When Office is not an Option: Policy Profiles in the UK's Final European Election Campaign

Brian Boyle, Sebastian Popa, and Zoltan Fazekas

May 2022

EARLY DRAFT PLEASE DO NOT CIRCULATE OR CITE WITHOUT PERMISSION

1 Introduction

What motivates candidates to run for political office is a fundamental question in political science, as who runs, and ultimately who wins, can shed light on core issues of persuasion, competition, and representation. One key source of motivation stems from the existence of office related benefits (Callander 2008, Wittman 1983), however, we know far less about how this type of incentive influences candidates' campaign strategies. More specifically, it is unclear whether this impacts candidates' policy or vote seeking communication strategies, and how they handle intra-party disagreements.

We exploit the unique situation that UK candidates faced in the 2019 European Parliament elections, where these direct office motivations were curbed. Despite initial plans for the UK to have exited the EU well before May 2019, multiple delays and extension requests resulted in the UK participating in what would be its final European contest. Candidates were thereby left to conduct an election campaign under relatively short notice, with a lack of long-term planning, to a body which they will soon no longer be a member of. We take a specific focus on candidates' online communications to assess how these factors influenced candidates' social media campaign strategies.

On the one hand, the hastily organised campaign and the dominance of Brexit as an issue, should have lead to a lack of party control and provided more freedom for individual candidates to build distinctive issue profiles beyond Brexit. Hence, under these circumstances we would expect to observe increased within-party heterogeneity when it comes to issues emphasised beyond Brexit. On the other hand, the inevitability of Brexit substantively diminished office-seeking motivations for those in eligible positions, and made their future political career dependent on following the party line. This in turn should lead to more similarity between highly electable and unpromising candidates in terms of their motivations to follow the party line, and thus reducing within-party differences across issue profiles.

We test these competing expectations using original data collected from the Twitter communications of approximately 5000 candidates competing in the 2019 EP elections across all EU Member States. We use a combination of human coding and machine learning across more than 10 languages, to compare patterns in candidates' twitter communications between the UK and other European nations in the 2019 European Parliament elections.

Leveraging this unique situation allows us to explore the underlying motivations of why candidates stand for office, which has direct implications for both the types of policy platforms offered to voters, as well as policy outcomes themselves.

2 2019 European Parliament Elections in the UK

In attempting to uncover the motivations for both candidates and parties to participate in this election, it is essential to first examine the national political context in light of the 2016 Brexit vote, its run-on effects on the mechanics of the campaign, and how this influenced the parties' positions in the lead up to the election.

In June 2016, the UK public voted to leave the European Union by a margin of 52-48%. The immediate aftermath of the referendum saw a new prime minister in Theresa May, replacing David Cameron who resigned in light of his decision to both hold the referendum, and his support for the losing remain side. By February of 2017, the UK parliament voted to pass the European Union bill that would allow the government to invoke article 50 of the Lisbon Treaty, thereby formally starting the process of leaving the EU. Article 50 was officially submitted on the 29th March 2017, triggering a two-year deadline for exiting the

EU, meaning the UK would have left the bloc before the European Parliamentary elections scheduled for May of 2019.

Theresa May's withdrawal agreement, which set out the specific terms for leaving the EU, failed to pass through the UK legislature when put to MPs in both January and March of 2019. With no agreed terms in place, the UK government submitted two requests to extend the withdrawal period in March and April of 2019, which ultimately pushed the exit deadline back to October 2019. One major consequence, however, was that the UK would still formally be a member of the EU at time of the European Parliament elections on 23-26 May 2019, and thus would be legally required to participate (Vasilopoulou 2020, Martill 2020).

The UK government's desire to exhaust all possible options to push through the withdrawal agreement and avoid this exact scenario, meant that the final decision to participate was left until the last possible moment, with Theresa May announcing the UK's participation in the contest on 7th May, just 3 weeks before the election. Set amongst the background of the relatively unplanned and short-term nature of the campaign, and the failure to reach agreement on the withdrawal process, the Brexit issue continued to dominate the campaign period for practically all UK parties.

The two largest parties were both internally divided on Brexit, with the Conservatives torn between those in favour of a more hard-line 'no-deal' Brexit, versus those who wished to retain common trade rules with the EU along the lines of the May's proposed withdrawal agreement (Vasilopoulou 2020). In light of the Conservative's failure to deliver Brexit within the initial two year time frame, and perhaps in anticipation of being punished by voters for these shortcomings, the conservative campaign was notably half-hearted (Martill 2020), with several MPs admitting to not campaigning, and the party opting to only publish a short leaflet rather than any formal manifesto (Vasilopoulou 2020).

Meanwhile, within the labour party, the leadership's platform of respecting the result of the referendum, but delivering a Brexit that retained closer ties with Europe, was at odds with large segments of party membership who instead favoured holding a second referendum on the issue (Vasilopoulou 2020). These pressures resulted in Labour campaigning on a manifesto based on opposing both a no-deal exit from the EU, as well as the government's proposed withdrawal agreement, and instead promoting their revised version of Brexit which emphasised a closer formal ties with the EU (Vasilopoulou 2020). Qualified support was also offered to the notion of holding a second referendum, but only in the event that no consensus could be reached on this revised Brexit vision, or if a general election was held beforehand (Vasilopoulou 2020).

The smaller parties within the UK were far more internally consistent when it came to their Brexit positions. The liberal democrats explicitly campaigned for holding a second referendum and supported remaining in the EU, a platform which was also shared by the Greens, SNP, and Plaid Cymru (Fella et al. 2019). The UK Independence Party (UKIP) meanwhile, stood on a clear anti-EU programme, having been one of the major proponents of the successful leave campaign in period leading up to the referendum itself.

Alongside these more longstanding smaller parties, two newly formed single issue groups competed in the 2019 elections, centering their campaigns almost entirely upon Brexit, from opposing ends of the spectrum. Change UK/The Independent Group, made up of several pro-European former Labour and Conservative MPs, formed in February 2019 and campaigned on the pledge of holding a second referendum on EU membership. Nigel Farage's Brexit party was created in November 2018 with the sole aim of ensuring that the UK followed through with leaving the European Union.

Coming into the 2019 elections then, the electoral landscape in the UK saw the two largest parties internally divided on Brexit, while for the smaller parties, the Brexit party occupied the majority of the territory on the anti-EU side, while the pro-European push for holding a second referendum was taken up by several parties including the Liberal Democrats, Green, SNP, Plaid Cymru and Change UK.

3 Party Campaigning strategy

In campaigns, parties can mix what they talk about in terms of core issues, i.e. issues they own (Ennser-Jedenastik, Gahn, Bodlos & Haselmayer 2021), and issues that are popular among the public (Klüver & Sagarzazu 2016, Ennser-Jedenastik, Gahn, Bodlos & Haselmayer 2021, Barberá et al. 2019). In the context of the less competitive second-order elections, parties have been shown to be less responsive to the public (Klüver & Sagarzazu 2016, Abou-Chadi 2018), and rather focus on more entrepreneurial strategies by highlighting fringe issues that give them the chance to increase their electoral margins (De Vries & Hobolt 2020). Furthermore, no matter which logic parties follow (i.e. responsiveness vs. issue ownership), or which issues they choose to engage with as a result, the communication strategy of the party should consider some combination of office-seeking, vote-seeking and/or policy-seeking incentives.

The recent advent of social media use in political campaigns allows us to study these decisions relying on the continuous flow of campaign communication. Social media is by now a well established communication tool for political parties. While the issue-content of social media communication generally matches the issue profile established through more traditional forms of communication (Gilardi et al. 2021, Peeters et al. 2021), it is notable that (at least in comparison to TV ads) such communication is approximately 10 percentage points less likely to mention issues, while at the same time allowing for a greater issues diversity during campaigns (Fowler et al. 2021). In this context, previous research shows that parties and politicians use social media platforms strategically, to attract attention to issues and events that benefit them (Barberá et al. 2019, Fazekas et al. 2021), and contribute to informing the both the public and their party supporters (Munger et al. 2022, Popa et al. 2020). These practices are particularly common during electoral campaigns when parties communicate about highly salient issues, especially when such issues can contribute to their electoral success (Munger et al. 2022, Nulty et al. 2016). In this regard, we know that parties can shape the salience of given issues by setting the agenda, but they can also use communication strategies to downplay issues that are not to their advantage (Fazekas et al. 2021). More often than not, parties use social media to amplify their overall message (Silva & Proksch 2021).

In parallel with these insights where parties are regarded as leaders rather than followers in terms of agenda setting, evidence from the highly polarised U.S. context suggest that parties on social media respond to the issues the public cares about, especially their own supporters (Barberá et al. 2019). However, in other contexts, the ability to be effective in responding to the public's issue preferences is limited to large parties that have the resources to actively monitor social media (Kruschinski & Haller 2017, Zuiderveen Borgesius et al. 2018). But even large European parties are reluctant to show responsiveness across all issues, as they only appear to adopt such strategies for issues they consider less salient (Ennser-Jedenastik, Gahn, Bodlos & Haselmayer 2021). Given that we can observe lover levels of responsiveness in a scenario of low contestation (Binzer Hobolt & Klemmensen 2008) such as that of second-order elections (Klüver & Sagarzazu 2016, Abou-Chadi 2018), we assume parties in the 2019 EP election campaign pay less attention to responsiveness.

In the context of EP elections, office-seeking motivation or possible office related benefits are generally unclear beyond possible incumbency and influence at supranational levels, but in the context of the 2019 EP elections in the UK the immediate benefits of office have been reduced to at best minimal, since future UK MEPs were expected to serve a fraction of their full term. In contrast, vote-seeking incentives were still important, as electoral performance in these EP elections could serve as indicators of party standing strength and potential re-confirmation of Brexit stances. Theresa May's resignation at least partially motivated by the disastrous results of the Conservative Party in these EP elections are a testament to such electoral stakes. Nevertheless, in the context of a hastily announced election and relatively unorganised campaign with little resources committed to it, parties did not seem to be motivated to heavily invest in gaining votes. Furthermore, the responsiveness expected to underlie vote-seeking strategies is less constraining on party strategies during EP elections. Thus, vote-seeking strategies that might guide parties to follow a coherent public issue responsiveness strategy were less dominant in the UK.

Additionally, in the context of Brexit, while it was important for parties to clarify their position on this issue, there was little room for building a more complex issue profile and highlight the issues parties own (Vasilopoulou 2020, Martill 2020). Hence, policy seeking rationales where dominated by debates around Brexit and the EU, with less strategic focus on other issues. Overall, and specifically in the UK, this results in party strategies that should be less constrained by strategic considerations that would normally lead to a coherent campaign communication aimed at either responsiveness or building party issue profiles. Hence, the often documented congruence between the issues profiles of parties and individual politicians on social media (Gilardi et al. 2021) should be reduced under such circumstances. Practically, the inevitability of the Brexit process leaves the UK candidates for the 2019 EP elections in an unexpected position from which to carry out their campaign, they have more freedom to pursue individualised communication strategies, albeit they are constrained by the main issue dominating the campaign. And thus at the party level we would expect to observe that beyond Brexit, there is more heterogeneity within UK parties when it comes to building coherent policy in comparison with other European parties, or in other words:

H1 Candidates from UK parties will be less likely to discuss their party's core issues, in comparison to candidates from other EU countries.

4 Candidate campaigning strategies

While party strategy can be centrally decided, individual candidates, do have at least some freedom within the boundaries set by parties (Ennser-Jedenastik, Haselmayer, Huber & Fenz 2021). The same is true for social media communication where messages generally have a strong partisan appeal, but individual politicians use social media to express a broader range of opinions than through traditional tools of communication (Silva & Proksch 2021). Beyond building a specific issue profile, social media communication also allows candidates to build ideological profiles that are distinct from the official party stance (Barberá 2015, Ceron 2017). More generally, social media allows candidates to build individualised and personalised campaigns even in closed lists electoral settings (Chadwick 2017, Karlsen & Enjolras 2016).

Previous research shows that Twitter allows politicians to engage with multiple issues, and this engagement is guided by both their party issue ownership, but especially by their own issue specialisation (Peeters et al. 2021). This finding points to the fact that social media is a useful tool that allows candidates to build individual profiles without ignoring the issues that are important for the party. But the way in which specific candidates combine the two incentives is primarily guided by their position within the party hierarchy. On the one hand, top-office holders guided by office-seeking motivations are expected to be more responsive to public opinion and thus engage more with the issues that the public considers important (Ennser-Jedenastik, Haselmayer, Huber & Fenz 2021). On the other hand, those at the lower levels of the party hierarchy should be more focused on promoting the issues the party owns as they are more policy-oriented, are more motivated to follow the party line in other to advance in the party, and their communication efforts are mainly targeting the intra-party audience (Ennser-Jedenastik, Haselmayer, Huber & Fenz 2021) [1].

EP elections are generally considered second-order and thus might be considered less

[1] Though it is important to mention that previous research only found a difference between the two groups when it comes to responsiveness towards the public, and not in terms of addressing issue ownership (Ennser-Jedenastik, Haselmayer, Huber & Fenz 2021)

attractive for the top-tier of the party. But even in this case, there is a clear differentiation between those who are elected and those who are not. Candidates that hold viable positions are more likely to represent the top-tier of the party as they previously had national political experience (Aldrich 2018). As viable candidates are higher up in the party hierarchy, we would expect they are also more likely to be guided by office-seeking motivations. Furthermore, having secured an eligible spot liberates candidates from strictly following the issue emphasis of the party, as they are (at least for the following five years) less dependent on the party in terms of their future. Thus, in the context of EP election candidates, we would generally expect a clear differentiation in terms of issue profiles between those who hold viable and all other candidates from the same party. But yet again, the specific context of the Brexit process substantially diminished the office-seeking motivation of the candidates that hold viable positions. In addition, they are in a similar position to candidates that have low chances of getting elected in terms of their motivation to follow the party line, as their chances for any future career is dependent on the party leadership and local party organisation. Thus, we formulate the following hypothesis:

H2 There is less difference in terms of their issue profile between viable and not-viable candidates in the UK compared to other countries.

5 Method

The main data source for the analysis comes from the 'Political Campaigning on Twitter During the 2019 European Parliament Election Campaign' dataset (Stier et al. 2020). This contains the twitter communication for European Parliament candidates from all 28 member states, for parties which received more than 2 per cent of the national vote share in the 2019 EP election. Candidate twitter accounts were collected as part of the Euromanifesto Study in the two months leading up the election. The candidate tweets, as well as any public retweets, replies, or direct mentions of candidates were purchased directly from Twitter after the election, which helps ensure the completeness of the data as opposed to using the Twitter API. The data collection period captured tweets between 23 April to 30 May 2019 (Stier et al. 2020). The full dataset contains over 16,000,000 tweets, across 28 countries and 31 languages. Of this total, the final dataset contains 495,266 tweets by European Parliamentary candidates, across 30 languages.

Following this initial data collection stage, a team of 17 research assistants were hired to begin classifying a sample of the tweets across a range of features, covering 11 different languages. In the sampling process, tweets were first split between those made by candidates standing in the 2019 European Parliament elections, and tweets from the public. Candidate tweets were then grouped by the country each candidate was standing in, and by language, with the majority language in each country being coded. Lacking geo-location data for all remaining tweets, public tweets were simply grouped by language. Each coder was tasked with annotating 9,000 tweets in their assigned language. Due to our analytical focus on the campaigning strategies of political elites, the sample given to each coder was weighted so that 75% of tweets were by political candidates, with the remaining 25% from the public.

Student research assistants were hired who were either native speakers, or fluent in their assigned language and had lived in that country for a significant period of time. Due to a lack of applications in some languages, tweets in Portuguese, Swedish, and Dutch were coded by a single coder each. The other eight languages were coded in pairs, with the first 2,000 tweets being categorised by both coders, allowing for inter-coder reliability checks, and the remaining 7,000 tweets assigned to one coder each.

Table 1 outlines the language distribution for the full set of tweets, as well as the sampling breakdown for each language. For each country/language, a random sample of tweets were drawn from both the candidate and public tweets according to the 3:1 candidate to public ratio outlined above. In the case of Germany, Greece and Hungary, the number of tweets by political candidates was small enough that they could all be manually labelled. In these instances, the full set of candidate tweets were given to coders, with the remaining tweets randomly sampled from the public to make up the 9,000 total for each coder.

As the dataset contains retweets, (which under the coding scheme would appear identical to both the original tweet and other retweets) to avoid instances of coders labelling the same tweet multiple times, any duplicate tweets were removed from the sample, and were then merged back in after the manual coding process was complete.

Coders were provided with a codebook, outlining the classification process for each feature to be labelled in the tweet text. A batch of 150 randomly selected tweets in English were labelled by the researchers according to the codebook, to be used as a gold standard measure

Table 1: Tweet language distribution and sampling

		All tweets		Sampled tweets	
Country	**Language**	**Candidates**	**Public**	**Candidates**	**Public**
UK	English	131,332	5,113,760	13,500	4,500
France	French	62,403	2,911,611	13,500	4,500
Spain	Spanish	52,824	2,328,691	13,500	4,500
Italy	Italian	17,826	1,834,711	13,500	4,500
Poland	Polish	43,770	1,048,559	13,500	4,500
Netherlands	Dutch	13,793	433,309	7,500	2,750
Germany*	German	13,156	371,372	13,156	4,500
Greece*	Greek	4,349	72,301	4,349	32,000
Hungary*	Hungarian	326	2,118	326	2,118

*All candidate tweets were manually coded for these countries

as a set of initial training tweets. Coders were given a one hour training session, where the entire coding scheme was discussed in detail, and were shown how to use the survey software, as well as going through examples of coding tweets in practice.

Each coder was then given the set of 150 training tweets to label, and provided detailed feedback on their classifications of the tweets. After this, coders were given tweets in their assigned languages in batches of 500, and then 1,000, with detailed feedback provided for the first 2,000 tweets labelled by each coder. Coders were asked to label each tweet across a range of features, all of which are outlined in Appendix 2. If a tweet was labelled as mentioning a political issue, coders were then asked to specify which type of issue from a list of 10 potential options:

1. Economy
2. Environment
3. Immigration
4. Brexit
5. EU
6. Support for democratic values
7. Opposition to democratic values
8. Anti-elitism
9. Crime and justice
10. Other (transport, health, education)

Appendix 2 contains examples tweets in the same format presented to coders, as well as a detailed breakdown of all the features included in the surveys, as well as the descriptions provided to coders on how to classify each tweet.

5.1 Text classification

For the languages (and countries) where we do not reach complete human coding of candidate tweets we apply classification methods to extrapolate the quantities of interest to unlabelled tweets. For each language, we represent the hand coded tweet text as a document-feature matrix (dfm). For each tweet, we define the text as being all text included in a tweet if an original tweet (1) or the reply/comment and the original tweet (2) if it is a reply or retweet with comment. We carry out several pre-processing steps. Every specific user mention or specific hashtag is replaced by a generic user or hashtag token to avoid any over-fitting in the future. Next, as text pre-processing steps we remove language specific stopwords, we apply language specific stemming, remove numbers, punctuation and special symbols, and then create all uni- and bi-grams. In addition, we remove very rare features, those that appear in fewer than 0.25 percent of the documents in each language, usually indicating appearance in at least three tweets.

Thus, the frequency of the uni- and bi-grams across documents will be our main predictor matrix and we train binary classification models for each issue separately, where 1 marks the presence of that issue in the tweet, 0 otherwise.[2] This process results in a sparse matrix with a high number of features and thus we rely on machine learning algorithms to overcome the dimensional issue. More precisely, for each issue we train Extreme Gradient Boosting (xgBoost) models.[3]

Within a randomly selected training set (stratified based on each issue's prevalence) we carry out grid-like parameter selection using five-fold cross-validation and then evaluate the best performing parameter combinations in a final model using out-of-sample metrics[4] based on the (randomly selected) test set. We report conventional performance metrics based on the test set in Table 2.

As seen the the summary table, there is substantial variation in the classifier performances across issues and languages. Unsurprisingly, when we have limited human coded examples (table contains test set counts), our classifiers struggle to accurately or precisely predict issue

[2]The "otherwise" category in each case will contain all other tweets: personal, campaign level, or other political issues.

[3]These models have been compared to various flavors of regularized regression, but have constantly outperformed these. Furthermore, previous work by (Fazekas et al. 2021) also relied on xgBoost classifiers for tweet level political content categorization.

[4]These parameters will vary across languages and issues and our focus is on tree depth and learning rate resulting in the best performance.

content. In turn, this also means that classifier performance and ultimately predictive accuracy is directly related to issue salience, which can generate potential bias for the unlabelled data.

Table 2: Issue classifier performances

Issue		Test freq	Accuracy	Precision	Recall	F-score
Economy	en	187	0.93	0.47	0.74	0.57
	fr	535	0.92	0.69	0.73	0.71
	pl	175	0.87	0.41	0.59	0.48
Environment	en	247	0.93	0.59	0.78	0.67
	fr	868	0.90	0.80	0.74	0.77
	pl	88	0.88	0.24	0.67	0.36
Immigration	en	34	0.95	0.13	0.53	0.21
	fr	144	0.96	0.52	0.82	0.63
	pl	15	0.95	0.04	0.27	0.08
Brexit	en	689	0.87	0.71	0.78	0.74
EU	en	122	0.92	0.29	0.70	0.41
	fr	357	0.91	0.51	0.68	0.58
	pl	147	0.87	0.36	0.73	0.48
Anti-elitism	en	27	0.99	0.36	0.15	0.21
	fr	108	0.91	0.17	0.56	0.26
	pl	47	0.93	0.14	0.36	0.20
Support for dem val	en	153	0.93	0.38	0.64	0.48
	fr	223	0.92	0.36	0.57	0.44
	pl	101	0.84	0.19	0.56	0.28
Opposition to dem val	en	1			0.00	
	fr	144	0.92	0.22	0.47	0.30
	pl	24	0.87	0.04	0.33	0.07
Crime and justice	en	78	0.96	0.33	0.45	0.38
	fr	184	0.90	0.30	0.75	0.43
	pl	205	0.88	0.49	0.70	0.58

In Table 3 we report highly predictive words of specific issues. For those language $\times$ issue combinations where no words are listed the current classifiers are better at finding highly predictive terms for the 0 category, rather than the issue. These are also the cases where classifier performance is very weak. For the issues where the performance is acceptable, the terms reveal good face validity. Finally, using the same text preparation steps and the trained classifiers, we predict issue content for the unlabelled data.[5]

[5]As when training the classifier and evaluating them on the test set, we select the probability thresholds based on best performance in the training data. This means that we do not necessarily label an issue being present if the predicted probability for the tweet is above 0.5, rather we account for variation in baseline

Table 3: Predictive terms from issue classifiers

Issue		Features
Brexit	en	brexit; remain; leav; deal; referendum
Crime and justice	en	polic; crime; rape
	fr	justic; polici; polic; terrorist; illgal
	pl	pedofilii; pedofili; pedofilw; sprawi; art
Economy	en	economi; tax; invest; trade; econom
	fr	fiscal; conomiqu; financ; emploi; tax
	pl	mld; euro; mln; pienidz; inwestycj
EU	fr	l'union; lue; l'ue; lunion; lunion_europenn
	pl	ue; unii; europejskiej; unia; unijn
Environment	en	climat; planet; rebellion; environ; carbon
	fr	cologiqu; climat; climatiqu; lcologi; l'cologi
	pl	wgla; zwierzt; klimatu; rodowiska; energetyczna
Immigration	en	immigr
	fr	migrant; l'immigr; migratoir; immigr; rfugi
Support for dem val	en	democraci
	fr	dmocrati; dmocratiqu
	pl	demokracji; mniejszoci; rwno; demokracj; rwne

In light of the preliminary results from the classifiers, we close this section by listing future steps to extend our approach or ameliorate some of the potential issues:

- We will re-train classifiers for each issue only based on tweets that contain some political issue communication. This will be then used in conjunction with a separate classifier focusing on a broader categorization of the content: personal, campaign, or political issue.

- Incorporation of uncertainty around predicted values, analogous to approaches proposed by Fong & Tyler (2021) for covariates.

- Systematic evaluation of pre-processing choices.

- Comparison between results based only on human coded data and those based on mixed data, including simulations based on those languages where human coding has full coverage.

frequency for performance. We use the same thresholds when applying the prediction to the unlabelled data.

5.2 Data

Before addressing the main research question, it is worth taking the time to explore the wider cross-national patterns present within this newly collected dataset. Figure 1 below illustrates the breakdown of political issues discussed by MEP candidates from the 9 countries currently included in the coding process.

Figure 1: Issues mentioned in 2019 EP candidate twitter campaign communication

MEP candidate tweets that mention at least 1 political issue.
Country totals may exceed 100% as multiple issues can be included per tweet.

The data used here are taken from the random sample of human coded MEP tweets in each country. Each bar represents the proportion of candidate tweets that mention a specific issue, out of the total number of issue tweets in each country. One important note is that the small number of candidate tweets present for Hungary (n = 326), explains much of the divergence from the other 8 countries in terms of issue profiles. What is immediately clear at a first glance, is the significant degree of cross-national variation in terms of the types of issues discussed by candidates. Unsurprisingly, the most striking difference is the dominance of the Brexit issue in the UK, which despite the wider consequences for its European neighbours, seemed to barely feature at all in non-UK candidate online communication. A second point, that will be discussed further in the latter section of the paper, is the substantial number of issues included in the 'other issues' category. While this contains many issues that are common to all of the countries such as healthcare, education, and foreign affairs, this open ended category also allows for a more nuanced examination of any country-specific issues that arose during the campaigns. The following section disaggregates this data to explore how these patterns of communication presented themselves at both the candidate and party level, in order to answer our key research questions.

5.3 Analysis

The dependent variable regarding how closely candidate follow their party issue platform is measured at the candidate level, and is calculated as the ratio of tweets that mention at least one of their party's top 3 issues, divided by the total number of tweets that mention any political issue. One important note is that due to the dominance of the Brexit issue in the UK campaign in 2019, in order to make appropriate cross-national comparisons we exclude tweets where the EU or Brexit was mentioned in isolation. Tweets that are solely about the EU or Brexit are therefore removed from the denominator when calculating this variable, but tweets that mention the EU *and* another issue are included. This allows us to compare the issue profiles of candidates and parties beyond Brexit.

The baseline measure of party issue profiles used for this variable was taken from the Chapel Hill Expert Survey (Bakker et al. 2020), where the three issues with the highest issue salience scores (excluding EU integration) were selected for each party.

We include a dummy variable for whether candidates are from the UK or any other participating countries, as well as an indicator of a candidate's electoral viability as used by Stier et al. (2020), which is calculated using survey-based predictions of the number of seats each national party is expected to win, and then grouping candidate's into three categories ranging from 'safe', 'doubtful', or 'unpromising' chances of electoral success. Due to the small numbers of candidates in the 'safe' seat category, this was converted into a dummy variable, comparing those with the lowest chance of election, versus both 'safe' and 'doubtful' candidates.

At the party level we include controls for the degree of intra-party dissent on EU integration, which takes the form of an interval level variable from the Chapel Hill Expert Survey (Bakker et al. 2020), ranging from 0 - the party was completely united, to 10 - the party was extremely divided. We also include a dummy indicator for each party's degree of Euroscepticism, taken from PopuList (Rooduijn et al. 2019), with 1 coded as Eurosceptic, as well as the vote share each party gained at the 2019 election as a proxy for party size. Finally, at the candidate level we include the total number of tweets (logged) of each candidate over the full campaign period. Due to the interval level dependent variable and the nested structure of the data (candidates nested within parties), we make use of multi-level linear regression models for our analysis.

6 Results

Table 4 displays the results for the models examining the similarity of candidate and party issue profiles across twitter campaign communication. The unit of analysis here is at the candidate level, with candidates nested in political parties. The models below include all candidate tweets (human and machine labelled). Appendix 1 includes sensitivity checks, comparing models using only the human labelled tweets, against the full human/machine labelled data. We include random a random intercept in the multi level model at the party level, and random slopes for the lower level interaction term (electoral viability) in the final modes which contains the cross-level interaction. For the purposes of testing the expectations of H1, the results below use a dummy variable at the country level, comparing the UK baseline against the other 8 countries in the analysis. Additional models which include the full set of county fixed effects are included in appendix 1.

Table 4: Mentions of Party Top 3 Issues - Candidate level models

Dependent variable:
Ratio of candidate issue tweets (excluding tweets solely about EU)
that mention the party's top 3 issues

	(1)	(2)	(3)
Intercept	0.478***	0.496***	0.497***
	(0.023)	(0.054)	(0.054)
UK (dummy)	−0.069	−0.085	−0.094
	(0.065)	(0.060)	(0.063)
Viability (unpromising)		−0.016	−0.017
		(0.008)	(0.009)
Party vote share		−0.001	−0.001
		(0.002)	(0.002)
Total tweets (log)		0.005	0.005
		(0.003)	(0.003)
Party EU dissent		0.015	0.015
		(0.012)	(0.012)
Eurosceptic party		−0.144***	−0.144***
		(0.042)	(0.042)
UK*Unpromising			0.010
			(0.023)
AIC	−1726.515	−1675.182	−1663.699
BIC	−1705.373	−1627.743	−1600.447
LL	867.258	846.591	843.850
Politicians	1459	1438	1438
Parties	55	54	54
Var: party	0.025	0.021	0.021
Var: residual	0.016	0.015	0.015
Var: party unpromising			0.000
Cov: party intercept unpromising			−0.000

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

Model 1 in Table 4 includes only the UK dummy variable, while model 2 adds in the electoral viability measure and the additional control variables. Finally, model 3 includes the interaction between the UK country variable and electoral viability. Across all three models, the coefficients for the UK dummy variable are negative, but are not statistically significant at the $p<.05$ level. As such, we cannot reject the null hypothesis for H1, that on average, UK candidates are less likely to tweet about their party's top issues, compared to candidates in other EU countries. When it comes to electoral viability, models 2 and 3, which compare unpromising candidates against the reference category of both 'safe' and 'doubtful' candidates, the coefficients for this term were again negative but not statistically significant.

In relation to H2, the interaction term between the UK and electoral viability included in Model 3 also failed to reach statistical significance at the $p<.05$ level. As such, for H2 which expected that the differences in issue profiles between candidates with higher and lower likelihoods of being elected will be reduced in the UK versus other European countries, we cannot reject the null hypothesis.

7 Next steps

Despite the null findings from the initial analysis outlined above, several important steps remain to be completed in relation to both the data collection process and operationalisation of our key variables. More explicitly, our next steps include:

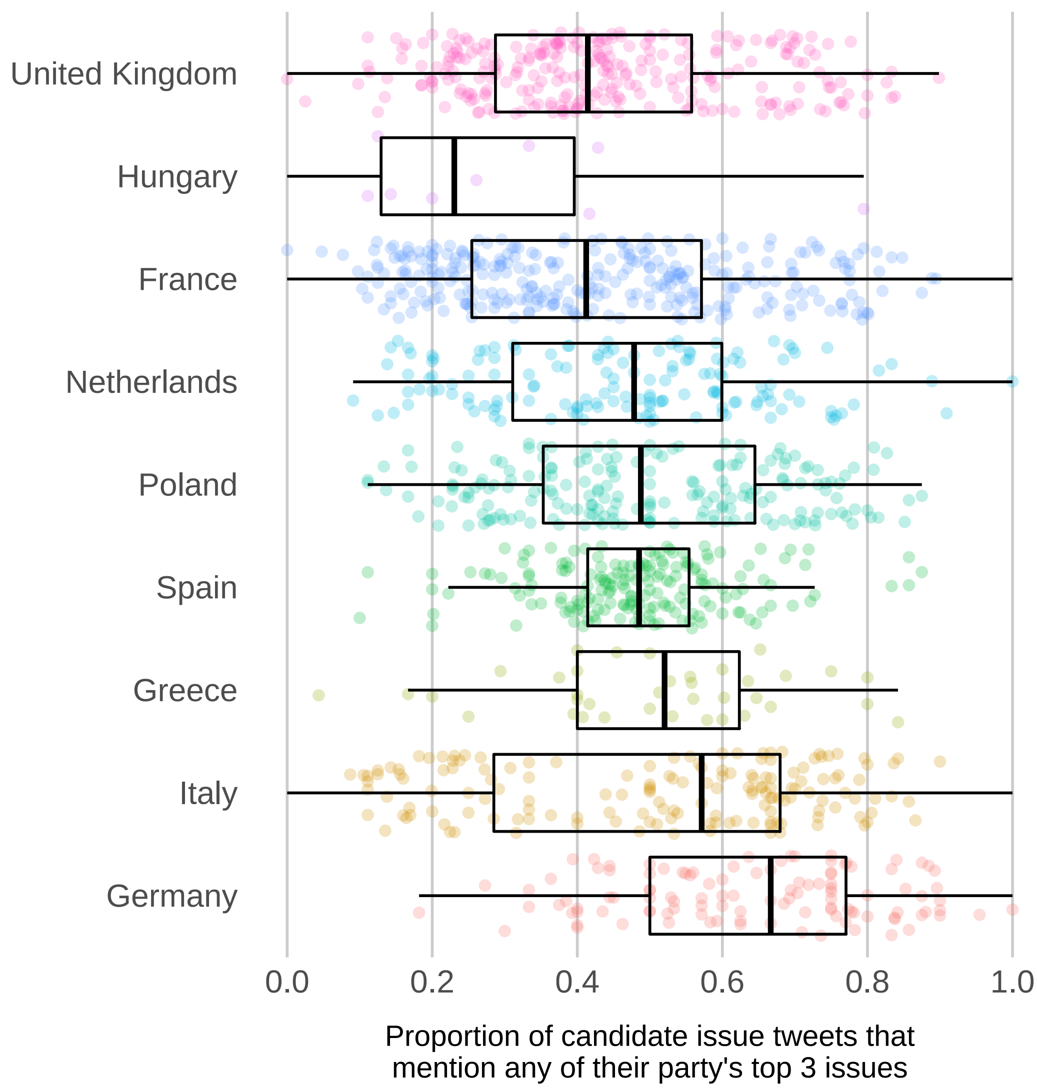
- Complete manual labelling for all 11 countries and extend comparison to all EU28
- Categorise open ended values coded under the 'other issue' category
- Measure party issue profiles based on official party twitter account and manifesto data
- Test if future political career trajectories, impact campaign strategies in the UK (i.e. 2014 vs. 2019 comparison)
- Compare parties and candidates with regards to responsiveness to the public.

The first two points in particular are likely to have a substantial impact on both the efficacy of the machine learning process, as well as the ability of our model to more accurately reflect the cross national patterns of candidate and party issue profiles. As illustrated by

Figure 2 below, and the direction of the coefficients in the full country fixed effects models in appendix 1, the majority of countries largely align with the the direction expected by H1. In Figure 2 we can see that with the exception of Hungary and France, all of the remaining countries candidates on average tended to tweet more about their party's core issues.

Figure 2: Candidate Tweets Mentioning Party's Top 3 Issues by Country

Dots represent values for each candidate, candidates with < 4 issue tweets excluded



This is also in light of the small number of tweets present for Hungary, as well as the fact that around 30% of all issue tweets came under the 'other issues' category (40% in France). Following these additional steps, as well as incorporating candidate social media data from 2014, will allow us to directly address the questions of what online political campaign communication looks like in situations when one of the central motivations for candidates to stand in an election is severely curbed.

References

Abou-Chadi, T. (2018), 'Electoral competition, political risks, and parties' responsiveness to voters' issue priorities', *Electoral Studies* **55**, 99–108.

Aldrich, A. S. (2018), 'National political parties and career paths to the european parliament', *JCMS: Journal of Common Market Studies* **56**(6), 1283–1304.

Bakker, R., Hooghe, L., Jolly, S., Marks, G., Polk, J., Rovny, J. & Steenbergen, Marco Vachudova, M. (2020), 2019 Chapel Hill Expert Survey, Version 2019.3, Technical report, Chapel Hill, NC: University of North Carolina, Chapel.
URL: *Available%0Aon chesdata.eu.*

Barberá, P. (2015), 'Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data', *Political analysis* **23**(1), 76–91.

Barberá, P., Casas, A., Nagler, J., Egan, P. J., Bonneau, R., Jost, J. T. & Tucker, J. A. (2019), 'Who leads? who follows? measuring issue attention and agenda setting by legislators and the mass public using social media data', *American Political Science Review* **113**(4), 883–901.

Binzer Hobolt, S. & Klemmensen, R. (2008), 'Government responsiveness and political competition in comparative perspective', *Comparative Political Studies* **41**(3), 309–337.

Callander, S. (2008), 'Political Motivations', *Review of Economic Studies* **75**(3), 671–697.

Ceron, A. (2017), 'Intra-party politics in 140 characters', *Party politics* **23**(1), 7–17.

Chadwick, A. (2017), *The hybrid media system: Politics and power*, Oxford University Press.

De Vries, C. E. & Hobolt, S. B. (2020), *Political Entrepreneurs*, Princeton University Press.

Ennser-Jedenastik, L., Gahn, C., Bodlos, A. & Haselmayer, M. (2021), 'Does social media enhance party responsiveness? how user engagement shapes parties' issue attention on facebook', *Party Politics* p. 1354068820985334.

Ennser-Jedenastik, L., Haselmayer, M., Huber, L. M. & Fenz, M. (2021), 'Who talks about what? issue strategies across the party hierarchy', *European Journal of Political Research* .

Fazekas, Z., Popa, S. A., Schmitt, H., Barberá, P. & Theocharis, Y. (2021), 'Elite-public interaction on twitter: Eu issue expansion in the campaign', *European Journal of Political Research* **60**(2), 376–396.

Fella, S., Uberoi, E. & Cracknell, R. (2019), European Parliament Elections 2019: Results and Analysis, Technical report.
URL: *https://commonslibrary.parliament.uk/research-briefings/cbp-8600/*

Fong, C. & Tyler, M. (2021), 'Machine learning predictions as regression covariates', *Political Analysis* **29**(4), 467–484.

Fowler, E. F., Franz, M. M., Martin, G. J., Peskowitz, Z. & Ridout, T. N. (2021), 'Political advertising online and offline', *American Political Science Review* **115**(1), 130–149.

Gilardi, F., Gessler, T., Kubli, M. & Müller, S. (2021), 'Social media and political agenda setting', *Political Communication* pp. 1–22.

Karlsen, R. & Enjolras, B. (2016), 'Styles of social media campaigning and influence in a hybrid political communication system: Linking candidate survey data with twitter data', *The International Journal of Press/Politics* **21**(3), 338–357.

Klüver, H. & Sagarzazu, I. (2016), 'Setting the agenda or responding to voters? political parties, voters and issue attention', *West European Politics* **39**(2), 380–398.

Kruschinski, S. & Haller, A. (2017), 'Restrictions on data-driven political micro-targeting in germany', *Internet Policy Review* **6**(4), 1–23.

Martill, B. (2020), 'The 2019 European Parliament election in the UK', *Italian Political Science Review* **50**(3), 368–381.

Munger, K., Egan, P. J., Nagler, J., Ronen, J. & Tucker, J. (2022), 'Political knowledge and misinformation in the era of social media: Evidence from the 2015 uk election', *British Journal of Political Science* **52**(1), 107–127.

Nulty, P., Theocharis, Y., Popa, S. A., Parnet, O. & Benoit, K. (2016), 'Social media and political communication in the 2014 elections to the european parliament', *Electoral studies* **44**, 429–444.

Peeters, J., Van Aelst, P. & Praet, S. (2021), 'Party ownership or individual specialization? a comparison of politicians' individual issue attention across three different agendas', *Party Politics* **27**(4), 692–703.

Popa, S. A., Fazekas, Z., Braun, D. & Leidecker-Sandmann, M.-M. (2020), 'Informing the public: How party communication builds opportunity structures', *Political Communication* **37**(3), 329–349.

Rooduijn, M., Van Kessel, S., Froio, C., Pirro, A., De Lange, S. Halikiopoulou, D., Lewis, P., Mudde, C. & Taggart, P. (2019), *The PopuList: An Overview of Populist, Far Right, Far Left and Eurosceptic Parties in Europe.*
URL: *www.popu-list.org*

Silva, B. C. & Proksch, S.-O. (2021), 'Politicians unleashed? political communication on twitter and in parliament in western europe', *Political Science Research and Methods* pp. 1–17.

Stier, S., Popa, S. A. & Braun, D. (2020), 'Political Campaigning on Twitter during the 2019 European Parliament Election Campaign', *GESIS Data Archive, https://doi.org/10.7802/1.1995* .

Vasilopoulou, S. (2020), 'Brexit and the 2019 EP Election in the UK', *Journal of Common Market Studies* **58**(S1), 80–90.

Wittman, D. (1983), 'Candidate Motivation : A Synthesis of Alternative Theories', *The American Political Science Review* **77**(1), 142–157.

Zuiderveen Borgesius, F., Möller, J., Kruikemeier, S., Ó Fathaigh, R., Irion, K., Dobber, T., Bodo, B. & de Vreese, C. H. (2018), 'Online political microtargeting: promises and threats for democracy', *Utrecht Law Review* **14**(1), 82–96.

Appendix 1: Sensitivity Check Candidate level Analysis

Table 5 - Models with all country fixed effects

Table 6 - Model comparison, human labelled vs human and machine labelled tweets

Table 5: Mentions of Party Top 3 Issues - All Country fixed effects

	(1)	(2)
Intercept	0.410^{***}	0.415^{***}
	(0.059)	(0.074)
Germany	0.217^{**}	0.206^{**}
	(0.084)	(0.076)
Greece	0.082	0.131
	(0.100)	(0.091)
France	0.019	−0.006
	(0.086)	(0.082)
Poland	0.073	0.163
	(0.097)	(0.092)
Spain	0.085	0.035
	(0.083)	(0.079)
Italy	0.062	0.117
	(0.091)	(0.084)
Netherlands	0.035	0.075
	(0.076)	(0.070)
Hungary	−0.069	−0.047
	(0.105)	(0.098)
Viability (unpromising)		-0.018^{*}
		(0.008)
Party vote share		−0.001
		(0.002)
Total tweets (log)		0.005
		(0.003)
Party EU dissent		0.017
		(0.013)
Eurosceptic party		-0.164^{***}
		(0.044)
AIC	−1779.851	−1730.525
BIC	−1721.986	−1646.583
LL	900.926	881.262
Politicians	1423	1403
Parties	55	54
Var: party	0.024	0.019
Var: residual	0.014	0.014

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

Table 6: Mentions of Party Top 3 Issues - Machine and human label sensitivity check

	HC+MC (1)	HC+MC (2)	HC+MC (3)	HC (4)	HC (5)	HC (6)
Intercept	0.478^{***}	0.496^{***}	0.497^{***}	0.525^{***}	0.586^{***}	0.584^{***}
	(0.023)	(0.054)	(0.054)	(0.027)	(0.069)	(0.069)
UK (dummy)	−0.069	−0.085	−0.094	−0.067	−0.083	−0.070
	(0.065)	(0.060)	(0.063)	(0.073)	(0.066)	(0.078)
Viability (unpromising)		−0.016	−0.017		−0.021	−0.018
		(0.008)	(0.009)		(0.014)	(0.015)
Party vote share		−0.001	−0.001		−0.002	−0.002
		(0.002)	(0.002)		(0.002)	(0.002)
Total tweets (log)		0.005	0.005		−0.000	−0.000
		(0.003)	(0.003)		(0.007)	(0.007)
Party EU dissent		0.015	0.015		0.019	0.019
		(0.012)	(0.012)		(0.013)	(0.013)
Eurosceptic party		-0.144^{***}	-0.144^{***}		-0.167^{***}	-0.168^{***}
		(0.042)	(0.042)		(0.046)	(0.046)
UK*Unpromising			0.010			−0.016
			(0.023)			(0.045)
AIC	−1726.515	−1675.182	−1663.699	−472.434	−433.231	−423.291
BIC	−1705.373	−1627.743	−1600.447	−453.197	−390.059	−365.729
LL	867.258	846.591	843.850	240.217	225.615	223.646
Politicians	1459	1438	1438	906	895	895
Parties	55	54	54	55	54	54
Var: party	0.025	0.021	0.021	0.030	0.023	0.025
Var: residual	0.016	0.015	0.015	0.029	0.030	0.030
Var: party unpromising			0.000			0.000
Cov: party intercept unpromising			−0.000			−0.002

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