Sentiment on the Campaign Trail: Gender Differences in Candidates’ Use of Emotive Language

Accepted for publication in the Journal of Politics

https://doi.org/10.1086/741451

Sentiment on the Campaign Trail: Gender Differences in Candidates' Use of Emotive Language

Tiffany D. Barnes
Professor, Department of Government, the University of Texas at Austin
tiffanydbarnes@utexas.edu, 0000-0001-6211-3527

Charles Crabtree
Senior Lecturer, School of Social Sciences, Monash University
K-Club Professor, University College, Korea University
charles.crabtree@monash.edu, 0000-0001-5144-8671

Akitaka Matsuo
Lecturer, Department of Government, the University of Essex
a.matsuo@essex.ac.uk, 0000-0002-3323-6330

Yoshikuni Ono
Professor, Faculty of Political Science and Economics, Waseda University
onoy@waseda.jp, 0000-0002-9222-4014

Abstract

How do candidates use emotive language during elections? Whereas government candidates and incumbents are incentivized to use positive language to incite support for the status quo, opposition candidates and challengers can strategically use negative sentiment to foster voter discontent. We argue that women candidates face a double bind, which makes it less likely they benefit from negative language and limits the strategies at their disposal. Leveraging approximately 165,000 tweets from 2,662 UK general election candidates, we show women are more positive and less negative than men, regardless of their government/incumbent status. Subsequent sentiment analysis of over a million replies indicates that women may avoid negativity because they are disproportionately penalized for negative emoting—garnering more negative replies and fewer likes than men. Together, these findings suggest women are not simply socialized to be more positive, but also, they are strategically motivated to behave in gender-typical ways to appeal to voters and avoid backlash on the campaign trail.

Key Words: women candidates; emoting; campaign; sentiment analysis

Supplementary material for this article is available in the appendix in the online edition.

Replication files are available in the *JOP* Dataverse (https://dataverse.harvard.edu/dataverse/jop). The empirical analysis has been successfully replicated by the *JOP* replication analyst.

Women encounter many obstacles on the path to office (Bauer 2020; Bernhard et al. 2021; Cassese and Holman 2018; Saha and Weeks 2022; Teele et al. 2018). To navigate the campaign trail, women candidates behave strategically, innovating to gain an edge (Wagner et al. 2017). For instance, women carefully craft their campaign messages, often emphasizing issues voters associate more favorably with women, to improve their electoral fortunes (Herrnson et al. 2003, Evans 2023). When it comes to campaign messages, *how* politicians use language—not merely the content—matters (Boussalis et al. 2021; Dietrich et al. 2019; Hargrave 2023; Hargrave and Blumenau 2022). Though experimental research has long shown campaign messages can manipulate emotional responses and influence voting behavior (Marcus et al. 2000), observational research has only recently demonstrated that parties (Crabtree et al. 2020) and legislators (Osnabrügge et al. 2021) leverage this strategy—deploying emotive language to sway voters. Despite growing attention to how parties and politicians use emotive language, extensive research on legislative speeches (Bäck and Debus 2019; Bäck et al. 2021; Childs 2004; Slapin et al. 2018; Slapin and Kirkland 2020; Vallejo Vera and Vidal 2022; Weeks 2009; Xydias 2024), and widespread public and scholarly interest in women's pathways to power (Crowder-Meyer 2020; Piscopo and Kenny 2020; Saha and Weeks 2022; Sweet-Cushman 2016), scholars have given far less attention to how individual-level characteristics—such as gender—shape politicians' strategic use of emotive language.

We consider the use of emotive language as an additional dimension of electoral campaigns. This dimension, called campaign sentiment, refers to "the emotive content of campaigns" (Crabtree et al. 2020, 1044).[1] Whereas *campaign content and focus* address what is said and about whom, *campaign sentiment* captures how it is said. In our usage, emotive language refers to a valence-based property of tone (positive – negative) that can be separated from the targets, issues, or rhetorical form (e.g.,

[1] We use 'campaign sentiment,' 'sentiment,' and 'emotive content' interchangeably, in line with the literature.

attack, contrast, critique). This approach differs from much of the negativity literature, which classifies the target and structure of attacks. Accordingly, negative tone need not be aggressive or uncivil, and an attack can be framed without overt negative affect. We thus do not measure aggression, incivility, or any related to the target of sentiment; instead, our focus is on tone.

We build on prior work and theorize that because positive emoting can elicit emotional responses in voters—thereby influencing their behavior at the polls—candidates have incentives to use emotive language strategically. Members of governing parties and incumbents have stronger incentives to employ positive emoting to bolster the status quo, whereas members of opposition parties and challengers are more motivated to use negativity to cultivate discontent. Not all candidates, however, benefit equally from emotive language. Women challengers and opposition members may face a double bind during campaigns (Schneider and Bos 2014), making it less likely that they benefit from negative emoting. Men are stereotyped as tough and aggressive, while women are expected to be compassionate, gentle, and likable (Bauer 2020; Cassese and Holman 2018; Sweet-Cushman 2016). Even when critiques are delivered without incivility, negative emoting can reduce perceived warmth and likability; thus, the same negative emoting that can aid men may trigger backlash for women. Consequently, women in opposition or as challengers, fearing sanction for violating gendered expectations (Krupnikov and Bauer 2014), may be less likely than men to engage in negative emoting—giving them one less tool on the campaign trail.

To test our theoretical argument, we developed a novel dataset capturing the emotive language from approximately 165,000 tweets posted by 2,662 candidates for the House of Commons across the 2017 and 2019 UK snap elections. The messages on these candidate accounts on Twitter (now X) are important. Unlike other types of political messages, the content of these messages is directly attributable to individual candidates, which means that candidates can be held accountable for the content they post; reputational benefits and costs for posted content can be immediate. Since most political candidates in the UK actively used Twitter during the period we study, we have sufficient variation across the candidate dimensions we care about: governing vs. opposition and incumbent vs. challenger status for both women and men. In line with cross-national research and recent studies from the UK, incumbents and governing-party members use more positive sentiment, while challengers and opposition members use more negativity (e.g., Crabtree et al. 2020; Duggan and Milazzo 2023). We build on this growing literature by showing that these patterns also vary by gender: women exhibit more positive and less negative sentiment than men with the same status. We argue below that this empirical pattern occurs because women face stronger negative reactions, a greater sanction, when they use negative sentiment.

Beyond documenting differences in campaign sentiment across candidates by gender, we assess effectiveness. If women emphasize positivity to avoid alienating voters, we should observe backlash from citizens when they violate gendered expectations and emote negatively. An analysis of over one million responses to candidate tweets shows that women are penalized more than men for negativity: negative tweets by women receive more negative replies and fewer likes, despite similar overall engagement. Thus, in line with evidence that women strategically adjust behavior to avoid backlash (Vallejo Vera and Vidal 2022), our findings suggest that women are not simply avoiding negativity because they are socialized to behave more positively, but rather women strategically deploy emotive language to navigate the double bind.

Combined, these results shed light on the complex ways campaigns are gendered and deepen our understanding of how politicians use campaign sentiment to advance their position. They also contribute to the accounts of women exercising agency, behaving "contra expectations of institutions characterized by party politics" (Childs 2023, 521) to improve their political fortunes. Women candidates strategically manage emotional appeals to minimize negative voter responses, yet *reactions* to their messages show the persistent gendered constraints they face. Though going negative remains a widely used campaign tactic—particularly for challengers and opposition (Duggan and Milazzo 2023; Nai 2020; Rossini et al 2024; Walter 2014)—our analyses show that women who adopt this strategy receive more negative responses than men. These disproportionate reprisals narrow women's strategic options and help entrench gender inequalities in electoral outcomes.

Gender and Campaigns

Political campaigns are inherently gendered (Anzia and Bernhard 2022; Cassese and Holman 2018; Ross et al. 2013; Smith 2021). Women are less likely to be recruited (Kenny 2013; Norris and Lovenduski 1995), often express lower political ambition (Allen and Cutts 2018; Crowder-Meyer 2020; Schneider et al. 2016), and run less often (Bernhard et al. 2021; Thomsen and King 2020). They also face gendered evaluations during campaigns (Saha and Weeks 2022). In some cases, women (and men) can benefit from feminine leadership styles (Bernhard 2022). Still, voters largely hold masculine expectations of candidates (Schneider and Bos 2014; Sweet-Cushman 2016), imposing often higher—and sometimes shifting—standards on women (Bauer 2020). As a result, women may be uniquely affected by attack advertising: they can perhaps weather some attacks better than men (Fridkin et al. 2009), but are more vulnerable to advertisements implying norm violations (Cassese and Holman 2018) or portraying them as aggressors (Krupnikov and Bauer 2014).

Given these barriers to evaluation, understanding how women navigate campaigns is crucial (Evans 2023). Evidence from U.S. elections shows that some women gain by running "as women" and emphasizing "women's issues" (Bernhard 2022; Herrnson et al. 2003), while others contend women do well with "masculine" strategies (Windett 2014) and may face backlash when campaigns reinforce stereotypes (Bauer 2015). Gendered messaging is not limited to women: men also deploy masculinity cues in UK campaigns (Smith 2022). Findings on gender and negativity are mixed and largely U.S.-based. More relevant to our current study on campaign sentiment, or emotive content, in the U.S. House context, women sometimes invoke more disgust (Macdonald et al. 2024), use more angry rhetoric (Russell et al. 2023), and tweet more negative (Evans et al. 2017), attack-style messages (Evans and Clark 2016).

How women deliver messages can be critical for navigating gendered dynamics. Scholars have long emphasized the importance of women's communication and speech-making for understanding political success (Bäck et al. 2014; Childs 2004; Costa 2021; Dietrich et al. 2019; Weeks 2009). Studies of the UK House of Commons, in particular, document a more "feminized style of politics," with women described as more measured and less shouty, defensive, or embattled (Childs 2004, 3; 5); kinder, gentler, and more collaborative (Norris 1996, 93); and more willing to listen to the other side (Bochel and Briggs 2000, 66–67; but see Hargrave and Langengen 2021). By contrast, much less is known about delivery during campaigns, and most evidence on women's campaign messaging comes from the U.S. context (Evans and Clark 2016; Macdonald et al. 2024; Russell, Evans, and Gervais 2023). While a growing UK literature tracks negativity at the party and leader levels across social media, online ads, and leaflets (Dommett et al. 2025; Duggan and Milazzo 2023; Rossini et al. 2024; Trumm et al. 2025), evidence at the candidate level—where messages are personally attributable and reputational costs are immediate—remains limited. We contribute by analyzing women's and men's use of emotive language in UK general elections, a setting where negative appeals are often outsourced to party or leader platforms.

Emoting on the Campaign Trail: A Political Strategy

How candidates communicate matters. Experimental research shows campaign sentiment can be manipulated to induce emotions and influence voting (Marcus et al. 2000). Campaigns use images, music, or vocal tone to trigger emotions and capture attention (Brader 2005; Lau and Rovner 2009; Rossini et al. 2024). Recent observational research shows that parties manipulate campaign sentiment (Dommett et al. 2025) through a range of materials including manifestos (Crabtree et al.

2020), and legislators use emotive language in parliamentary speech (Osnabrügge et al. 2021) to appeal to voters.

Electoral incentives influence whether candidates go positive or negative (Dommett et al. 2025; Rossini et al. 2024; Trumm et al. 2025). Generally, advantaged candidates tend to be more positive, as they benefit when voters are satisfied with the status quo. In the UK—a multilevel political system with multiparty parliamentary elections and relatively strong parties—candidates respond to both national and local incentives (Duggan and Milazzo 2023). At the national level, the governing party sets the policy agenda and elections function as a referendum on its performance; governing-party candidates therefore have incentives to emphasize achievements and adopt more positive sentiments than the opposition. At the local level, individual reputations matter: members of parliament (MPs) can propose private members' bills (Bowler 2010; Fleming 2022) and build constituency records (Duggan and Milazzo 2023); thus, elections also serve as a referendum on incumbent MPs (Pattie et al. 2017). In their "localised appeals" targeting specific districts (Dommett et al. 2025, 735), incumbents likewise have incentives to convey positive sentiment. Accordingly, both governing-party candidates and incumbents have strong reasons to use positivity in their campaign messaging to defend the status quo and garner more votes at the polls.

While using negative sentiment risks backlash (Nai and Martínez i Coma 2019; Walter and van der Eijk 2019), negativity is increasingly prevalent in modern UK election campaigns, and most candidates are motivated to employ it at some point (Rossini et al. 2024; Trumm et al. 2025). Opposition members and challengers confront this incentive more often (Duggan and Milazzo 2023; Walter et al. 2014). Since the governing party is responsible for crafting the UK-wide policy agenda, and incumbents control policies targeting their district, opposition members and challengers are incentivized to wield negative sentiment to foster dissatisfaction with the status quo and their opponents (Evans et al. 2017; Rossini et al. 2024). Negative emotional appeals can also heighten voter attention (Marcus et al. 2000) and provoke reevaluation of both candidates' qualifications and track records (Lau and Rovner 2009), prompting voters to consider alternatives (Druckman et al. 2010). Consequently, conventional wisdom suggests that opposition members and challengers are more likely to engage in negative emoting (Duggan and Milazzo 2023; Nai 2020).

Women Candidates and the Strategic Use of Emotional Sentiment

The conventional wisdom on campaign dynamics may be useful for anticipating how men leverage emotive language differently depending on their party and incumbency status. Women, however, confront different dynamics: they face a "double bind," where voters' expectations for how women should behave are incongruent with expectations for political leaders (Eagly and Carli 2007; Schneider and Bos 2014). Whereas men are expected to be strong and aggressive, women are expected to be positive, warm, gentle, and likable (Schneider and Bos 2019). Gender role congruity theory suggests that because politics is dominated by men, leaders are often associated with masculine traits, granting men an electoral edge (Eagly and Karau 2002).

Women politicians cannot simply resolve this incongruity by embodying masculine traits. Although voters tend to value masculinity in leaders (Bauer 2020), they also expect gender-congruent behavior (Eagly and Carli 2007). When women candidates conform to expected gender stereotypes, they are likely to be rewarded. Violating gender stereotypes can elicit negative responses (Schneider and Bos 2014; Vallejo Vera and Vidal 2022). To this point, women candidates are, on average, more likely to be rewarded for positive emotional displays, and punished for negative ones (Boussalis et al. 2021; Cassese and Holman 2018). Evidence also suggests that emoting can make women's messages more effective: women legislators are more likely to sway men to vote in favor of legislation when speaking with greater emotional intensity (Dietrich et al. 2019), and in campaign debates, voters respond more positively to women's positive displays (Boussalis et al. 2021).

Taken together, prior theoretical and empirical contributions in the politics and gender literature then suggest that women may have an incentive to use more positive sentiment than men. For women in the governing party and women incumbents, their incentives largely align: as governing party MPs/incumbents, they are motivated to engage in positive emoting, signaling support for the status quo, and to rely less on negativity. Above and beyond the incentive fostered by their party and incumbency status, as women, they have an incentive to deploy positive sentiment—eschewing negativity—to cultivate a warm, likable persona.

H1a: Among governing-party members and incumbents, women exhibit more positive sentiment in their campaign messages than men.

H1b: Among governing-party members and incumbents, women exhibit less negative sentiment in their campaign messages than men.

For women opposition candidates and challengers, the task of crafting effective messages may be more difficult. Opposition politicians face strong incentives to employ negative rhetoric. Yet, as Crabtree et al. (2020) argue, positivity and negativity are best understood as distinct dimensions of communication styles, rather than opposite ends of a single continuum. Politicians can therefore draw on both strategies at once, though gendered expectations likely shape how each is received, as the literature reviewed above shows. For male challengers and opposition MPs, negative emoting may reinforce stereotypes of strength and assertiveness. For women, however, the same behavior often entails violating gender norms, increasing the risk of voter backlash (Bauer 2017). Even when criticism is expressed without incivility, negative emoting can still diminish perceptions of warmth and likability—qualities more closely expected of women candidates. Indeed, experimental research finds women face harsher backlash for expressing negative emotions (Boussalis et al. 2021) or going negative in campaigns—especially when seen as the instigator (Herrnson et al. 2003; Krupnikov and Bauer 2014). Given the potential consequences of conveying negativity, we anticipate that despite the incentives fostered by their opposition/challenger status, women are likely to be motivated to curb their negativity and incorporate more positive sentiment into their campaign messages. Though women opposition MPs/challengers are likely more negative (and less positive) than women government MPs/incumbents owing to their institutional incentives, they are less likely than their male counterparts with the same institutional status to rely on negativity.

H2a: Among opposition members and challengers, women exhibit at least as much or more positive sentiment in their campaign messages than men.

H2b: Among opposition members and challengers, women exhibit less negative sentiment in their campaign messages than men.

Evaluating the Use of Emotive Language on the Campaign Trail

To evaluate our hypotheses, we developed a dataset capturing emotive language in tweets by candidates from parties competing in the 2017 and 2019 snap elections for the UK House of Commons. The United Kingdom provides a most-likely case setting to examine gender differences in campaign rhetoric. It combines a party-centered, multiparty Westminster-style system with single-member district elections, creating dual (and potentially cross-cutting) pressures for candidates to respond to both national party agendas and local constituency interests (Duggan and Milazzo 2023). Party cohesion is relatively high, with the party setting the national policy agenda (i.e., the UK-wide Westminster agenda in a multi-level polity), engineering a national campaign, and controlling the overarching message (Slapin et al. 2018). Still, voters value a direct link with their MP (Fleming 2022) and MPs work hard to develop a personal vote (Cain et al. 1987). Thus, elections also serve as a check on the incumbent (Pattie et al. 2017), cultivating an incentive for individual candidates to develop personal messages beyond the party line (Duggan and Milazzo 2023).

At the same time, the two contests we study—2017 and 2019—were both snap elections conducted amid intense Brexit-related polarization. These elections were unusual in both timing and tone: they occurred outside the normal electoral cycle, with parties scrambling to reframe long-standing ideological divides around Brexit, leadership competence, and national identity. This volatile environment amplified the reputational stakes for individual candidates, especially women, who faced heightened scrutiny as parties sought to project unity and discipline amid national turmoil. Both elections also marked peaks in social media campaigning, with Twitter emerging as a central venue for signaling partisanship, responsiveness, and professionalism under pressure. Studying these contests therefore provides a critical window into how gendered communication dynamics manifest under conditions of political crisis and strategic uncertainty.

Our focus on social media is also important. In the UK's party-centered and layered campaign ecology, the harshest attacks are often delegated to party, leader, or satellite actors (Dommett et al. 2025; Rossini et al. 2024), whereas constituency candidates remain directly and personally accountable to voters. Precisely because candidate appeals are immediately attributable to individual politicians, candidate Twitter communication is a most-likely arena for the gendered double bind, where reputational constraints are especially salient. Consistent with this view, evidence from candidate-level campaign materials in the UK shows that, when candidates do go negative, they more often target parties and issues than individuals (Trumm et al. 2025), and that their negativity varies with party strategy and constituency context (Duggan and Milazzo 2023). Our candidate focus thus complements party- and leader-level analyses rather than substituting for them.

This distinction is critical because it means both women and men have incentives to cultivate their own campaign strategy. If elections were only a referendum on the national party, candidates would have scant motivation to diverge from the party platform, making it very difficult to detect any gender differences in campaign behavior. But, since all candidates face personal-vote-seeking incentives, if women and men have different incentives to use emotive language, we should be most likely to observe these differences in settings like the House of Commons elections.

Observing Candidates' Campaign Messages: Evidence from Twitter

Twitter is a vital campaign tool in elections worldwide—including UK general elections (Bright et al. 2020; Smith 2021). Scholars increasingly use Twitter to understand politicians' messages both in office (Russell 2021), and during campaigns (Evans 2023; Russell et al. 2024). Candidate accounts offer personally attributable messages and elicit public, real-time reactions (e.g., replies), providing unique leverage for studying gendered reputational constraints. Twitter also offers important methodological advantages. First, it facilitates large-scale, candidate-level observation, filling a noted gap in UK-focused campaign research (Duggan and Milazzo 2023). Even under strong party discipline, candidates strategically deploy emotive rhetoric—and face gender-asymmetric audience responses. Second, Twitter is used to address a broad public (Stier et al. 2018), and research shows that candidates' online behavior closely mirrors traditional campaign styles (Larsson 2016; Lilleker et al. 2011; Stromer-Galley 2000). Thus, findings likely generalize beyond Twitter. Third, researchers can access the full corpus of campaign-period tweets, minimizing selection bias concerns—particularly important if candidate gender shapes message visibility in other contexts, risking post-treatment bias (Montgomery et al. 2018). Fourth, Twitter's free access removed financial barriers, contributing to its near-universal use among candidates in the 2017 and 2019 UK general elections (Bright et al. 2020) and enabling broad, gender-comparable coverage of campaign behavior.

Data Collection & Measuring the Use of Emotive Language

We selected all candidates for the House of Commons from the five parties with the largest vote shares in the 2017 and 2019 elections. Candidates competed across six parties: Conservatives, Labour, Liberal Democrats, Scottish National Party, UK Independence Party, and Brexit Party. We collected this sample of tweets through the "follow" functionality of the Twitter Streaming API, which allows users to supply a list of Twitter accounts and receive a stream of tweets related to these accounts. To maximize coverage, we crowd-sourced data collection and appended additional candidates to Democracy Club's candidate Twitter list, successfully identifying accounts for 70% of

2017 candidates and 87% of 2019 candidates from the included parties. We collected all candidate tweets from two weeks before the election up until a day before the election, for a total of 164,496 tweets posted by 2,662 candidates (75,760 tweets by 1,484 candidates in 2017 and 88,736 tweets by 1,829 in 2019).[2]

We coded tweet-level sentiment using a dictionary-based method that computes the proportion of positive and negative words in each tweet. Words were labeled positive, negative, or neutral with the Harvard-IV sentiment dictionary (*SentimentAnalysis* package in R), which has been repeatedly validated and widely used in social science (e.g., Dietrich et al. 2019). The dictionary contains 2,005 positive and 1,637 negative words. This approach yields two separate, continuous measures—a positive and a negative score—calculated as the proportion of words labeled positive or negative (See Appendix A1). Each ranges from 0 to 1; they are not mutually exclusive because tweets may contain both valences, and because many tokens are neutral, the two proportions do not sum to 1. For illustration, an average-positive tweet is 0.182: "We @ScotTories are the only party that Agriculture & Fisheries Industries can rely on for support through #Brexit. #GE2017." An average negative tweet is 0.091: "Farmers all over are in record debt due to the SNP's mess and mismanagement of CAP and our rural economy." Some tweets show both (0.125 positive; 0.062 negative): "Tories talk about how they're funding NHS but won't tell you this. @UKLabour will give the NHS the money it needs. #VoteNHS #VoteLabour2017." (see Appendix A8 for more examples). For candidate tweets, the average positive (negative) score is 0.18 (0.08) with a standard deviation of 0.15 (0.10).[3] To assess the construct validity of our measure, we replicate the main analyses using a large language model (LLM)-based alternative instrument (see Appendix A4) and

[2] Appendix A2 shows the number of candidates with Twitter accounts (Tables A1-A3) and level of twitter activity (Table A4) across groups. https://democracyclub.org.uk/
[3] Group means by gender, party status, and incumbency status are available in Table A5.

benchmark both scores against independent human-coded tweet sentiment (Appendix A5). Our results are robust to the alternative LLM classification approach.[4]

Independent Variables of Interest

To test our hypotheses, we created a binary indicator for whether a candidate self-identified as a woman (*Woman*) using the official election data from the House of Commons Library. Women represent 35.4 percent of Twitter-using candidates across the two elections. We also created two binaries: governing-party membership (*Governing Party*; 1 = governing, 0 = opposition) and incumbency (*Incumbent MP*; 1 = incumbent, 0 = challenger). Because we examine only two general elections, party effects cannot be fully separated from governing-party status; we therefore report models with party fixed effects in Appendix 7.[5] We include interactions between woman and governing-party status/incumbency, to evaluate our hypotheses.

Modeling Strategy

We estimate the relationship between campaign sentiment, incumbency status and gender using ordinary least squares regression models. The unit of observation is an individual tweet. We analyze separate regression models for positive and negative sentiments. The model incorporates fixed effects for election years and place—comprising the nine English regions and the three devolved

[4] Specifically, the main effects are replicated in both direction and substantive magnitude using the LLM-based approach, though interaction terms involving Woman attenuate and, in some specifications, do not reach conventional levels of statistical significance (see Appendix A4, Tables A10–A11, and Figures A1–A2 for details).

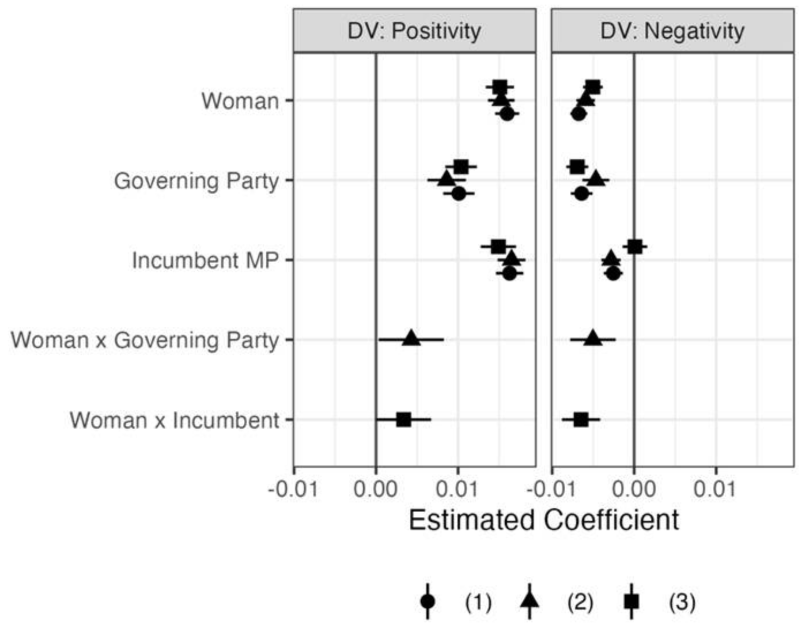
[5] Although we capture variation in incumbents' party affiliation, governing-party status is closely tied to party (Conservative Party candidates are coded as "members of the governing party) as we examine only the 2017 and 2019 UK general elections. This is a limitation of our sample period; however, we address inter-party dynamics by adding party fixed effects to the model as a robustness check in Appendix A7. Our inferences about women's more positive (less negative) tone and the steeper backlash to women's negativity remain unchanged.

nations (Scotland, Wales, and Northern Ireland). To address possible heteroscedasticity, we cluster standard errors by candidate (i.e., level of the treatment variable *Woman*).[6]

Analyzing the Use of Emotive Language on Twitter During Campaigns

Figure 1 shows the estimated coefficients (points) and 95% confidence intervals (bars) for key variables (Appendix Table A7 contains full model results and estimates for controls). The left (right) panel analyzes the use of positive (negative) sentiment. First, we observe that, on average, women use more positive and less negative emoting on Twitter than men. Second, consistent with expectations from previous literature, the left panel in Figure 1 displays a positive and significant relationship for members of the governing party and incumbents, indicating they engage in more positive emoting during campaigns than opposition MPs or challengers. In contrast, we observe opposition members and challengers use more negative sentiment. These incumbency-based results closely mirror prior findings (Crabtree et al. 2020) and fit well with conventional wisdom. They also provide strong face validity to our use of Twitter data for examining campaign sentiment.

[6] Results are unchanged when controlling for candidate age (available only for a subset of candidates) as a robustness check on potential candidate-level confounding; see Appendix A7.1.

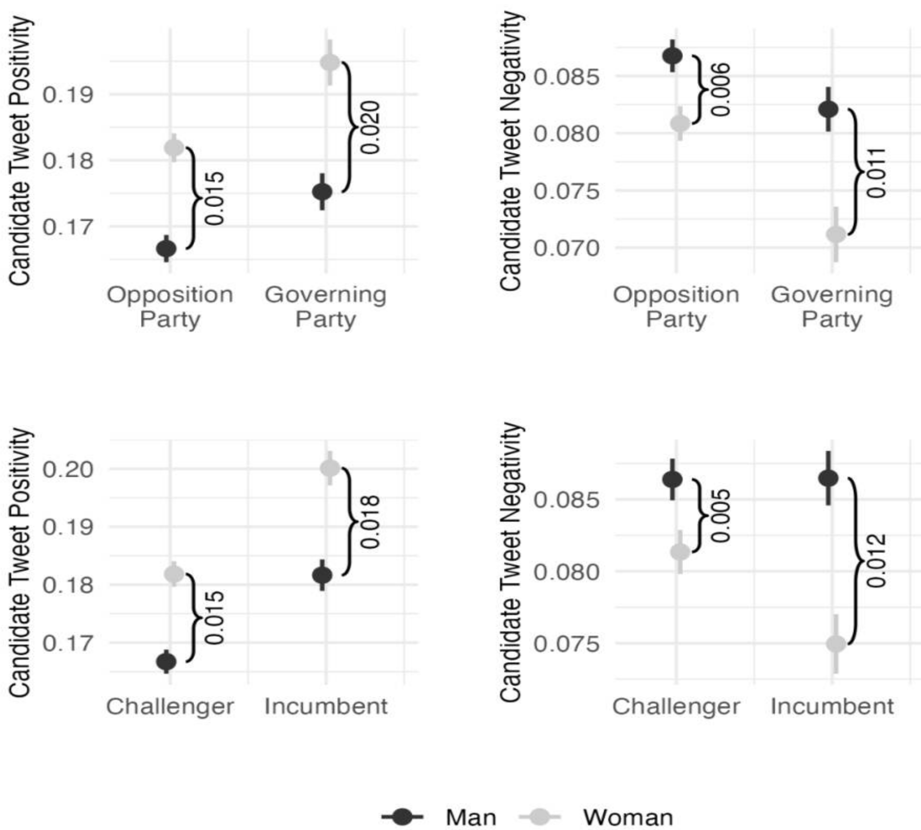


Note: The dependent variable in the left (right) panel is the proportion of positive (negative) tokens in candidate tweets (continuous values ranging from 0 to 1). Model 1(circles) excludes interactions, Model 2 (triangles) includes the Governing party and Woman interaction, and Model 3 (squares) includes the Incumbent and Woman interaction. The unit of analysis is the tweet.

Figure 1: Regression Coefficients for the Candidate-Tweet Sentiment Models

Alt Text: Figure 1 shows that women candidates use more positive and less negative sentiment in campaign tweets than men. Candidates from the governing party and incumbents also use more positive and less negative sentiment than opposition candidates and challengers. These patterns are similar across the three model specifications.

Our primary interest lies in how women navigate campaigns differently than men within the same incumbency/government status. To assess H1a/b and H2a/b, we turn to models including an interaction term between *Women* and *Governing Party* (triangles in Figure 1); and models including an interaction between *Women* and *Incumbent* (squares in Figure 1). The coefficient plot shows the relationship between women and government status is a significant correlate of both positive and negative emoting. And the relationship between women and incumbency status is a strong predictor of negative emoting. To illustrate these relationships and facilitate the interpretation of the interaction terms, Figure 2 plots the expected levels of positivity (left panel) and negativity (right panel) across different situations. The top panel focuses on the governing/opposition status of the candidates (based on Model 2 for both the positivity and negativity DVs) and the bottom panel displays candidates by their incumbency/challenger status (based on Model 3). Dark circles represent the expected values for men, while light circles display those for women.



Note: Points show predicted proportions of positive (left) and negative (right) tokens in candidate tweets (continuous values ranging from 0 to 1) by candidate type (top) and party status (bottom), for women (gray) and men (black). Top-row predictions use Models 2 and 5; bottom-row predictions use Models 3 and 6. Bars denote 95% confidence intervals.

Figure 2: Predicted Positivity and Negativity in Candidate Tweets

Alt Text: Figure 2 shows predicted levels of positive and negative sentiment in candidate tweets by gender and candidate status. Across all subgroups, women candidates use more positive and less negative sentiment than men. The gender gap is largest among governing-party candidates and incumbents. Similar patterns appear in both the governing–opposition and incumbent–challenger comparisons.

Looking first at the differences between women and men, the effect of gender is consistently significant across all sub-groups. That is, women are *consistently* more positive and less negative than men within the same governing-party/incumbency status (H1a, H1b) and within the same opposition party/challenger status (H2a, H2b). Importantly, when comparing across the four panels, the expected proportions of positivity and negativity for women and men candidates vary across the different dyads. Notably, we observe a consistent pattern for both women from the governing party and women incumbents wherein they use higher levels of positive sentiment than men. In fact, although the differences in positivity between men and women are always significant, the largest difference observed is among governing-party candidates, where women's proportion of positivity is about 0.020 higher than men's—roughly two additional positive tokens per 100 tokens. Because campaigns are streams rather than snapshots, this cumulates to about 12 additional positive words over a typical campaign (50 tweets, 595 tokens).

A similar gap is observed between women and men incumbents (0.018), which means about 1.8 more positive tokens per 100 tokens. This pattern provides additional support for our hypotheses: that women members of the governing party and women incumbents use higher levels of positive sentiment in their campaign messages than men. It is worth noting that women challengers are about as positive as men incumbents at the positivity level of 0.18. These trends further illustrate how women alter their behavior to be more positive, even when their challenger status would suggest they have fewer incentives to do so (consistent with H2a).

Next, we turn to the results for negative sentiment displayed in the right panel of Figure 2. First, we observe members of the opposition engage in more negative sentiment than members of the governing party. But there is an important variation by candidate gender. The level of negative sentiment deployed by men from the opposition party is 0.087 compared to 0.081 for women from the opposition party. Women's level of negative sentiment more closely resembles men's from the governing party. Still, women from the opposition party display more negative sentiment than women from the governing party. Thus, even though all opposition members have an incentive to engage in negative emoting, consistent with H2b, we observe women doing so less than men.

Finally, turning to the lower right panel, we examine the difference between challengers and incumbents. Although on average, challengers are more negative than incumbents, there is no significant challenger-incumbent gap for men. In other words, men incumbents use negative sentiment just as much as men challengers, indicating support for the idea that all candidates can benefit from negativity (Lau and Rovner 2009). Nonetheless, women do not engage in negative emoting to the same extent. We find strong support for H2b: women challengers exhibit less negative sentiment than men challengers. In fact, women challengers even express less negative sentiment than men incumbents, suggesting gendered campaign dynamics exert more influence on women's behavior than their challenger status.

Overall, the results show women are markedly less negative than men, suggesting women confront a different set of campaign incentives. The gender gap is larger between men and women from the governing party and incumbents—those candidates who have weaker incentives to convey negative messages—than members from the opposition party and challengers. Still, the gender gap persists among opposition MPs and challengers—indicating that even when women's opposition/challenger status should compel them to use negative sentiment, they are less likely than men to do so. We contend that women behave this way because they face a double bind and are more susceptible than men to adverse reactions when they engage in negative emoting.

Reactions to Men's and Women's Campaign Messages

We argue women have a strategic incentive to behave in gender-typical ways to shield themselves from backlash on the campaign trail. Our primary analysis shows that women are more likely than men to deploy positive sentiment and to avoid negative sentiment. On the one hand, our findings align with our argument that women politicians are aware they face competing expectations from voters—i.e., that politicians should exhibit masculine leadership traits, but women should display feminine, gender-typical traits (Bauer 2020; Schneider and Bos 2019). On the other hand, our results are observationally equivalent with the idea that women simply behave in more gender-typical ways because they are socialized to do so and not because they face different standards.

Our findings thus raise a second question: *Do women face adverse responses when they deploy negative sentiment?* If women's negativity is met with similar reactions to men's, women may be unnecessarily avoiding negative sentiment—a proven campaign strategy (Lau and Rovner 2009)—and their behavior may be better explained by gender socialization. Yet, if women encounter disproportionately negative reactions, their behavior may be a strategic attempt to traverse the double bind imposed on women candidates (Schneider and Bos 2014), and that women ultimately have one less tool in their campaign toolkit than men. Despite the importance of citizens' reactions to women's and men's use of emotional sentiment in their campaign messaging, we know very little about how citizens evaluate/react differently to men and women candidates in real world settings. We derive two competing expectations from existing research.

On the one hand, we may not expect women to elicit more negative responses from voters than men. It is possible that women only avoid negativity because they are socialized to do so. Prior work questions the assumption that voters respond differently to women candidates than to men (Brooks 2013). This literature also points to the conventional wisdom that when women run, they win, suggesting that women's underrepresentation in office owes more to lower levels of ambition than to discrimination (Bernhard, Shames, and Teele 2021; Lawless and Fox 2005; Schneider et al. 2016). Meta-analysis finds a small *bonus* for women candidates when reviewing experimental research comparing voter evaluations of identically-profiled men and women candidates (Schwarz and Coppock 2022). Evidence from survey experiments finds that voters from the UK evaluate men and women politicians similarly when they adopt a range of argument styles, not penalizing women for displaying aggression (Hargrave 2023). Similarly, survey experiments from the UK find little evidence that voters punish ambitious (Saha and Weeks 2022), corrupt (Eggers, Vivyan, Wagner 2018), or unproductive (Hargrave and Smith 2024) women politicians more than men. Evidence from women's participation during parliamentary debates shows their speaking styles are like men's. Hargrave and Blumenau (2022) suggest these similarities indicate women face less pressure from elites to conform to stereotypically 'feminine' communication styles. Together, these findings suggest that women candidates are not evaluated differently than men. If anything, voters may prefer women candidates. These findings thus cast doubt on the idea that women who deploy negative sentiment will be met with different reactions than men.

On the other hand, there is reason to believe women may be evaluated differently than men. Although voters may prefer women in some settings, the opposite may be true when women break with gender stereotypic behavior (Barnes, Beaulieu, and Saxton 2020; Bauer 2020; Ono and Yamada 2020). Research on constituent communication indicates women legislators tend to be evaluated more positively (negatively) than men when they provide high (low)-quality responses to constituents; evaluations of men, by contrast, are less sensitive to the quality of their responses (Costa 2021). Survey experiments in the UK also find women are viewed as less experienced (but more approachable) than identical men (Campbell and Cowley 2014). Other work shows women "face a disproportionate punishment for relying on negativity" during campaigns—but only when they are viewed as the instigator (Krupnikov and Bauer 2014, 167). Combined, this work suggests women are held to different standards—rewarded for conforming to gender stereotypes and punished for violating them. This body of research leads us to posit women candidates will elicit more negative emotions from voters when they use negative sentiment and receive more positive emotions when they emote positively.

To assess these competing expectations, and better understand whether women have a strategic incentive to adopt more positive and less negative sentiment to avoid reprisal from constituents—or whether they simply use more positive and less negative sentiment than men due to socialization—we hypothesize:

> *H3: Campaign messages containing positive/negative sentiment from women candidates will elicit more positive/negative responses, than similar messages from men.*

Evaluating Reactions to Women's and Men's Campaign Messaging

To evaluate reactions to women and men candidates, we examine Twitter replies to the candidates' tweets used in our prior analysis. Employing observational data from responses to real candidates in the run-up to elections is an important advance: although there is copious research on how subjects respond to experimental manipulations involving the gender of hypothetical candidates, with very few exceptions (Boussalis et al. 2021) we know little about how voters react to positive and negative sentiment from real candidates. This is an important distinction because experimental research evaluates subjects' responses to identical men and women candidate profiles in hypothetical settings; it does not account for the fact that, in practice, men and women candidates are not identical and that voters may prefer traditional [men] profiles (Teele, Kalla, and Rosenbluth 2018). Analyzing Twitter replies thus offers unique insights into citizens' responses to actual campaign sentiment. We start with the dataset from our prior analysis and build a second dataset consisting of responses to candidate tweets, collected via the Twitter Streaming API's "follow" functionality, which allows users to supply a list of Twitter user IDs and receive a stream of tweets related to these accounts—including tweets originating from them and associated interactions (replies, likes, and retweets).[7]

[7] Research shows candidates use Twitter to target a "mass audience" (Stier et al. 2018), but social media platforms such as Twitter and Facebook are not demographically representative of the general population (Mellon and Prosser 2017). Barberá and Rivero (2015) observe that conservative users are more active and participation overrepresents men, urban dwellers, and those with strong

This approach yields 1,282,340 responses to the original set of candidate tweets.[8]

Responses are not equally distributed across the original tweets. Specifically, simple descriptive statistics show negative candidate tweets tend to attract far more attention than positive ones: tweets above the median in positivity garner about 7.8 replies, whereas those above the median in negativity attract about 10.2 replies. These trends are similar for men and women candidates (albeit men's tweets receive slightly more replies). This increased attention to negativity helps explain candidates' incentives to use it in tweets. Consistent with research from Marcus, Neuman, and MacKuen (2000), negative emotional appeals make people more attentive to candidates' tweets. So, the puzzle remains as to why women candidates are more likely to avoid tweeting negatively than their men counterparts, even though negative tweets attract more attention for both genders.

Next, to address this puzzle and test our hypothesis (H3), we analyze the data using the same approach described above to quantify the level of positive and negative sentiment in the replies, using the Harvard IV dictionary. The dependent variables—reply positivity and reply negativity—are tweet-level proportions of tokens labeled positive or negative in replies. On this scale, the average positive (negative) score is 0.13 (0.11) with a standard deviation of 0.14 (0.13). Overall, candidate tweets are more positive and less negative than the replies they receive. We estimate a series of regression models that take the measure of positive and negative sentiment of replies as their outcomes. The unit of observation is a reply to a candidate tweet. There are two main independent variables regarding the original tweets they were replying to. First is candidate gender: *Woman* equals 1 (0) for women (men) candidates. Second is the sentiment expressed in the original tweet: *Candidate*

---

ideological preferences. We should exercise caution when generalizing results beyond social media, though.

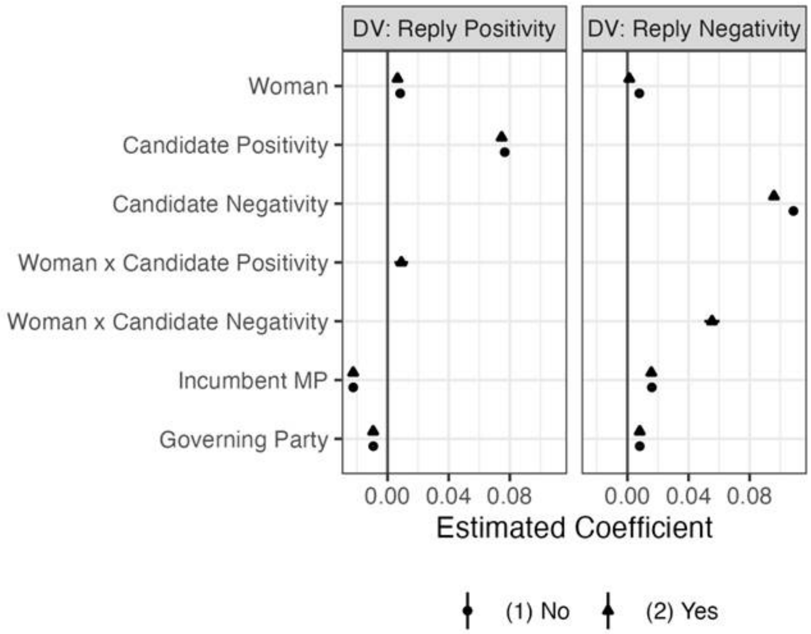
[8] A potential concern is whether automation—such as bots—drives reply patterns. To address this, we link replies to the Botometer-annotated Brexit corpus (Calisir and Brambilla, 2016-2019) and re-estimate our models excluding suspected bots. Our results are substantively unchanged (see Appendix A6; Table A14; Figure A5).

Positivity (*Candidate Negativity*) measures the magnitude of positive (negative) sentiment in the candidate's original tweet. Positivity and negativity are separate measures, each ranging from 0 (neutral) to 1 (the strongest level of sentiment).

We estimate two sets of models, one evaluating the positivity of replies and the other evaluating the negativity. Within each set, we evaluate the direct effect of the main independent variables—the candidate's gender and the positivity/negativity of the original tweet. Then, we introduce two-way interactions between candidate gender and the positivity/negativity of their original tweet to assess if women and men candidates elicit different responses. The models include the same set of control variables as in our prior analysis. We cluster standard errors at the candidate level—i.e., the level of the treatment assignment (*Woman*) in this research design.

Analysis of Replies to Candidate Tweets

Figure 3 plots the results (Appendix A Table A3). The coefficient on *Candidate Positivity* (left panel), shows that as candidate tweet positivity increases so too does reply-positivity. A similar, albeit slightly stronger relationship is observed between *Candidate Negativity* and the level of *Reply Negativity* (right panel). For one standard deviation increase in the positivity (negativity) of the candidate's tweet, the positivity (negativity) of the corresponding replies increases by 0.012 (0.011).



Note: The dependent variable in the left (right) panel is the proportion of positive (negative) tokens in replies (continuous values ranging from 0 to 1). Model 1 (circles) excludes interactions; Model 2 (triangles) includes interactions.

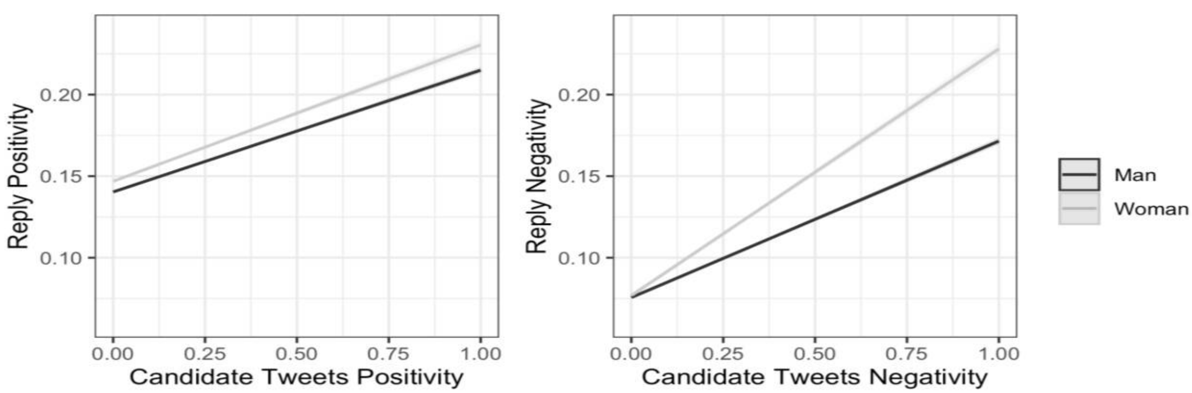
Figure 3: Regression Coefficients for Reply Sentiment Models

Alt Text: Figure 3 shows regression estimates linking candidate tweet sentiment to the sentiment of replies. More positive candidate tweets are associated with more positive replies, while more negative candidate tweets are associated with more negative replies. These relationships appear consistently across both model specifications.

Next, it is clear from the positive and significant coefficient for *Woman*, in the models without interaction terms (circles) that, all else equal, women candidates elicit stronger emotional replies than do men. They receive both more positive and more negative replies than men. Although it is notable that women, regardless of their actions, elicit more emotive replies, we are interested in the interaction between candidate gender and the sentiment of their original tweet for evaluating H3. We thus turn to the interaction between *Woman* and *Candidate Positivity*/*Negativity* plotted with triangles in Figure 3 (Appendix A Table A7 Models 2 and 4).

The interaction effects between *Woman* and *Positivity*/*Negativity* are positive and significant, bolstering our expectation that citizens respond more positively to women when they use positive sentiment, and more negatively when they engage in negative emoting. Still the magnitude of the relationship differs substantially depending on whether women express positive or negative sentiment. The coefficient for negativity (0.055) is over six times larger than for positivity (0.009).

We plot the expected values for reply-tweet positivity (left) and negativity (right) in Figure 4. In the panel on the left, we plot the expected level of *Candidate Positivity* on the x-axis, and positivity in reply to tweets on the y-axis. The results show that the baseline response to women candidates (when *Candidate Positivity* is 0) is more positive than the baseline response to men, indicating that people respond more positively to women candidates than to men. For both genders, as *Candidate Positivity* increases so does citizen positivity, showing that all candidates receive more positive emoting when they use more positive sentiment in their tweets. Importantly, at comparable levels of candidate positivity, replies to women are slightly more positive than replies to men—a small but statistically reliable difference, reflected in the slightly steeper slope for women. These patterns are consistent with our expectations that positive sentiment in women's messaging elicits more positivity than similar messages from men.



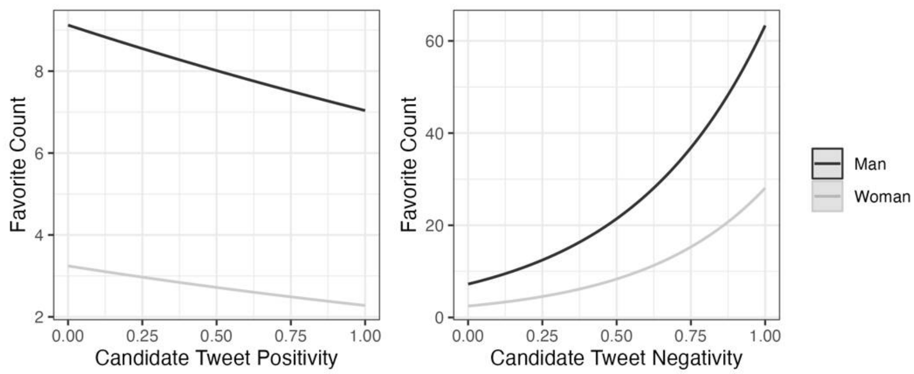
Note: The plot shows predicted proportions of positive (left) and negative (right) tokens in replies to candidate tweets (continuous values ranging from 0 to 1). Predictions use Models 2 and 4 from Figure 3. Shaded areas denote 95% confidence intervals.

Figure 4: Predicted Positivity and Negativity in Replies to Candidate Tweets

Alt Text: Figure 4 shows predicted levels of positivity and negativity in replies as candidate tweet sentiment increases. Replies become more positive as candidate positivity increases and more negative as candidate negativity increases. Across the range of candidate sentiment, replies to women candidates are slightly more positive and more negative than replies to men.

Next, turning to candidate negativity, the panel on the right plots the expected level of *Candidate Negativity* on the x-axis against the expected level of negativity in reply-tweets on the y-axis. The higher intercept for women candidates suggests they elicit more emotive responses regardless of the sentiment level in the original tweet. The positive and significant slope on the line for women and men candidates indicate that increases in negative sentiment from both women and men candidates are met with more negativity in replies. Two trends stand out. First, the relationship between candidate negativity and reply negativity is much steeper (for both genders), indicating that negative sentiment elicits stronger reactions than positive sentiment. Second, although the slope for positivity in the left panel is statistically different for women and men, the difference in the slope is quite modest. When it comes to negativity, however, women face much stronger reactions than men. In fact, the relationship between negative tweets and reply negativity is about 50% steeper for women than for men, indicating that campaign messages containing negative sentiment from women candidates elicit *much more* negative responses than similar messages from men. Taken together, the results presented in Figures 3 and 4 show support for H3: women are rewarded more than men for positive emoting but punished more than men for negative emoting.

It is possible, of course, that reply negativity is merely a product of candidate supporters piling-on and not indicative of citizens being more critical of (or negative toward) the candidate who originated the tweet. To further probe whether negativity in replies reflects backlash against women candidates who break with gender stereotypes (H3), we consider whether citizens are “favoriting”/ “liking” and replying to negative tweets by women and men at similar rates. We assess two sets of models: one uses the number of likes as the dependent variable, and a second takes the number of replies as the dependent variable. We consider the level of positivity/negativity in the original tweet and assess whether tweets are associated with more likes as positivity/negativity increases. We control for the same set of covariates as in the previous analysis, but because we are examining the number of likes/replies, we employ a Poisson count model (Appendix A Table A8).

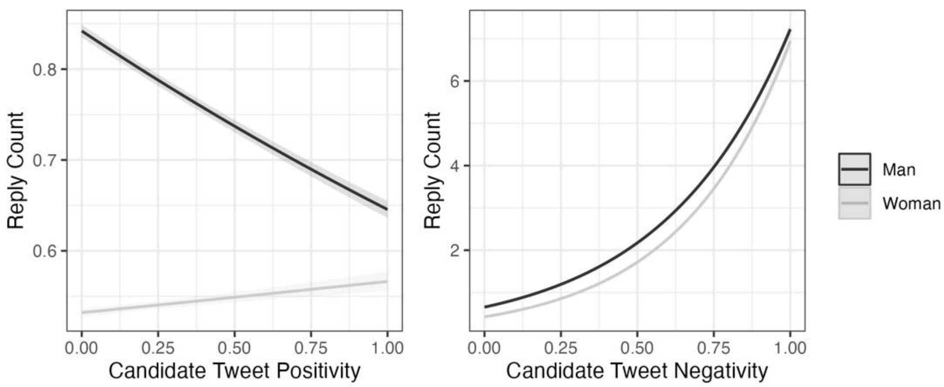


Note: The plot shows the predicted number of likes for candidate tweets. Predictions are generated from the Favorite Count Model in Appendix Table A8. Shaded areas denote 95% confidence intervals.

Figure 5: Predicted Number of Likes to Candidate Tweets

Alt Text: Figure 5 shows predicted numbers of likes for candidate tweets by candidate gender and tweet sentiment. Tweets from women candidates receive fewer likes than those from men across all levels of sentiment. Likes decrease as tweet positivity increases but increase as tweet negativity increases. The gender gap is especially large for negative tweets, which receive substantially more likes when posted by men than by women.

Figure 5 shows how the estimated number of likes given to a candidate's tweet varies by the candidate's gender and the sentiment of the tweet. First, tweets from women, regardless of the level of positivity/negativity, garner fewer likes than those from men. Second, tweet positivity is associated with fewer likes for both men and women—suggesting that positive tweets do not garner as much attention as negative ones. As positivity increases, the number of likes decreases. Third, consistent with previous research that shows negative messages receive more engagement than positive ones (Rossini et al. 2024), as negativity increases so does the number of likes a tweet garners. But, negative tweets by women attract far fewer likes than do negative tweets by men. And, as negativity increases, the disparity in the number of likes between genders grows. Whereas men's tweets see an average of 13.3 likes when negativity is at 95 percentile (Negativity score = 0.28), women's tweets garner about 4.9 likes, indicating that citizens are more likely to support or approve of negative tweets from men than from women. Given that reply sentiment is more negative for women than men, this striking difference suggests that negative replies to women may not be a product of supporters piling-on and instead may reflect genuine disapproval.



Note: The plot shows the predicted number of replies to candidate tweets. Predictions are generated from the Reply Count Model in Appendix Table A8. Shaded areas denote 95% confidence intervals.

Figure 6: Predicted Number of Replies to Candidate Tweets

Alt Text: Figure 6 shows predicted numbers of replies to candidate tweets by gender and tweet sentiment. Replies decrease as tweet positivity increases for men candidates but increase slightly for women candidates. Replies increase as tweet negativity increases for both genders. At low levels of negativity, men's tweets receive more replies than women's, but this gender gap disappears as tweets become more negative.

It is not the case that women's negative tweets receive fewer likes than men's simply because they receive less attention. When examining response rates, we observe most negative tweets from women draw nearly as many responses as the most negative tweets from men. Figure 6 shows how the estimated number of replies to a candidate's tweet varies by candidate gender and the sentiment of the original tweet. Whereas women's positive (and neutral) tweets garner fewer responses than men's, negative tweets from both genders solicit similar levels of responses. When tweet negativity is low there is a discernible gender gap, wherein men's tweets draw more responses. As tweets become more negative, the gender gap in replies disappears entirely. The similar numbers in replies (Figure 6) combined with the higher level of negativity in replies (Figure 4) and smaller number of likes (Figure 5) lend further evidence in support of the idea that negative tweets from women candidates are not as well received as those from men candidates.

Combined, the findings align with our expectation that women behave strategically in campaigns. Specifically, we argued that if women's behavior is strategic, rather than solely stemming from gendered socialization, we should observe backlash when they use negative emoting. Our reply analysis reveals that although negative emoting is generally met with negative responses, women face more negativity than men. Moreover, although the most negative tweets by women garner nearly as many responses as men's, they attract far fewer likes. Together, these patterns suggest that negativity in replies is not simply a product of supporters reinforcing the original tweet tone; rather, it likely reflects—at least in part—negativity aimed at the candidate (or her message). Our results are thus consistent with the idea that women strategically minimize negative sentiment to avoid backlash, and such behavior is not merely a byproduct of socialization.

These findings accord with recent research on women's strategic avoidance of backlash. In Ecuador, for example, Vallejo Vera and Vidal (2022) show that women lawmakers speak less often than men to reduce the risk of being interrupted during floor debates. As they explain, women limit floor time "to avoid the negative effect of interruptions in a context in which their perceived violation of gender roles comes at odds with their work-specific roles" (Vallejo Vera and Vidal 2022, 1385). Our work likewise finds that women limit their use of negative emoting to strategically sidestep potential backlash from the perceived incongruity between women and negative messaging.

Conclusions

Electoral candidates have many tools to navigate campaigns effectively. Campaign sentiment, or the strategic use of emotive language, is an important instrument for mobilizing support and invoking scrutiny, prompting voters to question candidates' positions and qualifications. We ask: Do women and men candidates use these tools differently during campaigns? Using Twitter data from 2,662 candidates across two UK general elections, we show that, despite incentives linked to governing and incumbency status, women deploy positive sentiment more frequently and use negativity less. Moreover, when women use negative sentiment, they are more likely than men to encounter negative reactions. These candidate-level Twitter observations offer a partial view of two Brexit-era contests, yet they contribute to research on how political actors craft messages to induce emotions (Crabtree et al. 2020) and appeal directly to voters (Osnabrügge, Hobolt, and Rodon 2021).

Women's greater positivity and the harsher response to their negativity have important implications for women politicians, campaign strategies, and our understanding of how emotive rhetoric influences voters. When taken at face value, stronger positive reactions to women's positive messaging suggest one path through the double bind—leaning into gender-typical roles. Such a strategy allows them to sidestep reprisal from voters by eschewing negativity. But, upon closer inspection, the double bind also limits women's range of available campaign strategies. Going negative can be a useful approach particularly in tight elections (Druckman et al. 2010; Duggan and Milazzo 2023); constrained access to this strategy may disproportionately disadvantage women. More generally, our findings contribute to our understanding of how women experience politics (Childs 2004), showing they do not benefit from the same campaign strategies as men.

Importantly, this limitation constrains all women candidates. Although opposition members and challengers may be best situated to benefit from negative emoting, there are instances where even incumbent candidates have an incentive to use negative sentiment. Negativity is an invaluable tool for disrupting the status quo and prompting voter scrutiny. As a matter of fact, our analysis shows male incumbents are about as likely as male challengers to use negativity, and governing-party men do so as frequently as opposition women. If women cannot effectively harness negativity, then women—including incumbents and governing-party members—have fewer tools for navigating elections. This challenge represents the classic double bind (Schneider and Bos 2014): negativity can help the average *male* candidate achieve several goals, but women are disproportionately punished for deploying it.

Our findings also bear on parliamentary deliberation and accountability. Parliament is the primary institution monitoring the government, and MPs are tasked with scrutiny and challenge. Yet when women engage in negative rhetoric, they are evaluated more harshly than men. If this dynamic plays out when women criticize the government, it may deter accountability among women MPs for fear of electoral costs; for those undeterred, it may disproportionately threaten their electoral fortunes. Future research should examine whether voters differentially evaluate men and women MPs who publicly scrutinize the government (e.g., at Oral Questions, on social media, or in news coverage) to clarify how gendered evaluations shape women's efficacy in office. Moreover, to understand how women adapt to different settings, it should also assess whether women MPs vary their rhetoric by visibility (e.g., high-profile debates vs. social media comments).

Finally, our results illuminate when individual candidates use negative campaigning. As Duggan and Milazzo (2023) note, the scarcity of large-scale, candidate-level data on campaign communication has steered UK research toward parties rather than individual candidates; echoing Rossini et al. (2024, 450), who observe "little research outside of the US that has investigated the use of communication strategies on social media," our candidate-level Twitter evidence shows that even in a party-centered, multiparty parliamentary setting with high party discipline—such as the UK—candidates do strategically deploy negative sentiment, and audience sanctions are uneven by gender. This pattern refines expectations derived from US presidential, two-party contexts. Our findings also suggest that the tone of political discourse may depend less on who participates and more on what voters reward; in this sense, the potential for changing campaign norms may rest as much on voter behavior as on descriptive representation.

Acknowledgements: The authors would like to thank Bethany Albertson, Susan Banducci, Matt Golder, Sona Golder, Thomas Gschwend, and Andrei Zhirnov for valuable comments. We are thankful to Maria E Navar for excellent research assistance. Earlier versions of this manuscript were presented at the 12th Annual Conference of the European Political Science Association, the Asian Polmeth VIII meeting, the 2023 Society for Affective Science Preconference, the Exeter QStep/NCRM Summer Research Methods Seminar Series, and seminars at the Research Institute of Economy, Trade and Industry (RIETI), the University of Exeter, the University of Houston, and the University of Arizona. The authors are also grateful to seminar and conference participants for their valuable comments. Crabtree would like to thank the Maureen and Mike Mansfield Foundation and Japan Foundation for supporting important research on this project via the U.S.-Japan Network for the Future. This research was conducted as part of the project "Advanced Technology and Democracy: Does new technology help or hurt democracy?" at RIETI.

Funding: Support for this research was provided by the Japan Society for the Promotion of Science (KAKENHI Grants-in-Aid for Scientific Research: 17K03523, 19H00584, 20H00059, 23KK0023, 24K00217, 25K00604), Waseda University (Grant for Special Research Projects: 2024C-005, 2025C-288).

Transparency Statement: During the preparation of this manuscript, the authors used OpenAI for additional sentiment coding to assess the robustness of our main findings shown in the main text; those results are reported in the online appendix. The authors also used ChatGPT and Grammarly for proofreading and language-editing. After using these tools, the authors reviewed and revised the text as needed and take full responsibility for the content of the manuscript.

References

Allen, Peter, and David Cutts. 2018. "An analysis of political ambition in Britain." *The Political Quarterly* 89(1):73-81.

Anzia, Sarah, and Rachel Bernhard. 2022. "Gender Stereotyping and the Electoral Success of Women Candidates." *British Journal of Political Science* 52(4): 1544-1563.

Bäck, Hanna, Marc Debus, and Jochen Müller. 2014. "Who Takes the Parliamentary Floor? The Role of Gender in Speech-making in the Swedish Riksdag." *Political Research Quarterly* 67(3): 504–518.

Bäck, Hanna, Marc Debus, and Jorge M. Fernandes, eds. 2021. *The Politics of Legislative Debates*. Oxford: Oxford University Press.

Bäck, Hanna, and Marc Debus. 2019. "When do women speak?" *Political Studies* 67(3):576-596.

Barberá, Pablo, and Gonzalo Rivero. 2015. "Understanding the political representativeness of Twitter users." *Social Science Computer Review* 33(6): 712–729.

Barnes, Tiffany, Emily Beaulieu, and Gregory Saxton. 2020. "Sex and corruption: How sexism shapes voters' responses to scandal." *Politics, Groups, and Identities* 8(1): 103-121.

Bauer, Nichole. 2017. "The effects of counterstereotypic gender strategies on candidate evaluations." *Political Psychology* 38(2):279-295.

Bauer, Nichole. 2020. *The qualifications gap*. Cambridge University Press.

Bauer, Nichole. 2015. "Emotional, Sensitive, and Unfit for Office?" *Political Psychology* 36(6):691–708.

Bernhard, Rachel, Shauna Shames, and Dawn Teele. 2021. "To emerge? Breadwinning, motherhood, and women's decisions to run for office." *American Political Science Review* 115(2):379-394.

Bernhard, Rachel. 2022. "Wearing the Pants (suit)?" *Politics & Gender* 18(2):513-545.

Bochel, Catherine, and Jacqui Briggs. 2000. "Do women make a difference?" *Politics* 20(2):63-68.

Boussalis, Constantine, Travis Coan, Mirya Holman, and Stefan Müller. 2021. "Gender, candidate emotional expression, and voter reactions during televised debates." *American Political Science Review* 115(4):1242-1257.

Bowler, Shaun. 2010. "Private members' bills in the UK parliament." *The Journal of Legislative Studies* 16(4):476-494.

Brader, Ted. 2005. "Striking a Responsive Chord: How Political Ads Motivate and Persuade Voters by Appealing to Emotions." *American Journal of Political Science* 49(2): 388–405.

Bright, Jonathan, Scott Hale, Bharath Ganesh, Andrew Bulovsky, Helen Margetts, and Phil Howard. 2020. "Does Campaigning on Social Media Make a Difference? Evidence From Candidate Use of Twitter During the 2015 and 2017 U.K. Elections." *Communication Research* 47(7): 988–1009.

Brooks, Deborah. 2013. *He Runs, She Runs: Why Gender Stereotypes Do Not Harm Women Candidates*. Princeton University Press.

Cain, Bruce, John Ferejohn, and Morris Fiorina. 1987. *The Personal Vote: Constituency Service and Electoral Independence*. Harvard University Press.

Campbell, Rosie, and Philip Cowley. 2014. "What voters want: Reactions to candidate characteristics in a survey experiment." *Political Studies* 62(4): 745-765.

Cassese, Erin, and Mirya Holman. 2018. "Party and gender stereotypes in campaign attacks." *Political Behavior* 40(3): 785-807.

Childs, Sarah. 2004. "A feminised style of politics? Women MPs in the House of Commons." *British Journal of Politics and International Relations*, 6(1):3-19.

Childs, Sarah. 2023. "Feminist institutional change: the case of the UK Women and Equalities Committee." *Parliamentary Affairs* 76(3):507-531.

Costa, Mia. 2021. "He said, she said." *Politics & Gender* 17(4):528-551.

Crabtree, Charles, Matt Golder, Thomas Gschwend, and Indriđi Indriđason. 2020. "It is not only what you say, it is also how you say it." *Journal of Politics* 82(3):1044-1060.

Crowder-Meyer, Melody. 2020. "Baker, bus driver, babysitter, candidate? Revealing the gendered development of political ambition among ordinary Americans." *Political Behavior* 42(2): 359–384.

Dietrich, Bryce, Matthew Hayes, and Diana O'Brien. 2019. "Pitch perfect: Vocal pitch and the emotional intensity of congressional speech." *American Political Science Review* 113(4):941-962.

Dommett, Katharine, Samuel A. Mensah, Junyan Zhu, and Tom Stafford. 2025. "Understanding the communicative strategies used in online political advertising and how the public views them." *The British Journal of Politics and International Relations* 27(3): 720-752.

Druckman, James, Martin Kifer, and Michael Parkin. 2010. "Timeless strategy meets new medium." *Political Communication* 27(1):88-103.

Duggan, Alan, and Caitlin Milazzo. 2023. "Going on the offensive: Negative messaging in British general elections." *Electoral Studies* 83: 102600.

Eagly, Alice H., and Linda L. Carli. 2007. *Through the Labyrinth: The Truth About How Women Become Leaders*. Boston, MA: Harvard Business School Press.

Eagly, Alice, and Steven Karau. 2002. "Role congruity theory of prejudice toward female leaders." *Psychological Review* 109(2):573.

Eggers, Andrew C., Nick Vivyan, and Markus Wagner. 2018. "Corruption, accountability, and gender: Do female politicians face higher standards in public life?" *Journal of Politics* 80(1): 321-326.

Evans, Heather. 2023. "Marketing Female Candidates as "Women." In *Political Marketing and the Election of 2020*, edited by Jody C Baumgartner, and Bruce I. Newman, pp. 37-48. Routledge.

Evans, Heather K., and Jennifer Hayes Clark. 2016. "'You tweet like a girl!' How female candidates campaign on Twitter." *American Politics Research* 44 (2): 326-352.

Evans, Heather K., Sean Smith, Alexis Gonzales, and Kayla Strouse. 2017. "Mudslinging on Twitter During the 2014 Election." *Social Media + Society* 3 (2):1–9.

Fleming, Thomas. 2022. "Partisanship and the Effectiveness of Personal Vote Seeking." *Legislative Studies Quarterly* 47(3): 505-538.

Fridkin, Kim L., Patrick J. Kenney, and Gina Serignese Woodall. 2009. "Bad for Men, Better for Women: The Impact of Stereotypes During Negative Campaigns." *Political Behavior* 31(1): 53–77.

Hargrave, Lotte. 2023. "A Double Standard? Gender Bias in Voters' Perceptions of Political Arguments." *British Journal of Political Science* 53(2): 327-345.

Hargrave, Lotte, and Jack Blumenau. 2022. "No Longer Conforming to Stereotypes? Gender, Political Style and Parliamentary Debate in the UK." *British Journal of Political Science* 52:1584-1601.

Hargrave, Lotte, and Tone Langengen. 2021. "The gendered debate: Do men and women communicate differently in the House of Commons?." *Politics & Gender* 17(4):580-606.

Hargrave, Lotte, and Jessica C. Smith. 2024. "Working hard or hardly working? Gender and voter evaluations of legislator productivity." *Political Behavior* 46: 909-930.

Herrnson, Paul, Celeste Lay, and Atiya Stokes. 2003. "Women running "as women": Candidate gender, campaign issues, and voter-targeting strategies." *Journal of Politics* 65(1):244-255.

Kenny, Meryl. 2013. *Gender and political recruitment: Theorizing institutional change*. Springer.

Krupnikov, Yanna, and Nichole Bauer. 2014. "The relationship between campaign negativity, gender and campaign context." *Political Behavior* 36(1):167-188.

Larsson, Anders Olof. 2016. "Online, all the time? A quantitative assessment of the permanent campaign on Facebook." *New Media & Society* 18(2): 274–292.

Lau, Richard R., and Ivy Brown Rovner. 2009. "Negative Campaigning." *Annual Review of Political Science* 12: 285–306.

Lawless, Jennifer, and Richard Fox. 2005. *It takes a candidate: Why women don't run for office*. Cambridge: Cambridge University Press.

Lilleker, Darren G., Karolina Koc-Michalska, Eva J. Schweitzer, Mariusz Jacunski, Nigel Jackson, and Thierry Vedel. 2011. "Informing, engaging, mobilizing or interacting: Searching for a European model of web campaigning." *European Journal of Communication* 26(3): 195–213.

Macdonald, Maggie, Whitney Hua, and Annelise Russell. 2024. "Constrained Communication and Negativity Bias: Gendered Emotional Appeals on Facebook." *Journal of Women, Politics & Policy* 45(2): 261-274.

Marcus, George, Russell Neuman, and Michael MacKuen. 2000. *Affective Intelligence and Political Judgment*. University of Chicago Press.

Mellon, Jonathan, and Christopher Prosser. 2017. "Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users." *Research & Politics* 4(3).

Montgomery, Jacob M., Brendan Nyhan, and Michelle Torres. 2018. "How Conditioning on Posttreatment Variables Can Ruin Your Experiment and What to Do about It." *American Journal of Political Science* 62(3): 760–775.

Nai, Alessandro. 2020. "Going Negative, Worldwide: Towards a General Understanding of Determinants and Targets of Negative Campaigning." *Government and Opposition* 55(3): 430–455.

Nai, Alessandro, and Ferran Martínez i Coma. 2019. "Losing in the Polls, Time Pressure, and the Decision to Go Negative in Referendum Campaigns." *Politics and Governance* 7(2): 278–296.

Norris, Pippa. 1996. "Women politicians: transforming Westminster?" In Jonie Lovenduski and Pippa Norris (eds), *Women in Politics*. Oxford University Press, p.91–104.

Norris, Pippa, and Joni Lovenduski. 1995. *Political Recruitment: Gender, race and class in the British Parliament.* Cambridge University Press.

Ono, Yoshikuni, and Masahiro Yamada. 2020. "Do voters prefer gender stereotypic candidates?" *Political Science Research and Methods* 8(3):477-492.

Osnabrügge, Moritz, Sara Hobolt, and Toni Rodon. 2021. "Playing to the gallery: Emotive rhetoric in parliaments." *American Political Science Review* 115(3):885-899.

Pattie, Charles, Todd Hartman, and Ron Johnston. 2017. "Incumbent parties, incumbent MPs and the effectiveness of constituency campaigns: Evidence from the 2015 UK general election." *British Journal of Politics and International Relations* 19(4):824-841.

Piscopo, Jennifer M., and Meryl Kenny. 2020. "Rethinking the ambition gap: gender and candidate emergence in comparative perspective." *European Journal of Politics and Gender* 3 (1): 3-10.

Rice, Douglas R., and Christopher Zorn. 2021. "Corpus-based dictionaries for sentiment analysis of specialized vocabularies." *Political Science Research and Methods* 9(1): 20-35.

Ross, Karen, Elizabeth Evans, Lisa Harrison, Mary Shears, and Khursheed Wadia. 2013. "The gender of news and news of gender." *The International Journal of Press/Politics* 18(1): 3-20.

Rossini, Patrícia, Rosalynd Southern, Emily Harmer, and Jennifer Stromer-Galley. 2024. "Unleash Britain's Potential (To Go Negative): Campaign Negativity in the 2017 and 2019 UK General Elections on Facebook." *Political Studies Review* 22(3): 449-470.

Russell, Annelise. 2021. *Tweeting is leading: how senators communicate and represent in the age of Twitter*. Oxford: Oxford University Press.

Russell, Annelise, Heather Evans, and Bryan Gervais. 2023. "Good and pissed: gendered emotional appeals on Twitter." *Journal of Information Technology & Politics* 20 (4): 422-436.

Russell, Annelise, Heather K. Evans, and Bryan Gervais. 2024. "Not ready to make nice: Congressional candidates' emotional appeals on Twitter." *Social Science Quarterly* 105(5):1848-1856.

Schneider, Monica, and Angela Bos. 2014. "Measuring stereotypes of women politicians." *Political Psychology* 35(2):245-266.

Schneider, Monica, and Angela Bos. 2019. "The application of social role theory to the study of gender in politics." *Political Psychology* 40:173-213.

Schneider, Monica, Mirya Holman, Amanda Diekman, and Thomas McAndrew. 2016. "Power, conflict, and community: How gendered views of political power influence women's political ambition." *Political Psychology* 37(4): 515-531.

Schwarz, Susanne, and Alexander Coppock. 2022. "What Have We Learned about Gender from Candidate Choice Experiments?" *The Journal of Politics* 84(2):655-668.

Saha, Sparsha, and Ana Weeks. 2022. "Ambitious women: Gender and voter perceptions of candidate ambition." *Political Behavior* 44(2):779-805.

Slapin, Jonathan B., Justin H. Kirkland, Joseph A. Lazzaro, Patrick A. Leslie, and Tom O'Grady. 2018. "Ideology, grandstanding, and strategic party disloyalty in the British Parliament." *American Political Science Review* 112(1):15-30.

Slapin, Jonathan, and Justin Kirkland. 2020. "The sound of rebellion: Voting dissent and legislative speech in the UK House of Commons." *Legislative Studies Quarterly* 45(2):153-176.

Smith, Jessica. 2021. "Bulldozing Brexit." *Journal of Elections, Public Opinion and Parties* 31(4):450-469.

Smith, Jessica. 2022. "Masculinity and femininity in media representations of party leadership candidates: men 'play the gender card' too." *British Politics* 17:408-429.

Stier, Sebastian, Arnim Bleier, Haiko Lietz, and Markus Strohmaier. 2018. "Election campaigning on social media: Politicians, audiences, and the mediation of political communication on Facebook and Twitter." *Political Communication* 35(1):50-74.

Stromer-Galley, Jennifer. 2000. "On-Line Interaction and Why Candidates Avoid It." *Journal of Communication* 50(4): 111–132.

Sweet-Cushman, Jennie. 2016. "Gender, risk assessment, & political ambition." *Politics & Life Sciences* 35(2): 117–126.

Teele, Dawn, Joshua Kalla, and Frances Rosenbluth. 2018. "The ties that double bind: social roles and women's underrepresentation in politics." *American Political Science Review* 112(3): 525-541.

Thomsen, Danielle, and Aaron King. 2020. "Women's representation and the gendered pipeline to power." *American Political Science Review* 114(4):989-1000.

Trumm, Siim, Caitlin Milazzo, and Alan Duggan. 2025. "Parliamentary candidates and their campaign messages at the 2019 General Election." *Politics* 45(2): 330-341.

Vallejo Vera, Sebastián, and Analía Gómez Vidal. 2022. "The politics of interruptions: gendered disruptions of legislative speeches." *The Journal of Politics* 84 (3): 1384-1402.

Walter, Annemarie S., and Cees van der Eijk. 2019. "Unintended consequences of negative campaigning: Backlash and second-preference boost effects in a multi-party context." *The British Journal of Politics and International Relations* 21(3): 612-629.

Walter, Annemarie S., Wouter Van der Brug, and Philip van Praag. 2014. "When the stakes are high: Party competition and negative campaigning." *Comparative Political Studies* 47(4):550-573.

Walter, Annemarie S. 2014. "Negative Campaigning in Western Europe." *Political Studies* 62(S1): 42-60.

Wagner, Kevin, Jason Gainous, and Mirya Holman. 2017. "I am woman, hear me tweet!" *Journal of Women, Politics & Policy* 38(4):430-455.

Weeks, Ana. 2009. "Women acting for women?" *Politics & Gender* 5(1): 45-68.

Windett, Jason. 2014. "Gendered Campaign Strategies." *American Politics Research* 42(4):628–655.

Xydias, Christina. 2024. *Beyond Left, Right, and Center*. Philadelphia: Temple University Press.

Biographical Statements:

- Tiffany D. Barnes is a Professor in the Department of Government in the University of Texas at Austin, Austin, TX 78712.
- Charles Crabtree is a Senior Lecturer in the School of Social Sciences at Monash University, Wellington Rd, Clayton VIC 3800 and a K-Club Professor in the University College at Korea University, Seongbuk-gu, Seoul, 02841, Korea.
- Akitaka Matsuo is a Lecturer in the Department of Government at the University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom.
- Yoshikuni Ono is a Professor in the Faculty of Political Science and Economics at Waseda University, Shinjuku, Tokyo, 169-8050, Japan.

Appendix for

"Sentiment on the Campaign Trail: Gender Differences in Candidates' Use of Emotive Language"

Contents

A1 Calculation of Sentiment Scores

The sentiment scores reported in this paper are computed using the `analyzeSentiment` function in the `R` package `SentimentAnalysis`. This function performs dictionary-based sentiment analysis following a bag-of-words approach, ignoring word ordering, sentence syntax, and grammar.

The function first tokenizes the text and then checks whether each token appears in the entries of a sentiment dictionary. Based on this lookup, each word is labeled as positive, negative, or neutral. The positive and negative sentiment scores are calculated as the proportion of tokens labeled as positive or negative, respectively, relative to the total number of tokens in the text.

For example, if a text contains four tokens, of which one is positive and two are negative, the positive score would be $1/4 = 0.25$, while the negative score would be $2/4 = 0.50$. As this example illustrates, the sum of the positive and negative scores cannot exceed 1. Most of the tweets in our dataset do not exclusively contain emotional content (that is, they contain neutral tokens), so for many tweets in our dataset the combined value is typically less than one.

The package includes several built-in sentiment dictionaries. We used the Harvard-IV sentiment dictionary, which contains 2,005 positive words and 1,637 negative words. This dictionary has been repeatedly validated and has been widely used in social science applications (e.g., Dietrich, Hayes and O'Brien 2019).

A2 Descriptive Statistics

Appendix A describes how candidates use Twitter, focusing on variation across our key explanatory variables. Tables A.1–A.3 show the number of candidates with Twitter accounts by party (A1), incumbency status (A2), and gender (A3). It is clear from Table A1 that candidates from the governing party (in both years, the Conservative Party) are less likely to have Twitter accounts compared to candidates from other parties (except for the UKIP candidates). Table A2 shows that the trend for incumbent candidates is the opposite of that for government candidates: incumbent MPs, as opposed to challengers, are more likely to have a Twitter account. Next, Table A3 shows women are more likely to have Twitter accounts, especially in the 2017 election (77.3 percent in 2017; 86.1 percent in 2019), than men (67.4 percent in 2017; 86.1 percent in 2019). We consider the difference in adoption rates across genders an interesting descriptive fact that warrants future explanation.

To assess the volume of content generated by different candidates, Table A4 shows the mean number of tweets by government/opposition, incumbency status, and gender. Government candidates post fewer messages (32.7 in 2017; 37.8 in 2019) than opposition candidates (48.9 tweets in 2017; 61.0 in 2019). Incumbents generate more tweets than challengers (averaging 65.5 vs. 37.6 in 2017; 78.7 vs. 46.9 in 2019). Women candidates are more likely to 'tweet' during the campaign than men, with a higher average number of tweets (54.9 vs. 40.1 in 2017; 66.3 vs. 48.0 in 2019).

Table A1: Candidates with Twitter accounts (by party)

Election	Party	Count	Proportion Has Twitter
2017	Conservative	653	0.743
2017	Labour	686	0.848
2017	Liberal Democrat	644	0.691
2017	Scottish National Party	61	0.984
2017	UK Independence Party	384	0.354
2019	Conservative	635	0.850
2019	Labour	646	0.930
2019	Liberal Democrat	611	0.876
2019	Scottish National Party	59	1.000
2019	UK Independence Party	44	0.341
2019	Brexit Party	275	0.840

Table A2: Candidates with Twitter accounts (by incumbency status)

Election	Status	Count	Has Twitter
2017	Challenger	1,821	0.650
2017	Incumbent	607	0.865
2019	Challenger	1,729	0.852
2019	Incumbent	541	0.939

Table A3: Candidates with Twitter accounts (by gender)

Election	Status	Count	Has Twitter
2017	Women	726	0.773
2017	Men	1,702	0.674
2019	Women	807	0.895
2019	Men	1,463	0.861

Table A4: Twitter Activity

(a) By government and opposition

Election	Government	Prop Active	Avg Num Tweets
2017	0	0.624	48.9
2017	1	0.640	32.7
2019	0	0.830	61.0
2019	1	0.787	37.8

(b) By incumbency

Election	Status	Prop Active	Avg Num Tweets
2017	Challenger	0.571	37.6
2017	Incumbent	0.801	65.5
2019	Challenger	0.794	46.9
2019	Incumbent	0.895	78.7

(c) By gender

Election	Gender	Prop Active	Avg Num Tweets
2017	Female	0.708	54.9
2017	Male	0.594	40.1
2019	Female	0.856	66.3
2019	Male	0.797	48.0

Table A5: Average sentiment by groups

(a) By government and opposition

Election	Government	Avg Pos	Avg Neg
2017	0	0.182	0.084
2017	1	0.200	0.075
2019	0	0.188	0.083
2019	1	0.203	0.072

(b) By incumbency

Election	Status	Avg Pos	Avg Neg
2017	Challenger	0.181	0.083
2017	Incumbent	0.201	0.078
2019	Challenger	0.189	0.081
2019	Incumbent	0.201	0.078

(c) By gender

Election	Gender	Avg Pos	Avg Neg
2017	Female	0.198	0.076
2017	Male	0.182	0.085
2019	Female	0.200	0.078
2019	Male	0.187	0.081

A3 Full Regression Tables

Table A6: Regression Coefficients for the Candidate Tweet Sentiment Models (Figure 1, Full model)

	DV: Positivity			DV: Negativity		
	(1)	(2)	(3)	(4)	(5)	(6)
Woman	0.016^{***}	0.015^{***}	0.015^{***}	-0.007^{***}	-0.006^{***}	-0.005^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Governing Party	0.010^{***}	0.009^{***}	0.010^{***}	-0.006^{***}	-0.005^{***}	-0.007^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Incumbent MP	0.016^{***}	0.017^{***}	0.015^{***}	-0.003^{***}	-0.003^{***}	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Woman x Governing Party		0.004^{*}			-0.005^{***}	
		(0.002)			(0.001)	
Woman x Incumbent			0.003^{*}			-0.006^{***}
			(0.002)			(0.001)
(Region) South East	-0.007^{***}	-0.007^{***}	-0.007^{***}	-0.004^{***}	-0.004^{***}	-0.004^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
(Region) South East	-0.007^{***}	-0.007^{***}	-0.007^{***}	-0.004^{***}	-0.004^{***}	-0.004^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
(Region) East Midlands	0.009^{***}	0.009^{***}	0.009^{***}	-0.009^{***}	-0.009^{***}	-0.009^{***}
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
(Region) London	0.001	0.001	0.001	−0.001	−0.001	−0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
(Region) North East	-0.011^{***}	-0.011^{***}	-0.011^{***}	−0.000	−0.000	0.000
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
(Region) North West	−0.000	−0.000	−0.001	0.000	0.000	0.001
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
(Region) Northern Ireland	-0.014^{***}	-0.014^{***}	-0.014^{***}	0.004	0.004	0.004
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
(Region) Scotland	-0.009^{***}	-0.009^{***}	-0.009^{***}	-0.007^{***}	-0.007^{***}	-0.007^{***}
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
(Region) South West	0.003	0.003	0.003	−0.002	−0.002	−0.002
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
(Region) Wales	-0.006^{**}	-0.006^{**}	-0.006^{**}	-0.014^{***}	-0.014^{***}	-0.014^{***}
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
(Region) West Midlands	0.002	0.002	0.002	0.001	0.002	0.001
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
(Region) Yorkshire and The Humber	−0.001	−0.001	−0.002	0.002	0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Election Year	0.004^{***}	0.004^{***}	0.004^{***}	−0.000	−0.000	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Intercept	-8.406^{***}	-8.464^{***}	-8.450^{***}	0.735	0.803	0.819
	(0.745)	(0.746)	(0.746)	(0.519)	(0.519)	(0.519)
R^2	0.008	0.008	0.008	0.003	0.003	0.003
Adj. R^2	0.008	0.008	0.008	0.003	0.003	0.003
Num. obs.	164496	164496	164496	164496	164496	164496

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table A7: Regression Coefficients for the Reply Sentiment Models (Figure 3, Full model)

	DV: Reply Positivity		DV: Reply Negativity	
	(1)	(2)	(3)	(4)
Woman	0.008***	0.007***	0.008***	0.001**
	(0.000)	(0.001)	(0.000)	(0.000)
Candidate Positivity	0.076***	0.074***		
	(0.001)	(0.001)		
Candidate Negativity			0.109***	0.096***
			(0.001)	(0.001)
Woman x Positivity		0.009***		
		(0.003)		
Woman x Negativity				0.055***
				(0.003)
Incumbent MP	−0.023***	−0.023***	0.016***	0.016***
	(0.001)	(0.001)	(0.000)	(0.000)
Governing Party	−0.009***	−0.009***	0.008***	0.008***
	(0.001)	(0.001)	(0.001)	(0.001)
Election Year	−0.006***	−0.006***	−0.003***	−0.003***
	(0.000)	(0.000)	(0.000)	(0.000)
Intercept	12.450***	12.369***	5.436***	5.473***
	(0.647)	(0.647)	(0.593)	(0.593)
R^2	0.014	0.014	0.014	0.014
Adj. R^2	0.014	0.014	0.014	0.014
Num. obs.	1281498	1281498	1281498	1281498
RMSE	0.134	0.134	0.125	0.125
N Clusters	302083	302083	302083	302083

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table A8: Count Models of Engagement

	Favorite Count	Reply Count
Woman	−1.056***	−0.492***
	(0.001)	(0.004)
Candidate Positivity	−0.260***	−0.266***
	(0.002)	(0.008)
Candidate Negativity	2.169***	2.401***
	(0.002)	(0.008)
Woman x Positivity	−0.094***	0.328***
	(0.004)	(0.013)
Woman x Negativity	0.260***	0.393***
	(0.005)	(0.014)
Incumbent MP	3.049***	2.831***
	(0.001)	(0.003)
Governing Party	−0.791***	0.838***
	(0.001)	(0.002)
(Region) South East	−0.238***	−0.289***
	(0.002)	(0.004)
(Region) South East	−0.238***	−0.289***
	(0.002)	(0.004)
(Region) East Midlands	−0.418***	−0.881***
	(0.002)	(0.006)
(Region) London	1.819***	1.210***
	(0.001)	(0.003)
(Region) North East	−0.214***	−0.785***
	(0.002)	(0.007)
(Region) North West	−0.300***	−0.693***
	(0.002)	(0.005)
(Region) Northern Ireland	−0.606***	−0.599***
	(0.008)	(0.022)
(Region) Scotland	−0.600***	−0.577***
	(0.002)	(0.005)
(Region) South West	−0.720***	−0.862***
	(0.002)	(0.005)
(Region) Wales	−1.389***	−1.522***
	(0.003)	(0.010)
(Region) West Midlands	−0.271***	−0.785***
	(0.002)	(0.005)
(Region) Yorkshire and The Humber	−0.378***	−0.564***
	(0.002)	(0.005)
Election Year	0.671***	0.573***
	(0.000)	(0.001)
Intercept	−1352.246***	−1155.869***
	(0.680)	(2.068)
Deviance	58945311.412	5465806.302
Num. obs.	164496	164496

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

A4 Robustness Check: Alternative Sentiment Coding using LLM

In the main text, we relied on a dictionary-based approach to measure the sentiment of candidate and reply tweets. Dictionary methods are widely used, transparent, and easy to understand. They can be sensitive, though, to the context in which words appear and may not always capture nuances in tone or framing. To assess the robustness of our findings, we conducted an alternative analysis using large language model (LLM)–based sentiment classification, which research shows can be a particularly effective way to capture sentiment (Miah et al. 2024). The following section summarizes the coding procedure and presents the main results, which we compare with those from the dictionary-based analysis in the main text.

For this sentiment analysis, we simplify the continuous polarity measure used in the main text into a binary classification. We adopt this scheme because it facilitates the construction of clear and simple instructions and makes subsequent analysis more straightforward. In the main text, we detected positive and negative sentiment separately, and we retained that framework here. The exact prompt used for sentiment analysis is provided in Appendix Subsection A4.5.

The model we employed was OpenAI's `gpt-4.1-mini`. Coding was conducted via batch processing through the OpenAI API, with each API call classifying 10 tweets. For candidate tweets, all 164,746 were processed. For reply tweets, we randomly sampled about 20% of the corpus (284,114) for coding.[A1] The temperature setting was fixed at 0. Reply tweets were processed on August 25, 2025, and candidate tweets on September 3, 2025.

A4.1 Correlations

Table A9 reports the correlations between dictionary-based sentiment and GPT-based sentiment, calculated separately for positive and negative sentiment in candidate tweets (and in replies).

Table A9: Correlation between Dictionary and GPT Sentiment

Type	Correlation
Candidate Positivity	0.329
Candidate Negativity	0.348
Reply Positivity	0.235
Reply Negativity	0.343

A4.2 Candidate Tweet Sentiment

Table A10 reports the results of re-estimating the candidate sentiment models presented in the main text, replacing the dictionary-based sentiment with LLM-based sentiment classifications. The dependent variable is a binary indicator of whether a tweet is positive or negative, and we modeled it using a linear probability model. As in the main text, robust standard errors are employed to account for possible heteroskedasticity. Regional dummy variables are also included in the estimation, but omitted from the table for space reasons.

Overall, the direction and significance of the coefficients are similar to those in the main analysis, although the interaction terms between woman and incumbency and woman and governing-party status, are not statistically significant here.

To examine the substantive implications of this difference, Figure A1 plots the corresponding predictions. As in the main analysis, we observe significant gender differences in the expected direction. Because we estimate linear probability models, we can interpret the gender effect as the difference in the probability that a tweet is classified as positive or negative. Tweets from women were about 0.1 more likely in probability terms to be positive and roughly 0.07 more likely to be negative.

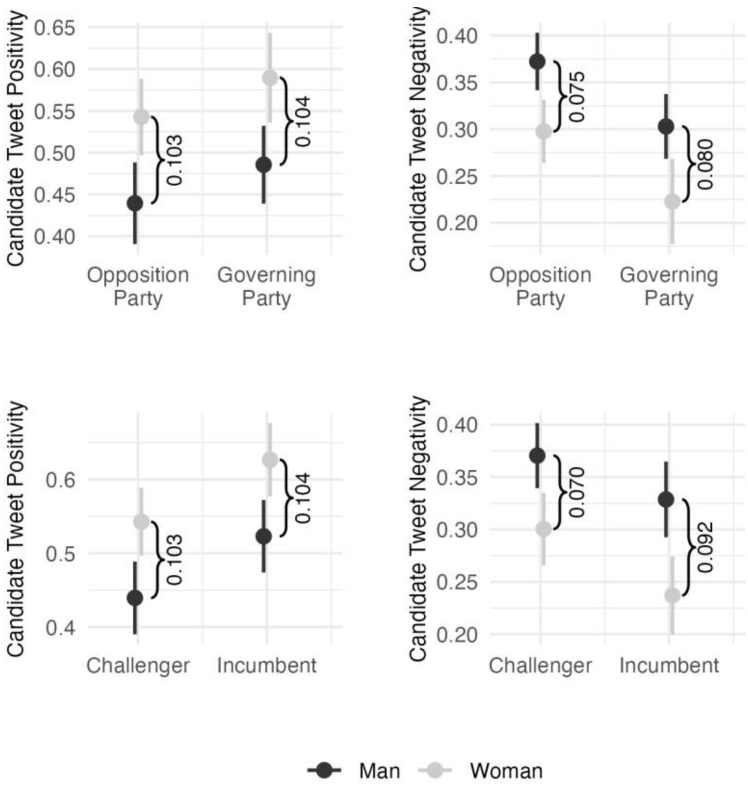
[A1]We processed 300,000 without noticing the sizable number of duplicates being included in the sample.

Table A10: Candidate Sentiment Model (GPT Sentiment)

	DV: Positivity			DV: Negativity		
	(1)	(2)	(3)	(4)	(5)	(6)
Woman	0.103***	0.103***	0.103***	−0.076***	−0.075***	−0.070***
	(0.012)	(0.014)	(0.015)	(0.010)	(0.012)	(0.013)
Governing Party	0.046***	0.046**	0.046***	−0.071***	−0.069***	−0.073***
	(0.012)	(0.015)	(0.012)	(0.011)	(0.013)	(0.011)
Incumbent MP	0.084***	0.084***	0.084***	−0.051***	−0.051***	−0.042**
	(0.011)	(0.011)	(0.014)	(0.010)	(0.010)	(0.013)
Woman x Governing Party		0.001			−0.006	
		(0.027)			(0.026)	
Woman x Incumbent			0.001			−0.022
			(0.023)			(0.021)
Intercept	−92.042***	−92.052***	−92.054***	−29.337**	−29.263**	−29.059**
	(11.778)	(11.765)	(11.772)	(10.236)	(10.201)	(10.210)
R^2	0.030	0.030	0.030	0.016	0.016	0.016
Adj. R^2	0.030	0.030	0.030	0.015	0.015	0.016
Num. obs.	164496	164496	164496	164496	164496	164496
RMSE	0.492	0.492	0.492	0.473	0.473	0.473

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Figure A1: Predicted Probability of Negative and Positive in Candidate Tweets (GPT Sentiment)



A4.3 Reply Sentiment

Next, we examine the sentiment of replies to campaign messages. As in the previous subsection, the dependent variable is a binary indicator of positive or negative sentiment, estimated with a linear probability model (Table A11). When we compare these results with the main-text analysis (Figure 3), we find that the estimates look very similar. The main difference is that the coefficient for the woman dummy becomes strongly negative in the model with negative replies as the dependent variable. This suggests that when a candidate's tweet is not negative, women are less likely to receive negative replies.

We then vary the sentiment of candidate tweets to predict the probability that replies take on positive

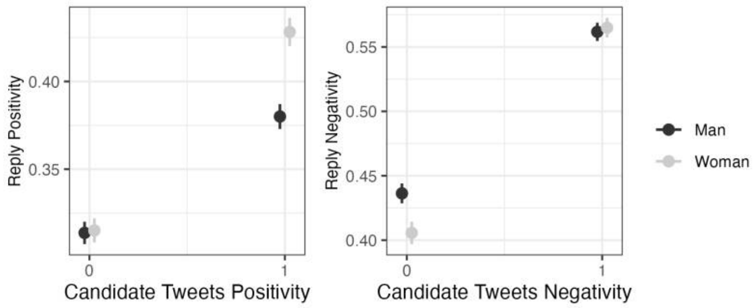
Table A11: Reply Sentiment Model (GPT Sentiment)

	DV: Reply Positivity		DV: Reply Negativity	
	(1)	(2)	(3)	(4)
Woman	0.025***	0.001	−0.012***	−0.031***
	(0.002)	(0.002)	(0.002)	(0.004)
Candidate Positivity	0.079***	0.066***		
	(0.002)	(0.002)		
Candidate Negativity			0.134***	0.125***
			(0.002)	(0.003)
Woman x Positivity		0.047***		
		(0.004)		
Woman x Negativity				0.034***
				(0.004)
Incumbent MP	−0.157***	−0.155***	0.165***	0.164***
	(0.003)	(0.003)	(0.004)	(0.004)
Governing Party	−0.101***	−0.099***	0.080***	0.080***
	(0.002)	(0.002)	(0.004)	(0.004)
Intercept	48.808***	46.764***	−16.023***	−14.015***
	(2.444)	(2.453)	(3.506)	(3.504)
R^2	0.053	0.053	0.041	0.041
Adj. R^2	0.053	0.053	0.040	0.041
Num. obs.	283877	283877	283877	283877
RMSE	0.367	0.366	0.457	0.457

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

or negative values (Figure A2). Compared with the main-text results (Figure 4), the overall claims remain consistent, but there are some differences. For positive replies, when candidate tweets are not positive, there is no gender gap. But when candidate tweets are positive, women are much more likely than men to receive positive replies (a difference of about 0.1 in probability terms). For negative replies, when candidate tweets are not negative, women are less likely to receive negative replies. However, this difference disappears when candidate tweets themselves are negative.

Figure A2: Predicted Probability of Sentiment in Replies to Candidate Tweets (GPT Sentiment)



A4.4 Favorite and Reply Counts

We replicated the engagement analysis from the main text using sentiment classifications generated by GPT, as described earlier in the appendix. Figure A3a reports the predicted number of likes, and Figure A3b shows the predicted number of replies with 95% confidence intervals. The confidence intervals do not appear in some panels because they are very small.

Several results align with the trends reported in the main analysis. In particular, tweets from women receive fewer likes than those from men, regardless of sentiment. Unlike the main-text findings, however, we see a slight tendency here for positive tweets to attract more likes, though the effect is small.

By contrast, negative content generates more engagement overall, but gender differences in baseline

Table A12: Model Coefficents

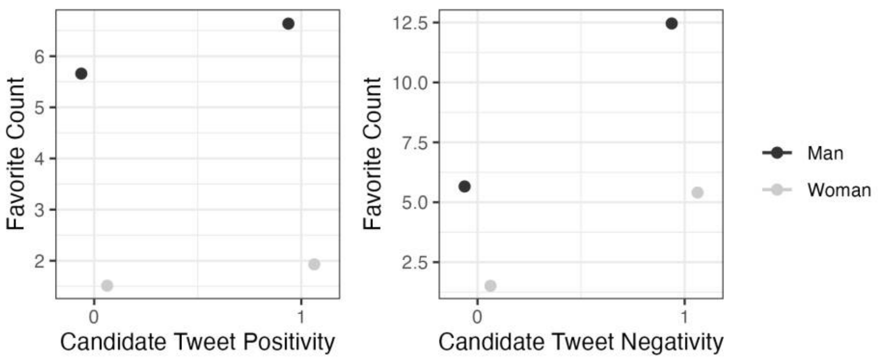
	Favorite Count	Reply Count
Woman	−1.319***	−0.384***
	(0.002)	(0.005)
Candidate Positivity	0.159***	0.319***
	(0.001)	(0.002)
Candidate Negativity	0.789***	0.832***
	(0.001)	(0.002)
Woman x Positivity	0.084***	−0.330***
	(0.002)	(0.005)
Woman x Negativity	0.484***	0.386***
	(0.002)	(0.005)
Incumbent MP	3.082***	2.862***
	(0.001)	(0.003)
Governing Party	−0.706***	0.914***
	(0.001)	(0.002)
Election Year	0.654***	0.567***
	(0.000)	(0.001)
Intercept	−1318.293***	−1144.384***
	(0.684)	(2.089)
Deviance	57777493.177	5360563.476
Num. obs.	164746	164746

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

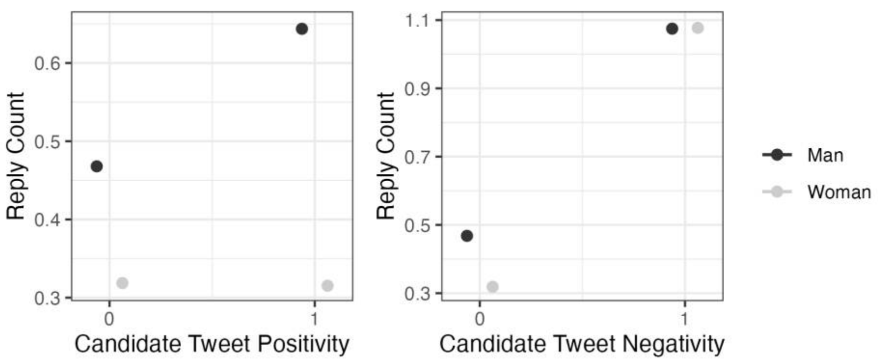
likes do not disappear when candidates post negative tweets; instead, the gap appears to widen. Turning to reply counts, consistent with the main analysis, positive tweets from women tend to receive fewer replies than those from men. For negative tweets, there is little difference in the number of replies between men and women, as reported in the main analysis.

Figure A3: Predicted Number of Likes and Replies to Candidate Tweets

(a) Likes



(b) Replies



A4.5 AI Prompts

A4.5.1 Candidate tweet coding prompt

```
You are an expert political scientist with extensive experience in natural language processing and
sentiment analysis. Classify the sentiment of tweets by UK political candidates based on their emotional
tone.

For each tweet, determine if it contains:
Positive sentiment (approval, support, enthusiasm, gratitude, optimism)
Negative sentiment (criticism, disapproval, anger, dissatisfaction, concern)

A tweet can: Be positive, negative, both, or neither (neutral/factual).

Instructions:

- If purely factual or procedural without emotion, classify as neither.
- Sarcastic praise counts as negative.
- Expressions of hope or determination can be positive, negative, or both.
- Prioritize emotional tone and framing.
- You return the JSON only. You should not provide any explanations.

Output your classifications clearly in this format:
[{"tweet": 1, "positive": true, "negative": true},
{"tweet": 2, "positive": true, "negative": false},
...]

Evaluate carefully, prioritizing accuracy in tone and intent.

[Tweets]
```

A4.5.2 Reply coding prompt

```
You are an expert political scientist with extensive experience in natural language processing and
sentiment analysis. Classify the sentiment of tweets replying to UK political candidates based on their
emotional tone.

For each tweet, determine if it contains:
Positive sentiment (approval, support, enthusiasm, gratitude, optimism)
Negative sentiment (criticism, disapproval, anger, dissatisfaction, concern)

A tweet can: Be positive, negative, both, or neither (neutral/factual).

Instructions:

- If purely factual or procedural without emotion, classify as neither.
- Sarcastic praise counts as negative.
- Expressions of hope or determination can be positive, negative, or both.
- Prioritize emotional tone and framing.
- You return the JSON only. You should not provide any explanations.

Output your classifications clearly in this format:
[{"tweet": 1, "positive": true, "negative": true},
{"tweet": 2, "positive": true, "negative": false},
...]

Evaluate carefully, prioritizing accuracy in tone and intent.

[Tweets]
```

A5 Robustness Check: Comparing with Human Coding

In this subsection, we evaluate the reliability of our dictionary-based sentiment measure by comparing it with sentiment coding conducted by human coders for candidate tweets. The coders consist of one of the paper's authors (Coder 1) and a research assistant (Coder 2, an undergraduate student at a well-known U.S. university). They coded a stratified random sample of 200 candidate tweets, with 100 sampled for positivity and another 100 for negativity. For the 100 tweets selected for positivity, 20 came from tweets with a positivity score of 0. The remaining 80 were drawn from tweets with positive scores, stratified by quartiles of the score, with 20 randomly sampled from each quartile. The same procedure was applied for negativity.

After sampling the 200 tweets, we shuffled them and instructed the coders to assign each tweet a value of 0, 1, or 2 for both positivity and negativity. The coding instructions closely followed the prompt used for AI coding, with the main difference being that coders were asked to classify sentiment on a three-point scale for each polarity. The relevant portion of the instructions is reproduced below.

Positive Sentiment
This includes approval, support, praise, enthusiasm, gratitude, or optimism.
Choose one of the following three levels:

- 0: Not positive – no discernible positivity or completely neutral/factual
- 1: Positive – clear positive emotion or praise
- 2: Strongly positive – strong or intense positive affect (e.g., pride, celebration)

Negative Sentiment
This includes criticism, disapproval, anger, dissatisfaction, concern, or alarm.
Choose one of the following three levels:

- 0: Not negative – no discernible negativity or completely neutral/factual
- 1: Negative – clear criticism or frustration
- 2: Strongly negative – intense negativity (e.g., outrage, severe condemnation)

Figure A13 presents a Pearson correlation matrix comparing the human coding results with other sentiment scores. Several points stand out. First, the correlations between human coding and the dictionary-based sentiment scores are not particularly low. For positivity, both coders show correlations above 0.4 with the sentiment scores. For negativity, the correlations are somewhat lower, but both remain above 0.3. Second, correlations with GPT are higher than those with the dictionary-based scores. If we interpert this test as an assessment of the convergent validity of each sentiment measure, then this seems to suggest that GPT's sentiment scores might have greater construct validity. However, since the instructions given to the human coders were closely aligned with the prompts given to GPT, we should interpret this result with some caution.

Table A13: Correlation Matrix of Various Sentiment Coding

(a) Positivity

	Dictionary	GPT	Coder 1	Coder 2
Dictionary	1.000	0.328	0.448	0.416
GPT	0.328	1.000	0.628	0.588
Coder 1	0.448	0.628	1.000	0.663
Coder 2	0.416	0.588	0.663	1.000

(b) Negativity

	Dictionary	GPT	Coder 1	Coder 2
Dictionary	1.000	0.295	0.375	0.320
GPT	0.295	1.000	0.695	0.736
Coder 1	0.375	0.695	1.000	0.696
Coder 2	0.320	0.736	0.696	1.000

A6 Robustness Check: Dealing with Bots for Reply Tweet Analysis

A6.1 Description

This section in the appendix investigates the extent to which our findings on replies may have been affected by the presence of bots. The analyses in the main text did not explicitly account for bot activity. This section therefore assess whether that omission affects our argument. Research on Twitter bots in the context of British politics, originating with studies of Brexit-related communication, has since expanded considerably.

With respect to the 2017 and 2019 general elections, accounts exhibiting highly automated behavior were clearly active. However, existing scholarship generally concludes that their influence was limited to the amplification of ongoing conversations rather than the initiation of new narratives (Bruno, Lambiotte and Saracco 2022; Kaminska et al. 2017). In the 2017 election, such accounts were more active around Labour-related hashtags than Conservative ones, yet their activity primarily followed the dynamics of human users and served to reinforce existing debates (Kaminska et al. 2017). In the 2019 election, where bot activity intensified during the latter stages of the campaign and peaked on election day, bots functioned mainly by retweeting, thereby increasing the visibility of particular narratives (Bruno, Lambiotte and Saracco 2022). In both elections, their influence thus operated through volume and diffusion rather than direct persuasion.

To detect bots, we draw on BotoMeter (Yang et al. 2020), a widely employed API-based tool that was also used by Bruno, Lambiotte and Saracco (2022). BotoMeter provides probabilistic classifications of whether a Twitter account is automated. As with other detection methods, its utility lies in identifying active bot accounts, but it cannot classify accounts that have since been deleted, suspended, or made private. Therefore, the tool cannot conclusively determine whether accounts in our study period were bots. Moreover, since the effective discontinuation of Twitter's free API in spring 2023, applying such tools to new data has become increasingly impractical. For these reasons, our analysis relies on bot scores generated with BotoMeter included in a publicly available dataset.

Specifically, we employ the dataset released by Calisir and Brambilla (2020), which provides bot annotations for Brexit-related Twitter activity. Using BotoMeter, they created a dataset documenting the presence of automated accounts in discussions about Brexit in the United Kingdom. Their dataset, comprising tweet IDs and user IDs collected between January 2016 and September 2019, overlaps with the temporal scope of our study. Given its coverage of politically active accounts active in UK political discourse, we expect it to encompass a substantial share of the users in our data.

In what follows, we merge their dataset with ours to assess the potential influence of bots on our results.

A6.2 Descriptive Statistics of Bot Scores

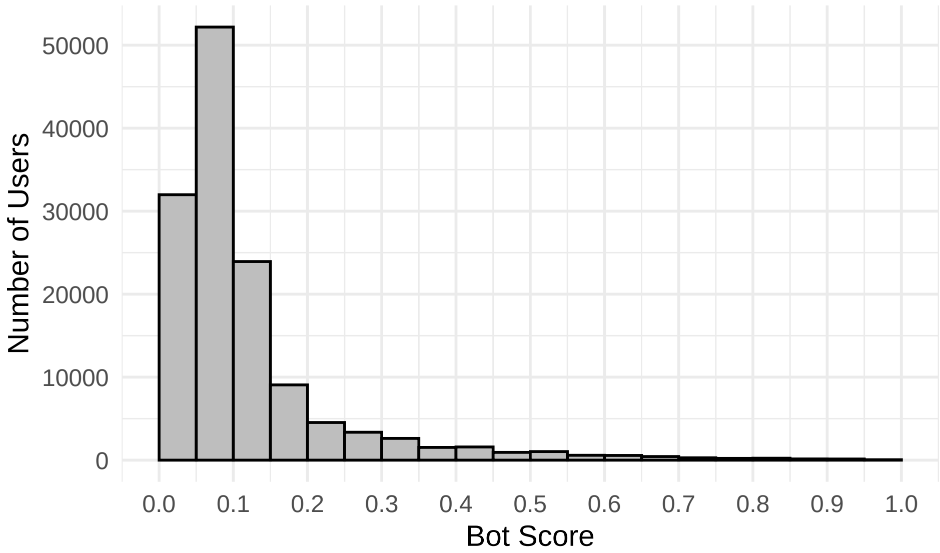
Our reply dataset contains 302,165 unique Twitter accounts, of which 135,430 (44.8%) appear in the dataset provided by Calisir and Brambilla (2020). In terms of tweets, nearly half of them (679,701, or 49.6%) have an associated bot score.

We next examine the distribution of the bot scores assigned to these accounts. The Bot Score generated by the BotoMeter API is a continuous variable ranging from 0 to 1. Figure A4 presents the distribution of these values. As the figure indicates, the majority of accounts fall toward the lower end of the scale, with scores below approximately 0.3.

Since the bot score is a continuous variable, it is necessary to establish a threshold to classify an account as a bot. Yang et al. (2020), in their description of Botometer's updated framework, emphasize that the choice of threshold is not fixed but depends on the evaluation context. Their analysis shows that thresholds around 0.48 maximize classification performance under cross-validation, while a lower threshold of approximately 0.32 yields better results for cross-domain testing across different datasets.

Because the accounts in our dataset differ from those used in Botometer's training data, it is reasonable to adopt the cross-domain threshold. Moreover, to classify accounts that may be bots under a more permissive criterion and thereby include a larger set of accounts in the bot category, we use 0.32 as our main threshold. With a threshold of 0.32, 6.8% of accounts are classified as bots, compared with 3.0% when the threshold is set at 0.48. In the analysis below, our conclusions about the influence of bots do

Figure A4: Distribution of Bot Scores



not change under either threshold.[A2]

A6.3 Compare models

In this subsection, we compare the results of the analyses based on the full sample presented in the main text with those obtained after excluding accounts suspected of being bots or not included in Calisir and Brambilla (2020) dataset.

A6.3.1 Regression Table

Table A7 presents the estimated results of the models. Four models are reported. Model 1 is the positive-sentiment reply model with an interaction term between gender and sentiment, the results of which were also presented in the main text. Model 2 applies the same specification as Model 1 but is estimated on the subsample excluding suspected bots. Model 3 is the negative-sentiment model using the full dataset, while Model 4 is the negative-sentiment model estimated on the subsample excluding bots.

As is evident from the table, the coefficients estimated using the full dataset and those from the no-bot subsample are nearly identical, even though the sample size is reduced by roughly half in the no-bot subsample. Notably, the results are congruent when it comes to the coefficient of the interaction term between candidate gender and sentiment, which is the focus of our analysis.

A6.3.2 Prediction Plot

While the table suggests that bots have only a minimal impact on our findings, we examine this more directly by comparing predicted reply sentiment when candidate tweet sentiment changes. These results are presented in Figure A5. In this figure, the first row corresponds to the positive sentiment models and the second row to the negative sentiment models; the left column shows results from the full dataset, and the right column shows results from the no-bot dataset. As the figure clearly illustrates, the results are virtually indistinguishable between the two samples.

Across both regression estimates and predicted reply sentiment plots, the results are substantively unchanged between the full and no-bot samples. These findings indicate that automated accounts do not likely drive our core conclusions, and that there is no evidence to show that bot activity has a substantial impact on the patterns we report.

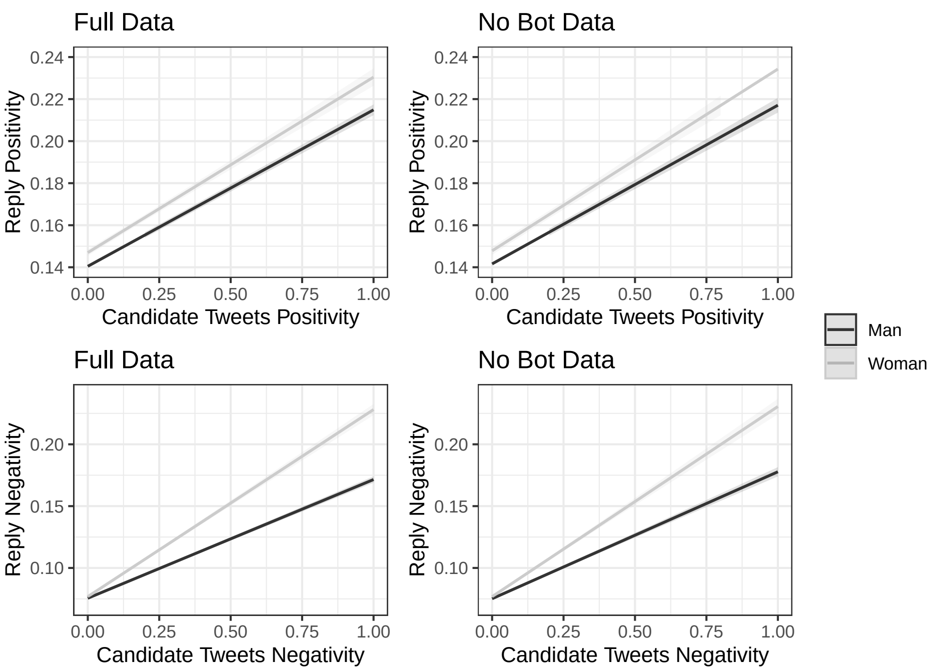
[A2]For reference, Bruno et al. (2022), who also rely on Botometer, employed a threshold of 0.42.

Table A14: Reply Sentiment Model (Comparing Full Data and No Bot Data)

	DV: Reply Positivity		DV: Reply Negativity	
	(1) Full Data	(2) No Bot	(3) Full Data	(4) No Bot
Woman	0.007^{***}	0.006^{***}	0.001^{**}	0.002^{**}
	(0.001)	(0.001)	(0.000)	(0.001)
Candidate Positivity	0.074^{***}	0.075^{***}		
	(0.001)	(0.002)		
Candidate Negativity			0.096^{***}	0.103^{***}
			(0.001)	(0.002)
Woman x Positivity	0.009^{***}	0.011^{**}		
	(0.003)	(0.004)		
Woman x Negativity			0.055^{***}	0.051^{***}
			(0.003)	(0.004)
Incumbent MP	-0.023^{***}	-0.022^{***}	0.016^{***}	0.016^{***}
	(0.001)	(0.001)	(0.000)	(0.001)
Intercept	12.369^{***}	14.458^{***}	5.473^{***}	5.661^{***}
	(0.647)	(0.844)	(0.593)	(0.814)
R^2	0.014	0.016	0.014	0.016
Adj. R^2	0.014	0.016	0.014	0.016
Num. obs.	1281498	635082	1281498	635082

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Figure A5: Predicted Reply Sentiment by Candidate Tweet Sentiment (Full Data vs No Bot Data)



A7 Robustness Check: Additional Control Variables (Age, Parties)

In this section, we re-estimate the models with control variables that were not included in the main text to check whether the results change. Specifically, we add two additional controls: the candidate's age and party dummies.

A7.1 Analysis with Candidate Age

In UK general elections, no unified database exists that provides the ages of candidates. For this reason, we rely on the candidate database published by Democracy Club.[A3] A limitation of this database is that it contains age information for only a relatively small number of candidates. Our candidate dataset contains 2,662 candidates, but we only have the ages of 1,560 (58.6%) of them. As will be seen in the later analysis, when restricting the analysis to only those candidates with known ages, some results change. However, we argue that this is due to sample subsetting and not the inclusion of the age variable itself.

A7.1.1 Candidate Sentiment Analysis

Table A15 presents the results when we include candidate age and compare them with the main-text analysis. Models (2)–(6) correspond to those reported in Table A6. To facilitate comparison, we place the Age model immediately to the right of the baseline specification. As the table shows, the direction and significance of the coefficients remain unchanged. The age variable is significantly negative in the positivity model and significantly positive in the negativity model. As expected from these results, the implications drawn from the predictions of the models including age do not differ from those in the main text (Figure A6).

Table A15: Candidate Sentiment Models with Age

	DV: Positivity				DV: Negativity			
	(2)	(2)-Age	(3)	(3)-Age	(5)	(5)-Age	(6)	(6)-Age
Woman	0.015***	0.020***	0.015***	0.021***	−0.006***	−0.009***	−0.005***	−0.009***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Governing Party	0.009***	0.012***	0.010***	0.010***	−0.005***	−0.003*	−0.007***	−0.004***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Incumbent MP	0.017***	0.019***	0.015***	0.021***	−0.003***	−0.003**	0.000	−0.003*
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Woman x Governing Party	0.004*	−0.006*			−0.005***	−0.003		
	(0.002)	(0.003)			(0.001)	(0.002)		
Woman x Incumbent			0.003*	−0.004			−0.006***	−0.000
			(0.002)	(0.002)			(0.001)	(0.002)
Candidate Age		−0.005***		−0.005***		0.002***		0.002***
		(0.000)		(0.000)		(0.000)		(0.000)
Intercept	−8.464***	−8.355***	−8.450***	−8.350***	0.803	−1.592	0.819	−1.650
	(0.746)	(1.226)	(0.746)	(1.228)	(0.519)	(0.876)	(0.519)	(0.877)
R^2	0.008	0.013	0.008	0.013	0.003	0.005	0.003	0.005
Adj. R^2	0.008	0.012	0.008	0.012	0.003	0.004	0.003	0.004
Num. obs.	164496	78020	164496	78020	164496	78020	164496	78020

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

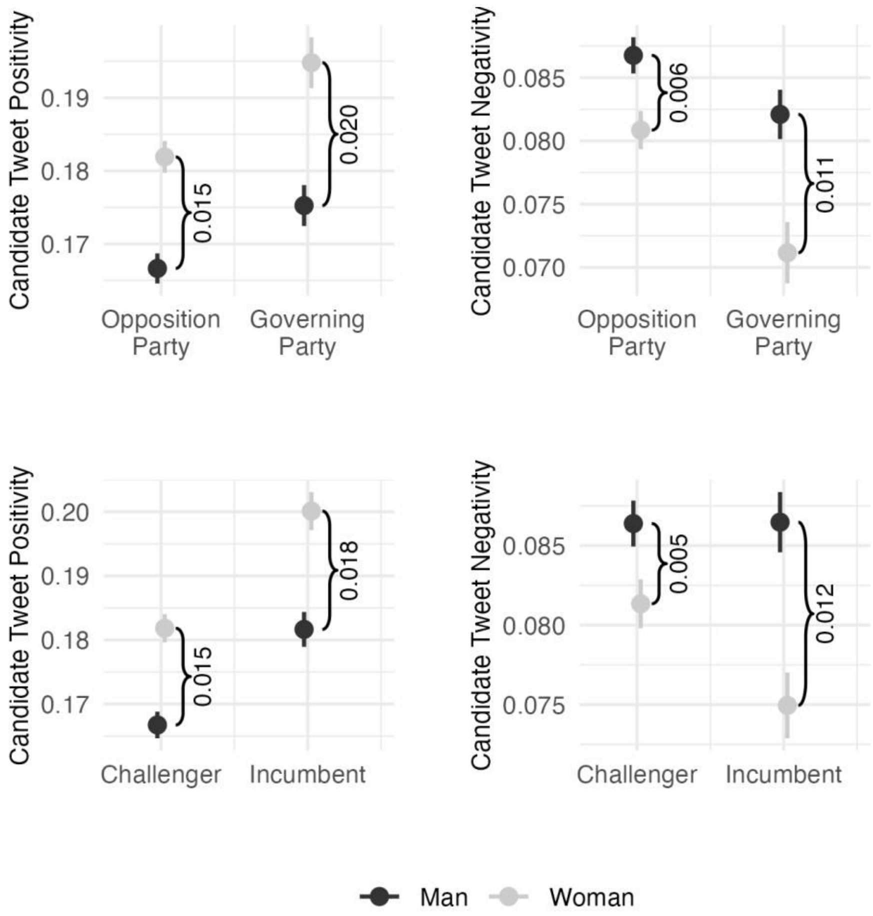
A7.1.2 Favorite and Reply Counts

The results of the like and reply count models differ in some respects from those in the main text. Figure A7 plots the predicted values. In each subfigure, the top row shows the models that include the Age control, while the bottom row shows the models without it. For the like count model, the results remain consistent with the main text (Figure A7a). However, for the reply count model, both the positivity and negativity specifications indicate that women candidates receive more replies than men when they post emotionally expressive tweets (Figure A7b).

As the second row of subfigures (models without the Age control) shows, the predictions change very little, indicating that the differences do not result from including candidate age but rather from sample

[A3]https://candidates.democracyclub.org.uk/

Figure A6: Predicted Probability of Negative and Positive in Candidate Tweets (with Age Control)



bias. Candidates with known ages do not represent a random sample of all candidates. For example, the proportion of candidates with known ages is lower for UKIP and the Brexit Party than for other parties (e.g., Brexit: 0.399, Conservatives: 0.525). Similarly, when comparing by gender, women are less likely than men to have recorded ages (0.400 to 0.499).

A7.2 Controlling Parties

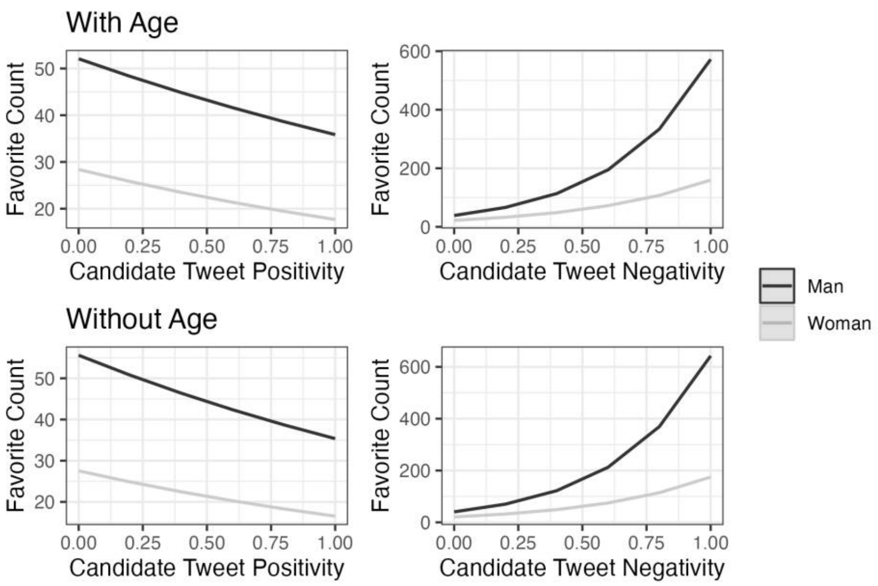
In the main-text, we included a governing-party dummy in the models for candidate tweets. In both general elections used in our study (2017 and 2019), the Conservative Party held government. This means that the dummy variable is effectively equivalent to a Conservative-versus-all-others indicator. This may have influenced the tests of our hypotheses regarding the effect of incumbency. To check this possibility, we re-estimate the models by replacing the governing-party dummy with party dummies for each major party and compare the results. The reference category is the Conservative candidates.

A7.2.1 Candidate Sentiment Analysis

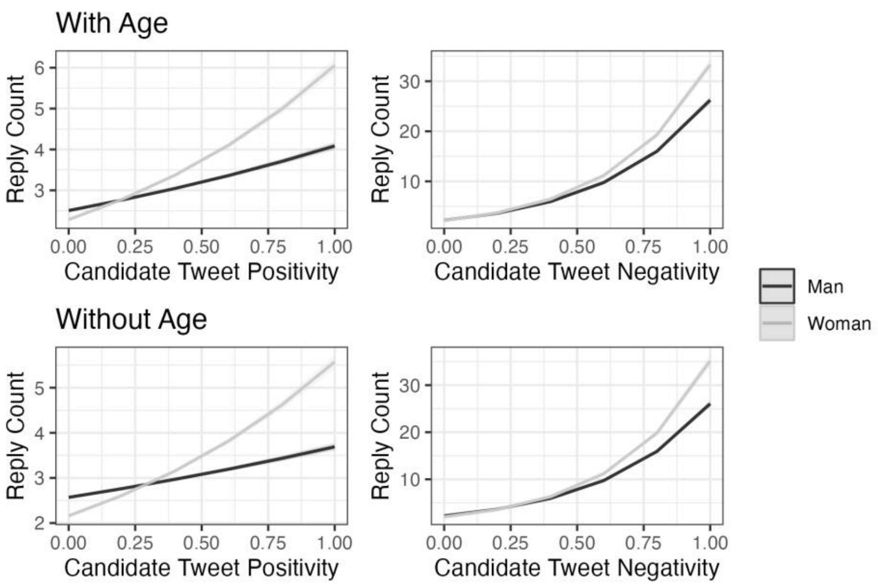
We first examine the models where the dependent variable is the sentiment of candidate tweets (Table A16). The table presents side-by-side results from the governing-party dummy model and the party-dummy model. The coefficients related to women, the woman dummy and its interaction with incumbency, remain consistent in terms of significance and direction, and their magnitudes do not change substantially. We also generated prediction plots, setting the candidate's party to Labour in these predictions. The predictions likewise show essentially the same results (Figure A8).

Figure A7: Predicted Number of Likes and Replies to Candidate Tweets

(a) Likes



(b) Replies



A7.2.2 Favorite and Reply Counts

We also estimated models that took the number of likes and replies as dependent variables. As in the previous subsection, the results are substantively the same (Table A17; Figure A9).

Table A16: Candidate Sentiment Model with Party Dummy

	DV: Positivity		DV: Negativity	
	(3)	(3)-Party	(6)	(6)-Party
Woman	0.015***	0.013***	−0.005***	−0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Governing Party	0.010***		−0.007***	
	(0.001)		(0.001)	
Labour		−0.005***		0.005***
		(0.001)		(0.001)
Liberal Democrats		−0.017***		0.013***
		(0.001)		(0.001)
SNP		0.003		−0.017***
		(0.002)		(0.002)
UKIP		−0.028***		0.016***
		(0.002)		(0.001)
Brexit Party		−0.027***		0.021***
		(0.002)		(0.001)
Incumbent MP	0.015***	0.008***	0.000	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)
Woman x Incumbent	0.003*	0.004*	−0.006***	−0.007***
	(0.002)	(0.002)	(0.001)	(0.001)
Intercept	−8.450***	−8.642***	0.819	1.479**
	(0.746)	(0.789)	(0.519)	(0.549)
R^2	0.008	0.010	0.003	0.006
Adj. R^2	0.008	0.010	0.003	0.006
Num. obs.	164496	164496	164496	164496

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Figure A8: Predicted Probability of Negative and Positive in Candidate Tweets (with Party Dummy)

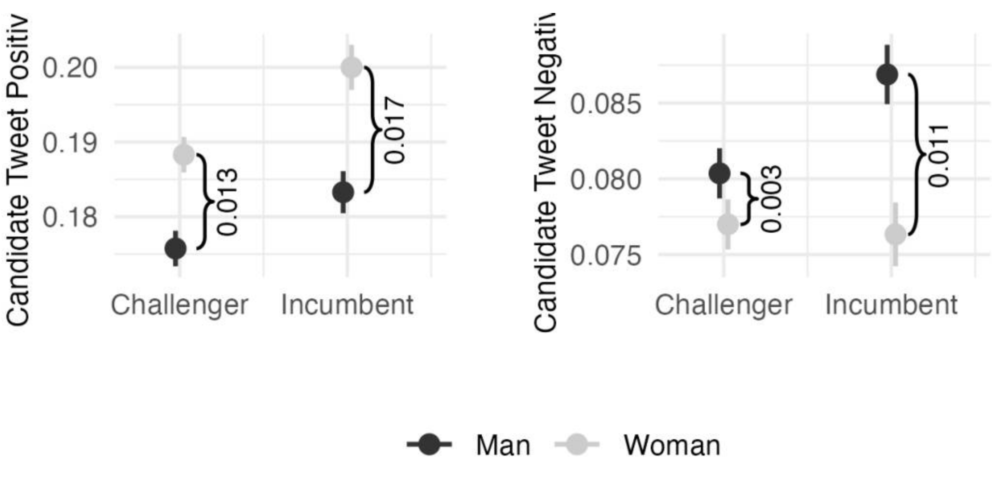


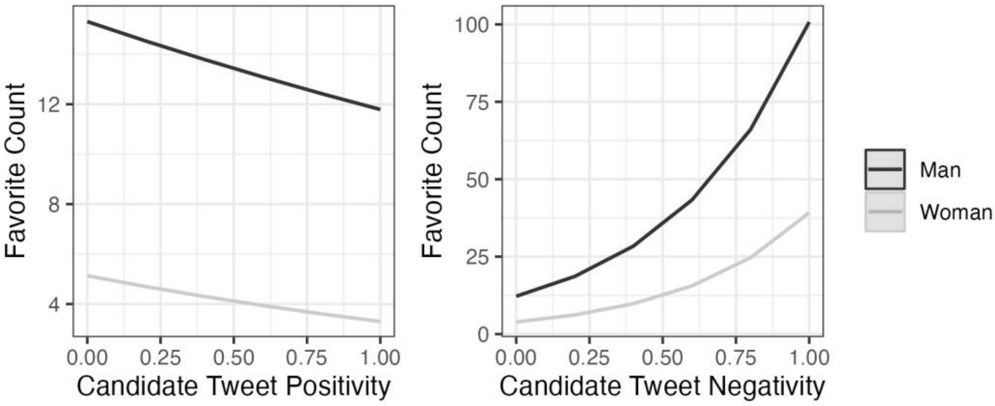
Table A17: Favorite and Reply Count Models with Party Dummy

	DV: Favorite Count		DV: Reply Count	
	Gov Pty	Pty Dummy	Gov Pty	Pty Dummy
Woman	-1.056^{***}	-1.110^{***}	-0.492^{***}	-0.482^{***}
	(0.001)	(0.001)	(0.004)	(0.004)
Candidate Positivity	-0.260^{***}	-0.261^{***}	-0.266^{***}	-0.251^{***}
	(0.002)	(0.002)	(0.008)	(0.008)
Candidate Negativity	2.169^{***}	2.111^{***}	2.401^{***}	2.388^{***}
	(0.002)	(0.002)	(0.008)	(0.008)
Woman x Positivity	-0.094^{***}	-0.180^{***}	0.328^{***}	0.307^{***}
	(0.004)	(0.004)	(0.013)	(0.013)
Woman x Negativity	0.260^{***}	0.198^{***}	0.393^{***}	0.326^{***}
	(0.005)	(0.005)	(0.014)	(0.014)
Incumbent MP	3.049^{***}	2.696^{***}	2.831^{***}	2.947^{***}
	(0.001)	(0.001)	(0.003)	(0.003)
Governing Party	-0.791^{***}		0.838^{***}	
	(0.001)		(0.002)	
Labour		0.956^{***}		-0.755^{***}
		(0.001)		(0.002)
Liberal Democrats		-0.337^{***}		-1.062^{***}
		(0.001)		(0.004)
SNP		0.418^{***}		-2.259^{***}
		(0.003)		(0.008)
UKIP		1.098^{***}		0.568^{***}
		(0.004)		(0.009)
Brexit Party		0.280^{***}		-0.460^{***}
		(0.002)		(0.007)
Election Year	0.671^{***}	0.713^{***}	0.573^{***}	0.579^{***}
	(0.000)	(0.000)	(0.001)	(0.001)
Intercept	-1352.246^{***}	-1437.017^{***}	-1155.869^{***}	-1168.269^{***}
	(0.680)	(0.691)	(2.068)	(2.113)
Deviance	58945311.412	57584621.809	5465806.302	5403019.820
Num. obs.	164496	164496	164496	164496

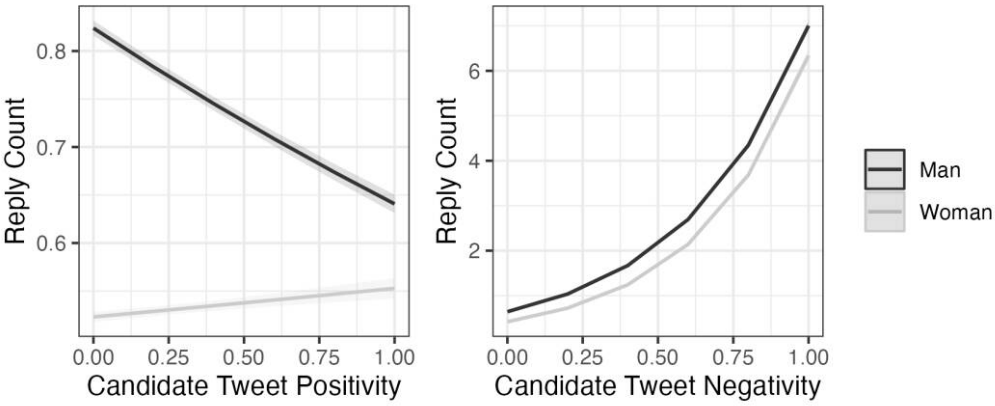
$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Figure A9: Predicted Number of Likes and Replicies to Candidate Tweets

(a) Likes



(b) Replies



A8 Example Tweets

Table A18 presents a set of example tweets selected to illustrate how our sentiment scores behave. For privacy reasons, URLs as well as the names and Twitter handles of individuals who are not public political figures have been anonymized. The tweets are sorted in descending order in the value of (Positivity – Negativity).

Table A18: Examples of Candidate Tweets by Sentiment

Positivity	Negativity	Candidate Info	Tweet Text
0.571	0.071	Male, Governing Party (Conservative), Incumbent	Excellent evening out on doorsteps meeting residents with my brilliant colleague Cllr Saroy Great response for @Conservatives @theresa_may https://***
0.429	0.000	Female, Opposition (Labour), Challenger	Thanks! Labour gives kids opportunities (including me when I was a kid & had free music lessons). VoteLabour June 8th https://***
0.467	0.067	Male, Governing Party (Conservative), Incumbent	NHS in London doing magnificent job over past 22 hours to help save lives. Remarkable dedication and bravery of NHS staff. https://***
0.400	0.000	Female, Governing Party (Conservative), Incumbent	Working for a Conservative majority government - supporting @jackmrankin who would be an excellent local MP. #ge2019 #GetBrexitDone https://***
0.333	0.000	Female, Governing Party (Conservative), Incumbent	Whether it's in the South, Midlands or North we stand with colleagues and agree the terrorists will not win, our hearts go out to all https://***
0.333	0.000	Male, Opposition (Labour), Incumbent	Enjoyed being out doorknocking in Seacroft this morning and in Gipton this afternoon with brilliant Labour volunteers! Great response! https://***
0.200	0.000	Female, Opposition (Labour), Incumbent	Labour will increase investment in teachers and schools- giving all our children a brighter future
0.182	0.000	Female, Governing Party (Conservative), Challenger	We @ScotTories are the only party that Agriculture & Fisheries Industries can rely on for support through #Brexit. #GE2017 #***
0.154	0.000	Female, Governing Party (Conservative), Incumbent	Talking education in Enfield North yesterday with #GE2017 candidate @nickdebois, now many more good/outstanding Enfield schools than in 2015 https://***
0.130	0.000	Female, Opposition (Labour), Incumbent	Great energy and amazing turnout in #chingford today @*** @uklabour #letsdothis #vvotelabour Please vote early and Vote Labour. Every day more and more people are voting for change and voting https://***

Table A18: Examples of Candidate Tweets by Sentiment *(continued)*

Positivity	Negativity	Candidate Info	Tweet Text
0.120	0.040	Female, Governing Party (Conservative), Incumbent	It's time to #GetBrexitDone in order to move forward so we can deliver a transformative one nation agenda- investing in our #NHS, putting more police on our streets & ensuring we have fairer funding for our schools & colleges. #Voteconservativeactually
0.125	0.062	Female, Opposition (Labour), Incumbent	Tories talk about how they're funding NHS but won't tell you this. @UKLabour will give the NHS the money it needs. #VoteNHS #VoteLabour2017 https://***
0.125	0.062	Female, Governing Party (Conservative), Challenger	Best way to start the day on @ScotTories #GE2019 street stall! For those who were at @DennistounCC hustings, I don't mess about! #savoury #sweet #mincepies #GlasgowNorthEast https://***
0.062	0.000	Male, Governing Party (Conservative), Incumbent	Another busy day on the campaign trail today: Stalbridge, Blandford, Marnhull, Sturminster Marshall. I'm enjoying, as always, the conversations and exchanges. #6moresleeps
0.000	0.000	Female, Governing Party (Conservative), Incumbent	Out campaigning in North Kensington this afternoon #Back*** #VoteConservative #*** https://***
0.000	0.000	Male, Governing Party (Conservative), Challenger	Busy day in WAK. Banchory, Kemnay & Sauchen before the only hustings of the campaign. Tomorrow...back to it. 6 days to #winbackwak #GE2017
0.000	0.000	Male, Opposition (Liberal Democrat), Challenger	On @BBCCovWarks radio from 6pm, tune in!
0.000	0.059	Male, Governing Party (Conservative), Incumbent	@*** If you vote Lib Dem you are voting to devalue the country's democratic decisions and let antisemitic revolutionary socialists take power. Think about that for a minute.
0.000	0.062	Female, Opposition (Labour), Incumbent	Today I was at Lord Blyton Primary School, another Tory Government would see an average cut of £662 per pupil by 2020 vote @UKLabour https://***
0.000	0.067	Male, Opposition (Labour), Incumbent	Johnson and Farage want to flog off our NHS to Trump this Christmas. We won't let them in East Hull. Vote Labour 12 December. https://***
0.000	0.091	Female, Governing Party (Conservative), Challenger	Farmers all over are in record debt due to the SNP's mess and mismanagement of CAP and our rural economy. #ScotDebates

Table A18: Examples of Candidate Tweets by Sentiment *(continued)*

Positivity	Negativity	Candidate Info	Tweet Text
0.000	0.100	Female, Opposition (Labour), Incumbent	At least 135,000 children in Britain to be homeless at Christmas In the 5th richest country in the world this is #disgraceful https://***
0.000	0.143	Male, Governing Party (Conservative), Challenger	Labour doubling down on the US resident, hard remain multi millionaire vote in Houghton and Sunderland South. #VoteConservative2019 to #GetBrexitDone https://***
0.000	0.231	Male, Governing Party (Conservative), Incumbent	No #MagicMoneyTree – we need a strong economy to fund strong public services – and we won't get that with Corbyn. #VoteConservative #BBCQT
0.000	0.286	Female, Governing Party (Conservative), Incumbent	The problem with Labour, is that they always run out of other people's money..... https://***
0.000	0.333	Male, Opposition (Labour), Incumbent	Boris Johnson lies again
0.000	0.364	Female, Governing Party (Conservative), Challenger	Agreed @***. I tried not to post on this today, but the blatant lies and unpatriotic apportioning of blame could not go unchallenged
0.000	0.600	Male, Governing Party (Conservative), Challenger	More misleading misinformation from Labour exposed... https://***

References

Bruno, Matteo, Renaud Lambiotte and Fabio Saracco. 2022. "Brexit and Bots: Characterizing the Behaviour of Automated Accounts on Twitter during the UK Election." *EPJ Data Sci.* 11(1):1–24.

Calisir, Emre and Marco Brambilla. 2020. "Twitter Dataset about Brexit.".
URL: *https://doi.org/10.7910/DVN/KP4XRP*

Dietrich, Bryce J., Matthew Hayes and Diana Z. O'Brien. 2019. "Pitch Perfect: Vocal Pitch and the Emotional Intensity of Congressional Speech." *American Political Science Review* 113(4):941–962.

Kaminska, Monica, John D. Gallacher, Bence Kollanyi, Taha Yassari and Philip N. Howard. 2017. Social Media and News Sources during the 2017 UK General Election. Technical report Oxford Internet Institute.

Miah, Md Saef Ullah, Md Mohsin Kabir, Talha Bin Sarwar, Mejdl Safran, Sultan Alfarhood and MF Mridha. 2024. "A multimodal approach to cross-lingual sentiment analysis with ensemble of transformer and LLM." *Scientific Reports* 14(1):9603.

Yang, Kai-Cheng, Onur Varol, Pik-Mai Hui and Filippo Menczer. 2020. "Scalable and Generalizable Social Bot Detection through Data Selection." *Proceedings of the AAAI Conference on Artificial Intelligence* 34(01):1096–1103.