Yannis Theocharis, Sebastian Adrian Popa, Olivier Parnet, and Kenneth Benoit. 2016. "Social Media and Political Communication in the 2014 Elections to the European Parliament." *Electoral Studies* 44: 429–44.

Papacharissi, Zizi. 2002." The Virtual Sphere: The Internet as a Public Sphere." *New Media & Society* 4 (1): 9–27.

Papacharissi, Zizi. 2004. "Democracy Online: Civility, Politeness, and the Democratic Potential of Online Political Discussion Groups." *New Media & Society* 6 (2): 259–83.

Provost, Foster, and Tom Fawcett. 2013. *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O'Reilly Media, Inc.

https://doi.org/10.1017/S1743923X25100111 Published online by Cambridge University Press

Rheault, Ludovic, Erica Rayment, and Andreea Musulan. 2019. "Politicians in the Line of Fire: Incivility and the Treatment of Women on Social Media." *Research & Politics* 6 (1): 2053168018816228.

Rossini, Patrícia. 2022. "Beyond Incivility: Understanding Patterns of Uncivil and Intolerant Discourse in Online Political Talk." *Communication Research* 49 (3): 399–425.

Siegel, Alexandra A. 2020. "Online Hate Speech." In *Social Media and Democracy: The State of the Field, Prospects for Reform*, eds. N. Persily and J.A. Tucker, 56–88. New York: Cambridge University Press.

Sobieraj, Sarah. 2020. *Credible Threat: Attacks Against Women Online and the Future of Democracy*. Oxford University Press, New York, US.

Sobieraj, Sarah, and Jeffrey M. Berry. 2011. "From Incivility to Outrage: Political Discourse in Blogs, Talk Radio, and Cable News." *Political Communication* 28 (1): 19–41.

Southern, Rosalynd, and Emily Harmer. 2021. "Twitter, Incivility and "Everyday" Gendered Othering: An Analysis of Tweets Sent to UK Members of Parliament." *Social Science Computer Review* 39 (2): 259–75.

Suler, Jane. 2004. "The Online Disinhibition Effect." *Cyberpsychology & Behavior* 7 (3): 321–6.

Sydnor, Emily. 2018. "Platforms for Incivility: Examining Perceptions across Different Media Formats." *Political Communication* 35 (1): 97–116.

Tenove, Chris, Heidi Tworek, Grace Lore, Jordan Buffie, and Trevor Deley. 2023. "Damage Control: How Campaign Teams Interpret and Respond to Online Incivility." *Political Communication* 40 (3): 283–303.

Theocharis, Yannis, Pablo Barberá, Zoltán Fazekas, Sebastian Adrian Popa, and Olivier Parnet. 2016. "A Bad Workman Blames His Tweets: The Consequences of Citizens' Uncivil Twitter Use When Interacting with Party Candidates." *Journal of Communication* 66 (6): 1007–31.

Theocharis, Yannis, Pablo Barberá, Zoltán Fazekas, and Sebastian Adrian Popa. 2020. "The Dynamics of Political Incivility on Twitter." *Sage Open* 10 (2): 2158244020919447.

Tucker, Joshua A., Yannis Theocharis, Margaret E. Roberts, and Pablo Barberá. 2017. "From Liberation to Turmoil: Social Media and Democracy." *Journal of Democracy* 28 (4): 46–59.

Van Vliet, Livia, Petter Törnberg, and Justus Uitermark. 2020. "The Twitter Parliamentarian Database: Analyzing Twitter Politics across 26 Countries." *PLoS One* 15 (9): e0237073.

Ward, Stephen, and Liam McLoughlin. 2020. "Turds, Traitors and Tossers: The Abuse of UK MPs via Twitter." *The Journal of Legislative Studies* 26 (1): 47–73.

Maarja Lühiste is a Reader (Associate Professor) in Comparative Politics and Gender at Newcastle University: maarja.luhiste@ncl.ac.uk

Stiene Praet is a Postdoctoral Researcher at the University of Antwerp: stiene.praet@uantwerpen.be

Sebastian Adrian Popa is a Reader (Associate Professor) in Comparative Politics and Public Opinion at Newcastle University: Sebastian.popa@ncl.ac.uk

Yannis Theocharis is a Professor of Digital Governance at the Technical University of Munich: yannis.theocharis@hfp.tum.de

Pablo Barberá is an Adjunct Associate Professor at the University of Southern California and a Research Scientist at Meta: pbarbera@usc.edu

Zoltán Fazekas is an Associate Professor of Business and Politics at the Copenhagen Business School: zf.egb@cbs.dk

Joshua A. Tucker is a Professor of Politics at the New York University: joshua.tucker@nyu.edu

Cite this article: Lühiste, Maarja, Stiene Praet, Sebastian Adrian Popa, Yannis Theocharis, Pablo Barberá, Zoltán Fazekas, and Joshua A. Tucker. 2025. "When Does Fame Not Matter? Examining Gender Differences in Politicians' Social Media Experiences." *Politics & Gender* 1–28. https://doi.org/10.1017/S1743923X25100111