**British Journal of
Political Science**

ARTICLE

Detecting Group Mentions in Political Rhetoric
A Supervised Learning Approach

Hauke Licht[1] and Ronja Sczepanski[2]

[1]Department of Political Science and Digital Science Center, University of Innsbruck, Innsbruck, Austria and [2]Centre for European Studies and Comparative Research, Sciences Po Paris, Paris, France
Corresponding author: Hauke Licht; Email: hauke.licht@uibk.ac.at

Abstract
Politicians appeal to social groups to court their electoral support. However, quantifying which groups politicians refer to, claim to represent, or address in their public communication presents researchers with challenges. We propose a supervised learning approach for extracting group mentions from political texts. We first collect human annotations to determine the passages of a text that refer to social groups. We then fine-tune a transformer language model for contextualized supervised classification at the word level. Applied to unlabeled texts, our approach enables researchers to automatically detect and extract word spans that contain group mentions. We illustrate our approach in two applications, generating new empirical insights into how British parties use social groups in their rhetoric. Our method allows for detecting and extracting mentions of social groups from various sources of texts, creating new possibilities for empirical research in political science.

Keywords: social groups; political rhetoric; computational text analysis; supervised classification

Introduction

The struggle of social groups to influence political processes and outcomes shapes politics around the world. Understanding the role of social groups in politics is thus a central theme in many fields of political science research, ranging from democratic representation research over political sociology to conflict studies. It is thus not surprising that the extant political science literature offers many hypotheses about how and why politicians relate themselves to social groups or talk about them in their public communication (for example, Chandra 2012; Huber 2021; Kitschelt 2000; Lieberman and Miller 2021; Saward 2006; Stückelberger and Tresch 2022; Thau 2019). However, quantitatively studying this facet of politics is currently limited by a lack of scalable measurement instruments allowing researchers to quantify group-based political rhetoric.

   This paper proposes a supervised text classification strategy for extracting social group mentions from large political text corpora. The first step is to define what constitutes a social group. For example, in the applications we present in this paper, we define social groups as collectives of people that share common attributes, such as economic circumstances, but also common values. Next, we tasked human coders with marking all passages in a sample of sentences that mention social groups. This second step results in a set of labeled sentences in which a varying number of words are marked as containing mentions of groups. We then use these annotations to fine-tune a transformer-based supervised token classifier. The classifier learns to predict whether

or not a word in a sentence belongs to a social group mention while accounting for the word's surrounding sentence context. The resulting classifier automates our manual word-level annotation procedure and enables reliable detection of group mentions in unlabeled texts.

We demonstrate the reliability, validity, and flexibility of our method in analyses of British parties' group-based rhetoric. Our approach proves very reliable in detecting mentions of social groups – even in social group references not contained in the training data or when transferred to German party manifestos and British parliamentary speeches. Our evidence further suggests that our approach is more reliable than dictionary-based mention detection. Our evidence also underscores the validity of our approach. The document-level indicators of the social group emphasis in parties' manifestos we obtain with our approach correlate strongly positively with comparable indicators obtained through manual content analysis by Thau (2019).

We illustrate the added value of our method in two applications. First, we study differences in the social group focus of British parties. Regarding the salience of social group mentions across policy topics, we find that both the Labour Party and the Conservative Party tend to emphasize social groups more when they discuss (re)distributive policy issues compared to regulatory policy issues. However, we find that this tendency is more pronounced for Labour. Further, we apply an inductive feature extraction method (Monroe et al. 2008) to the group mentions extracted by our classifier to reveal differences in the words and phrases that distinguish British parties' social group mentions. This analysis shows that parties do not only focus on different social groups but also use different terms to refer to these groups and demonstrates that a main advantage of our method lies in its ability to locate and extract verbatim group mentions from large text corpora. Second, we apply our method to study the relationship between group mentions and emotional rhetoric in British parties' manifestos. We show that sentences mentioning social groups are more emotional in tone than sentences without such mentions, suggesting that these two rhetorical strategies tend to be linked in parties' campaign communication.

Our findings and applications demonstrate that our method equips researchers with new flexibility in their analyses of social groups' roles in political rhetoric. At present, the quantitative study of group appeals is limited to a community of highly dedicated researchers endowed with significant resources. Our method opens new possibilities for expanding this literature, for example, by complementing existing studies that focus on how voters respond to group-based political rhetoric (Hersh and Schaffner 2013; Holman et al. 2015; Robison et al. 2021; Weber and Thornton 2012) with new studies examining whether and how politicians use these as part of their electoral strategies (for example, Stückelberger and Tresch 2022; Thau 2021). Moreover, our method may facilitate the broader adaption of measures of group-based political rhetoric in related fields that investigate party-voter linkages, including work on political representation, issue competition, party branding, party types, and affective polarization. For example, because our approach allows us to locate where social groups are mentioned in a text, researchers can study differences in how politicians talk about specific target groups (for example, refugees, women, the unemployed, ethnic minorities, etc.).

Social Groups in Political Rhetoric

Social groups are at the heart of political science theory. Politicians have many reasons to emphasize social groups by directly referring to them in their public communication (Conover 1988; Miller et al. 1991). Talking more or less about social groups allows parties and their representatives to show which groups are important to them and which are not (Conover 1988; Dolinsky et al. 2023; Gadjanova 2015; Horn et al. 2021; Howe et al. 2022; Nteta and Schaffner 2013; Stückelberger and Tresch 2022; Thau 2019). Mentioning a social group frequently can be a way to signal responsiveness to it and make its members 'feel seen' and represented in politics (Pitkin 1967; Robison et al. 2021; Saward 2006). Further, emphasizing social groups in their public

communication can allow politicians to mobilize groups' sentiments, identities, and grievances (Goodman and Bagg 2022; Miller et al. 1991; Stückelberger and Tresch 2022).

But group-based rhetoric is also about shaping groups' opinions, interests, and perceptions (Goodman and Bagg 2022; Miller et al. 1991; Stückelberger and Tresch 2022). For example, how elites talk about social groups can affect how positively or negatively these groups are viewed by others – often with consequences for how deserving these groups are perceived to be by the public (O'Grady 2022; Slothuus 2007). Thus, political parties and their representatives can shape groups' standing in society. Moreover, research has shown that connecting groups to an issue position can alter their opinion on the topic (Huber et al. 2024). Therefore, which groups politicians appeal to can also affect how citizens perceive their political and social world.

While it is thus of central interest to political scientists to understand when, why, and how politicians mention social groups, scholars tend to disagree on how to conceptualize a social group. Some limit their conception of a social group to include only collectives of people who share socio-economic circumstances or socio-demographic characteristics (Dolinsky et al. 2023; Huber 2021) that provide a source of identification for group members (Miller et al. 1991). Others, like Howe et al. (2022), advocate for a more open conception, arguing from a constructivist perspective that a social group can be any collective of people who share some attribute, including common values and life experiences (cf. Chandra 2012; Wolkenstein and Wratil 2021). For example, attributes like 'hard-working' and 'moral righteousness' can be central to people's conceptions of their in- and out-groups (Sczepanski 2024; Zollinger 2022). And even groups that are objectively based on socio-structural attributes, such as their place of residence, often place cultural, not socio-structural, factors at the centre of their in-group conceptions, such as specific values or a certain way of life (Zollinger 2024). These differences in conceptualizations have important implications. The socio-economic definition focuses on boundary drawing in line with the distribution of material resources and 'objective' demographic characteristics. By contrast, more abstract group references also focus on symbolic, discursively constructed boundaries such as 'honest people' (Lamont and Molnár 2002; Mierke-Zatwarnicki 2023).

In this study, we opt for the broader and more inclusive conceptualization. Our goal is to detect references to social group categories in political speech and text. Thus, we cannot apply group members' identification as a criterion. More importantly, even symbolic boundaries can turn into social boundaries and eventually political cleavages if they are politicized (cf. Enyedi 2005). By capturing references to all social categories that might turn into meaningful social and political boundaries, we thus account for politicians' agency in the social construction of groups.

Yet, regardless of whether researchers opt for a more narrow or broad definition of a social group, quantitative studies of political elites' group-based rhetoric are still relatively rare. A lot of research has focused on citizens' perceptions of group appeals and their feelings of being represented as a group (Holman et al. 2015; Jackson 2011; Kam et al. 2017; Robison et al. 2021; Valenzuela and Michelson 2016; White 2007). By contrast, research on the 'supply' of group-based rhetoric is currently largely limited to a handful of studies in the party politics literature (for example, Dolinsky 2022; Horn et al. 2021; Howe et al. 2022; Huber 2021; Stückelberger and Tresch 2022; Thau 2019, 2021) and research on ethnic politics (for example, Lieberman and Miller 2021; Nteta and Schaffner 2013). We attribute this to a central empirical challenge in studying social groups in political speech and text: detecting them in large amounts of texts and across contexts.

Detecting Mentions of Social Groups in Political Texts

We argue that one of the main reasons comparative research on political actors' use of group-based rhetoric is limited in scope lies in the methodological challenges researchers confront when trying to detect and extract social group mentions in large political text corpora. As outlined next, these challenges are largely due to social group mentions' linguistic characteristics. These characteristics, in turn, limit the reliability and scalability of existing content-analytic
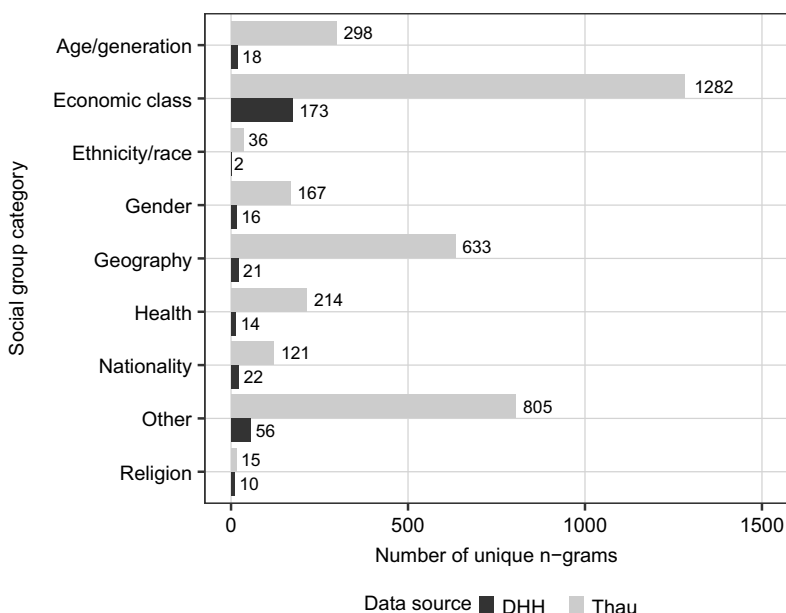
Figure 1. Unique n-grams in human-annotated data collected by Thau (2019) and in the Dolinsky-Huber-Horne (DHH) dictionary compiled by Dolinsky et al. (2023) by social group category.

measurement approaches. We introduce a supervised token classification approach to group mention detection that overcomes these challenges.

Characteristics of Group Mentions in Political Texts

One of the central methodological challenges in identifying mentions of social groups in political text and speech is that they are linguistically extremely diverse. First, the number of social groups that can be referred to in a given political context is typically large. The list is already long if one considers only groups that are defined based on socio-demographic characteristics such as age or generation, gender, race, or ethnicity (cf. Chandra 2012). And if one considers that objective membership in different group categories is often nested and intersectional, the list grows further. For example, a mention of 'people living and working in rural areas' refers to members of the rural population who are workers. As a case in point, Thau (2019) conducted a manual content analysis of group appeals in British party manifestos and identified more than 2.7 thousand unique ways in which the Conservative and Labour parties referred to economically or socio-demographically defined groups (see Figure 1 and Table F2).

Second, political actors do not only refer to groups using socio-demographic markers but also discursively construct groups by emphasizing people's shared values, norms, circumstances, and commonalities in other attributes. For example, phrases like 'the needy in our country' and 'the wretched of the world' (see Table 1), 'those with the broadest shoulders,' or 'those who work hard and do the right thing' do not refer to clearly circumscribed socio-demographic groups, but they likely still appeal strongly to people with corresponding self-conceptions and identities (Bornschier et al. 2021). Drawing again on the data collected by Thau (2019), we argue that this phenomenon should not be neglected. Thirty-one per cent of social group appeals in his data were assigned to the 'other' social group category as the mentioned groups did not fit into any of his economic or socio-demographic group categories (see Table F2).

Table 1. Examples of group mentions in sentences drawn from British mainstream party manifestos: the highlighted text spans the identified groups mentioned in each sentence

We seek to bring about a fundamental change in the balance of power and wealth in favour of *working people* and *their families*.

Eight years of meanness towards the needy in our country and towards *the wretched of the world*.

The welfare of *the old, the sick, the handicapped* and *the deprived* has also suffered under Labour.

Labour recognises the special needs of *people who live and work in rural areas*.

A third reason why social group mentions in political texts are linguistically extremely varied is that for any given social group, there are various lexically different ways to refer to it. For one, there are many indirect ways to refer to a group. For example, the phrases 'the unemployed' and 'those out of work' refer to the same social group. For another, many references to groups use descriptive language, such as 'the first generation to know we are destroying the environment, and the last generation with a chance to do something about it before it is too late'.

Established Methods and their Limitations

The linguistic diversity of social group mentions in political rhetoric has two important methodological implications. First, as illustrated in Table 1, the phrases used to mention, refer to, or address social groups in political text often span multiple words. Second, any sentence can mention no, one, or several social groups. Consequently, reliable detection and extraction of social group mentions require identifying the words used to refer to or describe social groups in a text while not knowing *a priori* how many unique mentions it contains, where the mentions are located in the text, and how many words a given mention spans.

To cope with these challenges, researchers studying groups-based rhetoric based on political text currently have two options: manual content analysis and automated dictionary measurement. These two approaches are well-established in the applications to sentence- and document-level classification (cf. Barberá et al. 2021; Quinn et al. 2010). However, both approaches have clear limitations when applied to extract group mentions from large text corpora.

Manual content analysis identifies group mentions in political texts by tasking coders to locate and extract the relevant text segments referring to groups (for example, Huber 2021; Stückelberger and Tresch 2022; Thau 2019, 2021) or by indicating this information at the sentence level (Hopkins et al. 2024; Horn et al. 2021). As in other applications (cf. Grimmer and Stewart 2013; Quinn et al. 2010), this approach can be considered the most valid compared to semi- or fully automated methods. Human coders can read and interpret texts, allowing them to spot simple group mentions but also more complex ones, like the abstract or descriptive multi-word examples included in Table 1 above.

However, manual content analysis is relatively costly (but see Benoit et al. 2016). Researchers need to hire annotators. Moreover, collecting manual annotations is time-consuming for large corpora.[1] Consequently, studies that have applied manual content analysis to study group-related rhetoric use either text corpora of limited size, focus on a small set of political parties, and/or limited periods.

The dictionary approach is more resource-efficient as it enables detection mentions of predefined groups automatically by searching for matches to a list of group keywords (cf. Dolinsky et al. 2023). The only required human input to dictionary-based measurement is to define a list of keywords that reflect the potential ways the social group(s) of interest are mentioned in a corpus.

---

[1]The alternative to mention-level annotation is to task coders to only indicate whether or not a sentence contains one or more references to a group (category). This makes annotation more time-efficient. But applying this approach, researchers also miss the opportunity to record group mentions' exact wording, limiting their ability to gain more detailed knowledge, for example, about how exactly politicians appeal to groups.

However, considering how linguistically varied social group references are in political texts, we should expect that compiling a comprehensive list of relevant keywords will be very challenging in many applications, especially since group mentions usually span multiple words, are often indirect, and potentially discursively invoke groups in abstract ways. For example, a dictionary might contain the keyword 'the unemployed' but fail to recognize semantically similar phrases like 'those out of work'.[2] Figure 1 underscores this argument, showing across different social group categories how the number of keywords and keyword patterns in a dictionary compiled by Dolinsky et al. (2023) for detecting social group mentions in British party manifestos compares the number of unique mentions Thau's coders have identified. This shows that even experts in group appeals research who have employed an iterative strategy to identify relevant keywords and patterns arrive at much shorter lists of phrases than is possible through direct human annotation of the target corpus.

In the supplementary materials, we present analyses that support our argument and justify our concerns. First, we apply the dictionary from Dolinsky et al. (2023) to our and Thau's human-annotated texts,[3] finding modest precision but poor recall at both mention and sentence levels (see Tables G1 and G2). Additionally, our analyses suggest that semi-automated dictionary expansion techniques are not a simple solution. For instance, when using a pretrained word embedding model to find relevant keywords (cf. King et al. 2017), many multi-word phrases are missing from the model's vocabulary. We estimate that considering the top $k = 10$ most similar words for each 'seed' keyword would require reviewing 1,412 words and phrases (see Table G3). Furthermore, skipping human review in dictionary expansion (cf. Osnabrügge et al. 2021b) by adding all k most similar words do not improve reliability (see Table G4), as it increases recall but reduces precision (see Figure G3).

To summarize, manual content analysis allows valid measurement of social group mentions in political texts but is resource-intensive, and, when adopting a sentence-level classification approach, it means discarding empirically interesting variation. By contrast, dictionary-based measurement promises resource efficiency, but it limits reliable detection demonstrably, likely especially so for groups without clear-cut membership criteria and groups that can be referred to in many lexically different ways.

A Supervised Token Classification Approach

We propose a method that allows researchers to automatically identify and extract mentions of groups in political texts with a limited manual labeling effort. Our method applies supervised learning to detect and extract mentions of social groups in political texts. It strikes a favourable balance between the objectives of reliable and valid detection on the one hand and scalability on the other.

After theoretically defining the concept, the first step of our supervised learning approach is to task human coders to highlight all mentions of social groups in a set of sentences sampled from a target corpus. This step mirrors the procedures adopted in existing manual content analysis studies. However, what distinguishes our approach is that we preserve the verbatim mentions of groups where and how they occur in texts.[4] The first row in Figure 2 illustrates what the annotations we collect look like. By tasking coders with highlighting all group mentions in a

---

[2]We note that searching for relevant keywords iteratively (cf. Dolinsky et al. 2023; Muddiman et al. 2019) risks over-fitting to the subset of the corpus the researcher has reviewed to build their dictionary and thus limit generalization. Including only indicator words (here, for example, 'those' and 'work') would lead to many false-positive classifications. Checking for co-occurrences of such words in documents (for example, 'those' + 'work', 'those' + ...) could partially remedy this concern. However, the number of keywords that require inclusion increases rapidly with the length of relevant expressions, while increasing the number of keywords in a dictionary often reduces precision due to polysemy.

[3]Lena Huber, a co-author of Dolinsky et al. (2023), shared their dictionary for UK party manifestos with us.

[4]This contrasts with Thau (2019), for example, who has 'cleaned' and 'harmonized' (i.e., post-processed) the real-occurring mentions before recording them in a spreadsheet without their original sentence context.
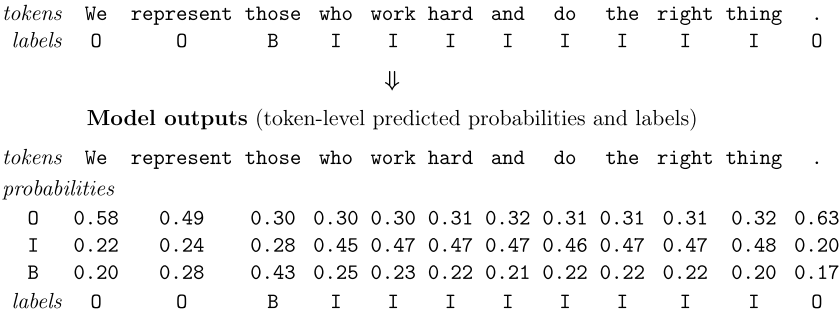
```
              Input data (text with group mention annotations)
        "We represent those who work hard and do the right thing."

                                    ⇓

        Model inputs (tokenized text and corresponding token-level labels)
 tokens   We  represent those who work hard and  do  the right thing  .
 labels   O      O       B    I    I    I    I    I   I    I     I    O

                                    ⇓

        Model outputs (token-level predicted probabilities and labels)
 tokens      We  represent those who work hard and  do  the right thing  .
 probabilities
    O      0.58    0.49     0.30 0.30 0.30 0.31 0.32 0.31 0.31 0.31  0.32 0.63
    I      0.22    0.24     0.28 0.45 0.47 0.47 0.47 0.46 0.47 0.47  0.48 0.20
    B      0.20    0.28     0.43 0.25 0.23 0.22 0.21 0.22 0.22 0.22  0.20 0.17
 labels    O        O        B    I    I    I    I    I    I    I     I    O

                                    ⇓

              Post-processed outputs (extracted group mentions)
    span               "those who work hard and do the right thing"
```

Figure 2. From sentence annotation to extracted mention. Highlighted spans are converted into token-level labels. Labels 'B' and 'I' indicate tokens that are at the 'beginning' or 'inside', the 'O' those outside of a group mention. The token classifier predicts label probabilities, which indicate a token's most likely label. Predicted mentions can be determined from token-level predicted labels.

sentence, we can determine the characters that belong to individual group mentions. This means that in each labeled sentence, no, one, or several spans of characters might be marked as mentioning a group (see Table 1 for examples).

In the second step, we use this information as data for supervised learning. Specifically, we train a supervised classifier for token classification. Token classification means to assign each word in a sentence a single label from some predefined categories. Enabling this requires converting the annotations into word-level labels. This is illustrated in the second panel of Figure 2. From the annotations we have collected in the first step, we know for each group mentioned in a sentence at which character it starts and ends. Tokenizing the sentence into words, we can determine for each word in the sentence whether or not it belongs to a mention of a group. Further, for words that belong to such a mention, we can determine whether the word is at the beginning of the mention or inside of it. As shown in the second row of Table 2, words that do not belong to a mention are labeled 'O' to indicate that they are *outside* of a social group mention. By contrast, words at the beginning or inside of a mention are labeled 'B' respectively 'I' (cf. Ramshaw and Marcus 1995).

With word-level labels at hand, the supervised token classification task is to predict each word's label in a sentence. Provided with multiple labeled sentences in this format, we fine-tune a transformer-based neural network for this task. This approach is commonly applied in named entity recognition, and it has already been adopted for event data extraction (Skorupa Parolin et al. 2022) and the section of references to the people and the elite in German parliamentary speeches (Klamm et al. 2023). Relying on a pretrained transformer-based model like BERT (Devlin et al. 2019), RoBERTa (Liu et al. 2021), or DeBERTa (He et al. 2021) for this task allows accounting for words' sentence context when learning to predict their labels. This is impossible with standard bag-of-words methods (cf. Timoneda and Vallejo Vera 2025).

The result of this second step is a fine-tuned token classification model that can be applied to detect and extract mentions of social groups in political texts. As shown in the third panel in

Figure 2, the label class that receives the highest predicted probability for a word is treated as its predicted label. And, as shown in the last panel of Figure 2, this classifier output can be parsed to extract the words belonging to the (predicted) group mention(s) in a sentence.

In the third step, the fine-tuned supervised token classifier can be applied to unlabeled texts to identify and extract mentions of social groups that have not been in the training data. This enables automated labeling and extraction of group mentions in large text corpora.

Our proposed method contrasts with established approaches to quantifying group-based rhetoric in political texts in three ways. First, it contrasts with dictionary-based measurement in that we presume that recognizing concrete group mentions in a text is more reliable than selecting indicative words or phrases *a priori*. Second, in contrast to the manual content analysis approach, we leverage the benefits of automation through supervised learning. This saves researchers the time and costs associated with manual content analysis (cf. Barberá et al. 2021; Grimmer and Stewart 2013). Third, in contrast to sentence-level classification approaches, we annotate, model, and predict the text passages that represent group mentions at the word level. Consequently, our approach preserves the lexical diversity and linguistic variability of group mentions as they occur in political texts, which will enable more detailed analyses of group-centered political rhetoric.

Evaluation and validation

To evaluate and validate our method, we first focus on detecting and extracting social groups mentioned in British parties' election manifestos. In section 'Transfer to Other Parties, Domains, and Countries', we then extend our focus and present additional analyses of social group mentions in parliamentary questions in the UK House of Commons as well as in German parties' manifestos.

Reliability: Evaluation in British Party Manifestos

We first focus on detecting and extracting social groups mentioned in British parties' election manifestos. Our case selection is motivated by substantive as well as methodological considerations. From a substantive perspective, we are interested in comparing parties' social group mentions across elections and parties, for example, to study what distinguishes the groups mentioned by parties with different ideological profiles and programmatic platforms. From a methodological point of view, studying cases that have already been studied in parts in the influential work by Thau (2019) allows us to assess whether the measurements we obtain with our supervised learning method align with those obtained through manual content analysis and, in turn, allows us to assess the validity of our approach.

Table 2. Summary of test set performances of DeBERTa group mention detection classifiers fine-tuned and evaluated on our corpus of labeled UK manifesto sentences. Values (in brackets) report the average (90 per cent quantile range) of performances of 25 different classifiers fine-tuned in a 5-times repeated 5-fold cross-validation scheme. Columns distinguish between different evaluation schemes (i.e., different ways to compute the eval. metrics)

	Mention level			
	seqeval	cross-span avg.	within-sentence avg.	Sentence level
F1	0.83 [0.77, 0.90]	0.87 [0.81, 0.92]	0.88 [0.79, 0.94]	0.95 [0.92, 0.97]
Precision	0.81 [0.73, 0.89]	0.88 [0.81, 0.92]	0.89 [0.79, 0.94]	0.95 [0.90, 0.97]
Recall	0.85 [0.78, 0.91]	0.88 [0.82, 0.92]	0.89 [0.81, 0.94]	0.96 [0.93, 0.98]

Note: seqeval is the strict metric proposed by Ramshaw and Marcus (1995) and implemented by Nakayama (2018).

Data and methods

Our data set records forty-six electoral manifestos from the two largest British parties – the Labour Party and the Conservative Party – from the elections of 1964 to 2019 and the manifestos of the Democratic Unionist Party (DUP), the Green Party of England and Wales (Greens), the Liberal Democrats (LibDem), the Scottish National Party (SNP), and the United Kingdom Independence Party (UKIP) for the elections in 2015, 2017, and 2019. We have split the raw texts of the manifestos into sentences (see Table 2) and sampled 8,596 sentences from this corpus for annotation, stratifying by party and (election) year, and, where possible, by the manifesto chapter (see Table B1).[5]

To collect annotations of social group mentions in these documents, we have designed a custom coding scheme. The focal category of our coding scheme is the 'social group' category. In our application, we define a social group as a collective of people with one or more common characteristics. As discussed in Section 2, we deliberately adopt a broad conceptualization. In addition, we include four other categories in our coding scheme ('political group', 'political institution', 'organization etc.', and 'implicit social group reference', see Table B3 in the Supplementary Material), and an 'unsure' category.[6] We included these additional categories for three reasons. First, when developing the coding scheme, we found that additional categories helped our annotators recognize the conceptual boundaries of the 'social group' category. Second, collecting annotations for these categories allows us to demonstrate that our method is similarly reliable in detecting other types of groups. Third, we wanted our data to be as reusable as possible for other researchers.

We have collected annotations from two trained research assistants using the doccano online annotation tool (Nakayama et al. 2018). As shown in Table B2 in the Supplementary Material, we have collected annotations from both coders for more than 30 per cent of sentences because it is a well-known limitation of content-analytic annotation procedures like ours that individual coders can make mistakes or some text passages might be ambiguous (cf. Krippendorff 2004).[7] As shown in Table B5, the intercoder agreement is very high in our sample of doubly annotated sentences. The median (mean) sentence-level agreement in sentences with at least one social group annotation by either coder is 95.7 per cent (90.8 per cent) and 95.2 per cent (91.5 per cent) in sentences without any social group annotation but at least one other group annotation. This indicates that our coding instrument and procedure indeed elicit highly reliable annotations. Moreover, analyzing the sentences with disagreements, we find that in a sizeable number of sentences (24–45 per cent), our coders' disagreements stem from mismatches in the exact beginning, end, or beginning and end of individual group mentions (see Table B6).

Because we have collected annotations from two coders for some sentences, we need to aggregate these annotations into a single set of word-level labels per sentence. As described in Supplementary Material B.1, we follow the rich computer science literature on annotation aggregation (cf. Chatterjee et al. 2019) and fit a Bayesian sequence combination model (Simpson and Gurevych 2019). This results in word-level labels for all 8,576 human-annotated sentences in our annotated British manifesto sentences.

To prepare the labeled data, we first removed all 'unsure' annotations so that the corresponding words are treated as if they are not part of any type of group mention.[8] We have then converted sentences' word-level labels into the IOB2 (inside–outside–beginning) label scheme (Ramshaw and Marcus 1995). This means that tokens at the beginning of a mention receive a special label. In

---

[5]This sampling strategy ensures that data from all election years and parties are represented equally in our training data. Stratifying by manifesto chapter, moreover, enhances the topical coverage of our labeled dataset.

[6]We acknowledge that the background of us researchers as well as the ones of the research assistants might lead to biases in the conceptualization and coding of social groups. Please refer to Supplementary Material I for the positionality statement.

[7]See Supplementary Material C for examples of ambiguity during coding and how the ambiguity was resolved with the coders.

[8]Examples of such annotations are 'organised youth activities' in the sentence 'To ensure there are more things for teenagers to do we will double the availability of organised youth activities on Friday and Saturday nights' and 'foreign lorries' in the sentence 'We will charge foreign lorries for the use of British roads with our Brit Disc scheme.'

particular, we distinguish between tokens at the beginning of social group mentions (B-SG) and tokens inside them (I-SG). Together with the 'outside' (O) label reserved for tokens outside of a mention, this results in eleven label classes.

We have used the resulting labeled sentences to fine-tune DeBERTa and RoBERTa models (He et al. 2021; Liu et al. 2021) for token classification and report the result of the fine-tuned DeBERTa model if not stated otherwise.[9]

Results

To assess the reliability of our approach in detecting social group mentions in held-out sentences, we compare token classifiers' predicted labels against the labels we obtained from our coders' annotations.[10] In Table 2, we report the results of 5-times-repeated 5-fold cross-validations of DeBERTa token classifiers fine-tuned on labeled sentences in our UK party manifesto corpus.[11] Cross-validation allows us to summarize the results of twenty-five different classifiers fine-tuned on different data splits to present robust estimates of classifiers' out-of-sample performance.[12]

Focusing on classifiers' reliability in detecting social group mentions,[13] we first turn to their average mention-level performance (column 'cross-span avg.'). We compute mention-level recall, precision, and the F1 score estimates by comparing predicted to 'true' word-level labels within observed and predicted group mentions and averaging these estimates across social group mentions in the test set.[14] Looking at classifiers' performance at the mention level, they correctly classify on average 87 per cent of words that belong to social group mentions in the human-labeled data (recall). Conversely, our classifiers are correct 88 per cent of the time when they predict that a word belongs to a social group mention (precision). This amounts to an average mention-level F1 score of 87 per cent.

This high level of reliability in detecting social group mentions in held-out texts translates into very reliable classification at the sentence level. To compute sentence-level performance from word-level predictions, we determine for each group category in our coding scheme whether there is at least one annotation in the 'true' respectively predicted labels and compare them within sentences. We then count a sentence as correctly classified if at least one word was labeled correctly for the given group type. According to this standard, our classifiers correctly classify on average 96 per cent of sentences that contain at least one social group mention (recall). In expectation, this amounts to only four misclassifications per 100 sentences that contain one or more social group mentions.

Table 2 also reports the so-called seqeval metric, which considers a classifier's predictions at the mention level only correct if it predicts the correct label for *every* word in a given human-labeled mention. Instances where the classifier's prediction begins too late or early, ends too early or late, etc., are considered classification errors (see Supplementary Material D). Even according to this rather strict standard, our classifiers correctly predict 85 per cent of social group mentions (recall), 82 per cent of the social group mentions they predict are correct (precision), and this amounts to an average F1 score of 0.83. We note, however, that based on our review of our coders'

---

[9]We have run an experiment to compare the development set performance when fine-tuning four pretrained models with varying hyper-parameters: BERT (base), DistillBERT, RoBERTa (base), and DeBERTa v3 (base). DeBERTa performed best, followed by RoBERTa (see Table E1).

[10]Supplementary Material D introduces quantitative evaluation in token classification applications.

[11]We have iterated over five random seeds to control the initial train/test split and then iterated in a 5-fold splitting over the training data to train five different classifiers per random seed on different train/development splits.

[12]Note that we have grouped by manifesto when splitting the data to prevent data leakage and increase the ecological validity of our analysis. This means that all the labeled sentences in a manifesto are either in the training, validation, or test sets. Depending on the random seed, this approach resulted in training sets with 6,108 to 6,245 labeled sentences, validation sets with 809 to 896 labeled sentences, and test sets with 1,480 to 1,574 labeled sentences.

[13]Table 5 reports the results for all group categories.

[14]We illustrate and explain our mention-level evaluation metric in Supplementary Material D.

annotations, minor disagreements on the exact beginning or end of group mentions are often inconsequential for capturing the essence of true group mentions. The strict standard the seqeval metric applies thus arguably results in overly conservative classification reliability estimates.[15]

The out-of-sample classification performances reported in Table 2 indicate that our supervised token classification approach to social group mention detection yields highly reliable measurements. In the Supplementary Material, we report additional evidence that supports this conclusion. First, the classifiers evaluated in Table 2 achieve similar levels of reliability in the other group types included in our coding scheme (see Table E2). Second, assessing the effect of the number of training samples on out-of-sample classification performance, we find that similar levels of reliability as those reported in Table 2 can be achieved when fine-tuning on only 4,000 labeled sentences (see Supplementary Material E.1.1). Third, we present evidence that our classifiers generalize well, as they can detect social group mentions not contained in their training data relatively reliably (see Supplementary Material 5). Fourth, in section 'Transfer to Other Parties, Domains, and Countries', we show that with very little additional labeled data, our classifiers fine-tuned on British party manifestos can be transferred relatively reliably to a different domain (parliamentary speech, cf. Osnabrügge et al. 2021a) and, with some reliability losses, also to another language (Licht 2023).

Convergent Validity with Measurements by Thau (2019)

We next demonstrate that the measurements generated with our approach also converge with those Thau (2019) has obtained through manual content analysis. Thau (2019) has tasked trained coders with manually coding group-based appeals made in UK Labour and Conservative party manifestos (1964–2015). Part of this task is identifying the explicit mentions of targeted social groups.

We use Thau's data to validate our approach in two ways. First, we assess whether the social group mentions Thau's coders have identified are also detected by our supervised token classification approach. To answer this question, we have matched the group mentions extracted by Thau's coders to the manifesto sentences from which they were retrieved,[16] applied a group mention detection classifier fine-tuned on our human-labeled UK party manifesto sentences,[17] and computed the average mention-level recall per group category in Thau's coding scheme.[18] As shown in Figure F1, our classifier performs overall consistently in his group categories, achieving average recall values above 0.90 in most categories. As discussed in greater detail in Supplementary Material F, the three exceptions to this pattern are explained by how our coding instructions diverge from Thau's.

Second, we use Thau's data to compare document-level indicators obtained with our automated method to those obtained with his manual approach. Specifically, we count the number of social group mentions in each party manifesto according to his records and our classifier's predictions and compare how they correspond. Figure 3 shows a high positive correlation between our and Thau's estimates. Moreover, our counts are systematically higher, which is expected since Thau has coded group-based appeals, and a group-based appeal implies a group mention but not *vice versa*.

---

[15]We illustrate and discuss the implications of this strict evaluation standard in Supplementary Material D and contrast it to our more permissive mention-level metric.

[16]We describe this procedure in Supplementary Material F.

[17]We fine-tuned a RoBERTa token classifier on 80% of labeled sentences in our UK parties' manifesto corpus.

[18]We focus on recall because Thau has coded group-based appeals. A group-based appeal implies a group mention but not *vice versa*. Hence, our classifier might detect mentions outside of group-based appeals.
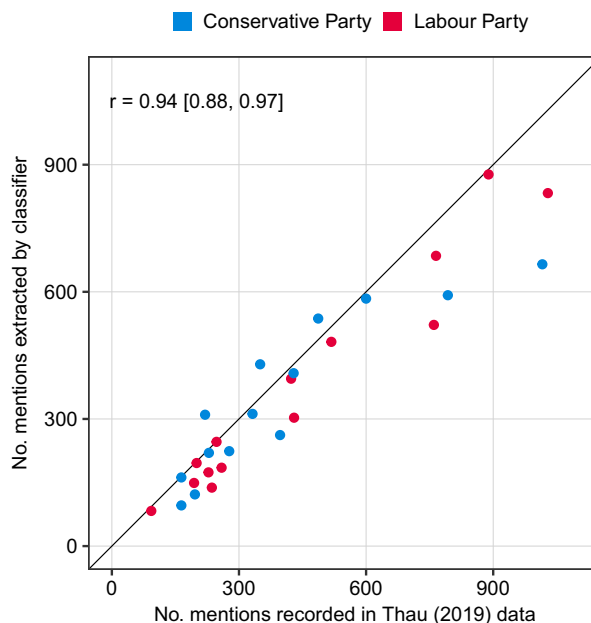
Figure 3. Cross-validation of RoBERTa group mentions detection classifier's predictions against data collected by Thau (2019). Figure compares the numbers of social group mentions identified in a manifesto by Thau (2019, see x-axis) and our classifier (y-axis) in Labour and Conservative party manifestos (1964–2015). Colors indicate parties. The correlation coefficient (with 95 per cent confidence interval) is shown in the top left of the plot panel.

Transfer to Other Parties, Domains, and Countries

The results presented thus far underscore the reliability and validity of our supervised group-mention detection method. However, applied researchers might want to adopt our approach to study group-based rhetoric in texts from other domains, countries, or languages. After all, in comparative politics and neighbouring fields, researchers typically want to compare political elites' communication behaviour across contexts.

To demonstrate the practical utility of our method, we assess the 'transferability' of the models we train on UK party manifesto data. By transferability, we mean the degree to which a classifier fine-tuned on labeled data from a 'source' context reliably classifies data from a 'target' context, which we consider an important dimension of generalization.

We examine the transferability of the classifiers obtained with our method in three scenarios. First, a *cross-party transfer* scenario in which we use labeled data from the Conservative and Labour Party manifestos as source data and that of the smaller British parties in our corpus (DUP, Greens, SNP, and UKIP) as target data. Second, we examine *cross-lingual transfer* using British parties' English-language manifestos as source documents and German parties' German-language manifestos as target documents (cf. Licht 2023). Third, we examine *cross-domain transfer* using British parties' manifestos as source documents and sentences from British *House of Commons* speeches as target documents (cf. Osnabrügge et al. 2021a). The datasets for these experiments are described in Supplementary Materials A and B.

In all three scenarios, we study zero- and few-shot transfer. Zero-shot transfer means classifying sentences from a 'target' context using a classifier solely fine-tuned on labeled sentences from the 'source' context. Few-shot transfer, in turn, means to continue fine-tuning this classifier on a few labeled sentences from the target context before applying it to other sentences from this context.

To examine how well our classifiers transfer in these scenarios, we have sampled the labeled data from the source context 50:50 into training and test splits. We have then started with evaluating the zero-shot setup by evaluating a classifier fine-tuned only on labeled sentences from the source context (for example, British manifestos) in the target-context test set (for example, sentences from German party manifestos).[19] By also evaluating the classifier in a source-context test set, we can compare the zero-shot transfer performance to the baseline of no transfer. We have then used portions of labeled sentences in the target-context training split to incrementally continue fine-tuning the classifier. For each scenario, we repeated this process with five different random seeds and averaged results across runs to account for uncertainty in fine-tuned classifiers' performances.

The results from these experiments are reported in Figure 4. The data points in the left-hand plot panels report the results without transfer, that is, from evaluating the classifiers in held-out test set examples from the respective source contexts. In the right-hand plot panels, the data points at x-axis values of 0, in turn, report the results for zero-shot transfer. Across scenarios, we find that zero-shot transfer comes with reliability losses. This should caution applied researchers against applying our pretrained classifiers to detect social group mentions in texts from other domains or languages. However, at least in the cases of cross-party transfer, the reliability losses are relatively modest.

However, Figure 4 also shows that the reliability of transfer to the target context can be improved through few-shot fine-tuning, that is, continuing to fine-tune the classifier pretrained on source-context sentences with a few labeled sentences from the target context. In all three transfer scenarios, classifiers' reliability in classifying target-context examples improves compared to the zero-shot baseline when continuing to fine-tune the classifier with a few hundred labeled sentences from the target context. As a point in case, in the cross-party transfer experiment (Figure 4a), continuing to train it with only 176 labeled sentences (10 per cent of the target corpus) allows matching the F1-score achieved in the source-context test set. Continuing to train with more labeled data from the target context does not improve classification performance in the target context further.

We find a similar initial improvement for cross-domain transfer from UK manifestos to parliamentary speech (see Figure 4c). However, as we continue to adapt the source-context classifier with more and more labeled parliamentary speech sentences, the classifier's target-context performance becomes more uncertain.

The results for cross-lingual transfer are not as strong (see Figure 4b). This might be explained by the fact that, in this setup, we transfer not only across languages but also party systems and political cultures. Nevertheless, even in the few-shot cross-lingual transfer experiment, 10 per cent of the labeled target corpus (361 labeled sentences) already yields substantial performance improvements relative to the zero-shot baseline.

Overall, our findings on the transferability of our classifiers suggest that, in practice, researchers can start with our pretrained classifiers and adapt them to their target context with a few labeled examples. We thus believe that our approach enables even less well-endowed researchers to size the scalability advantage of our proposed approach. Further, our results suggest that by fine-tuning using a small but diverse and potentially multilingual set of labeled sentences from different domains or countries, our approach could enable reliable detection and retrieval of social group mentions across political contexts. Our results thus highlight our approach's great promises for large-scale comparative research projects.

Applications

To illustrate the added value of our approach, this section presents two substantively motivated analyses of the measurements we have generated for British parties' election manifestos. These

---

[19]We fine-tuned DeBERTa models for cross-party and cross-domain transfer and XLM-RoBERTa models (Conneau et al. 2020) cross-lingual transfer.

(a)
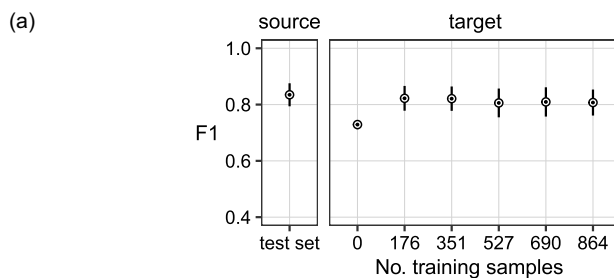
Cross-party transfer: From UK Labour and Conservative party manifestos ("source") to other UK parties' manifestos ("target").

(b)
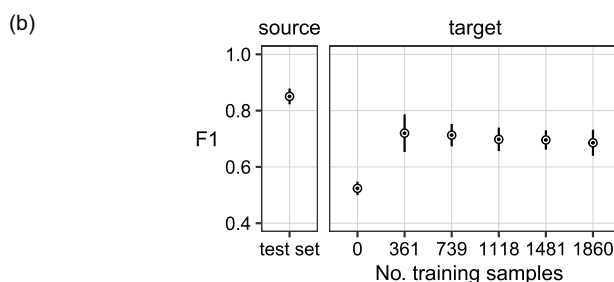
Cross-country/-lingual transfer: From British parties' manifestos ("source") to German parties' manifestos ("target").

(c)
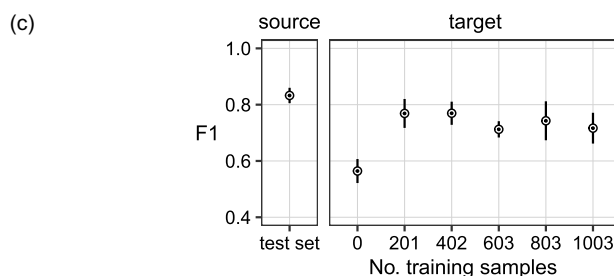
Cross-domain transfer: From British parties' manifestos ("source") to their speeches in the *House of Commons* ("target").

Figure 4 Summary of test set performances in cross-party, cross-lingual, and cross-domain transfer, respectively. The y-axis indicates the performance of classifiers trained on annotated manifesto sentences from the source context (for example, British manifestos) when evaluated on sentences from the target context (for example, German manifestos) in terms of the seqeval F1 score. Points (line ranges) report the average (pm 1 std. dev.) of performances of 5 different classifiers trained with different random seeds. Cross-party and cross-domain transfer results are based on fine-tuning DeBERTa models, and cross-lingual transfer results are based on fine-tuning XLM-RoBERTa models.

analyses show that our automated social group mention detection and extraction method allows testing theoretical claims and generating novel empirical insights. First, we study differences in British parties' social group focus regarding how much they emphasize groups in different policy areas and what distinguishes the groups they mention. Second, we show that sentences that contain mentions of social groups are more likely to include emotional language than sentences without group mentions.

What Distinguishes British Parties' Social Group Focus?

We first examine differences in British parties' social group focus in how much they emphasize groups in different policy areas, using data from the UK *Comparative Agendas Project* (CAP; Jennings et al. 2011). Specifically, we classify manifesto sentences according to the CAP policy topic they discuss[20] and then estimate the prevalence of social group mentions in Labour and Conservative party manifestos sentences (1983–2015) by CAP category.[21]

From the group appeals literature, we know that political parties combine policy and group appeals to cater to voters (cf. Huber et al. 2024; Robison et al. 2021; Thau 2023). Since group mentions can reflect parties' attempts at addressing groups' interests and shaping their opinions, we generally expect more mentions in policy areas marked by (re)distributive conflict, such as social welfare, compared to discussions about regulatory issues like the economy (Majone 1997). But we also expect differences in the emphasis parties place on social groups in policy areas due to divergent incentive systems for acquiring issues (Petrocik, 1996) and group yield (Huber 2021).
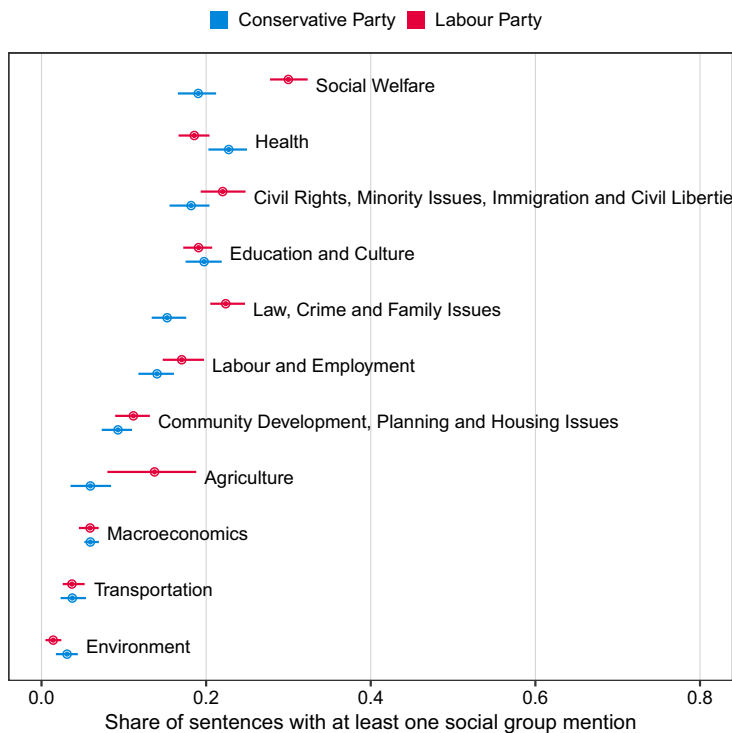

Figure 5. Social group mentions in Labour and Conservative party manifestos (1983-2015) by *Comparative Agendas Project* (CAP) policy topic. *Note:* Sentences CAP-coded using multiclass classifier trained on *human-labeled* manifestos of same cases (Jennings et al., 2011) Infrequent CAP policy topics grouped into the 'other' category. Topic 'Immigration' recoded to topic 'Civil Rights, Minority Issues, Immigration and Civil Liberties.'

---

[20]We fine-tuned a RoBERTa model for policy topic classification with human-coded quasi-sentences from UK Labour and Conservative party manifestos (1983-2015) in the CAP data. Note that we have collapsed the topics 8 ('Energy'), 15 ('Banking, Finance and Domestic Commerce'), 16 ('Defence'), 17 ('Space, Science, Technology and Communications'), 18 ('Foreign Trade'), 19 ('International Affairs and Foreign Aid'), 20 ('Government Operations'), and 21 ('Public Lands, Water Management, Colonial and Territorial Issues'), into one 'other' category because they were extremely sparsely populated. Moreover, we have assigned sentences originally coded to the 'Immigration' topic to the 'Civil Rights, Minority Issues, Immigration and Civil Liberties' topic.

[21]We focus on these parties and elections to avoid out-of-sample classification relative to the labeled CAP data.

Figure 5 presents evidence that supports both expectations. The overall salience of social group mentions in different policy topics aligns with our expectations. Distributive and redistributive policy areas (for example, social welfare, education, and civil rights) are more likely to include social group references than sentences about regulatory matters (for example, transportation, environment). In addition, we observe differences between parties in the degree to which they emphasize social groups when addressing these policy issues. Labour mentions social groups more in their manifestos than the Conservatives when talking about the topics of 'Social welfare' and 'Law, Crime, and Family issues'. A reverse pattern emerges tentatively in their discussion of macroeconomics topics. This suggests that parties emphasize social groups more in areas considered their core competencies, indicating an association between emphasis on social groups and issue ownership (Petrocik 1996).

Next, we analyze how British political parties distinguish themselves through their references to social groups. Previous studies emphasize that it is not only important *whether* groups are mentioned but also *which* groups (Huber, 2021; Thau, 2021) and *how* they are referred to (Graf et al. 2023).

To investigate this, we employ the 'fightin' words' method by Monroe et al. (2008) to the social group mentions identified and extracted by a RoBERTa classifier fine-tuned on our labeled British party manifestos corpus. In this analysis, we focus on manifestos from 2015 to 2019 to allow the inclusion of smaller British parties.

The 'fightin' words' algorithm (Monroe et al. 2008) is a bag-of-words method for quantifying differences in word choices between speakers, parties, or any other binary indicator. We use this method to compare the parties' social group mentions between pairs of parties. Specifically, we apply it to the predicted group mentions extracted from parties' manifestos after removing common stop words, retaining uni- and bi-grams, and adding skip-grams.

Figure 6 summarizes our findings. The x-axis shows term frequency. The y-axis displays z-scores that quantify how distinctive the words a party uses to refer to social groups when comparing pairs of parties. Higher z-scores indicate more distinctive words.

Analyzing Conservative and Labour manifestos, Labour emphasizes workers' and disadvantaged groups like people [with] disabilities,' refugees,' women, and 'BAME' (Black, Asian, and minority ethnic) and LGBT communities. By contrast, the Conservative Party focuses on 'ordinary working [people]', 'working families', 'British people', and the middle class (for example, 'doctors', 'entrepreneurs', and 'professionals').

Examining Greens and UKIP along the GAL-TAN dimension, Greens refer distinctively to age- and gender-based groups and disadvantaged communities, while UKIP, like the Conservatives, focuses on 'the nation' and 'British people', also mentioning immigrants and criminals.

Comparing Labour and the SNP, the center-periphery issue of Scottish independence is evident, with the SNP mentioning '[the] people [of] Scotland', 'Scottish', 'Scotland's' people, citizens, etc., as well as 'Scots'.

The insights into differences in British parties' social group focus we have generated in the analyses above underscore the practical value of our method. Automating the detection of references to social groups at a very granular level of measurement, our method allows detailed insights into how parties' group- and issue-based appeals correspond. Further, by extracting the exact words with which parties refer to social groups, our method facilitates inductive discovery and analysis of party rhetoric based on a limited set of human-annotated sentences.

What is more, our evidence presented in section 4.3 suggests that researchers can also harness these advantages of our method for analyzing texts from other domains and languages. For example, this promises new insights into individual legislators' group-based rhetoric.
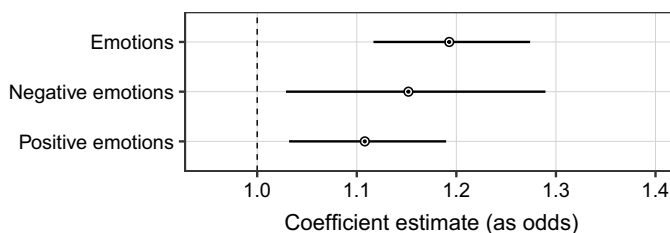
Figure 6. Different pairs of parties in terms of the words and phrases that distinguish the social groups the mention in their manifestos for the elections 2015, 2017, and 2019. **Note:***z*-scores indicate words 'distinctiveness' and have been obtained by applying the 'fightin' words' method proposed by Monroe et al. (2008) to the social group mentions retrieved by our classifier.

Is Group-Based Rhetoric Linked to Emotional Appeals?

Like directly mentioning social groups, emotional language is a powerful rhetorical strategy to appeal to voters (Crabtree et al. 2019; Gennaro and Ash 2022; Osnabrügge et al. 2021b). However, we do not know whether parties combine these two strategies in their campaign communication or use them separately.

We investigate the link between group-based rhetoric and emotional appeals through logistic regression analysis.[22] We use our sentence-level corpus of automatically labeled Labour and Conservative party manifestos from 1964 to 2019. Our dependent variable measures whether a sentence includes emotional language based on the *Linguistic Inquiry and Word Count* (LIWC) dictionary (Pennebaker et al. 2015). Specifically, we classified a sentence as containing emotional language (coded 1) if at least one word matched the list of positive and negative emotion words in the LIWC. If a sentence contained no emotional words, we coded it as zero (0). Further, we have created two additional indicators using only positive and negative emotion words, respectively. These alternative outcomes allow us to assess whether positive, negative, or both emotions contribute to the overall association.

Our main explanatory variable measures whether a sentence mentions one or more social groups, and we classify all sentences that contain at least one (predicted) social group mention as 1s and all others as 0s. To account for potential confounders, we control for parties' positions on the economy and cultural topics using Manifesto Project Data indicators (Lehmann et al. 2022), whether a party was the prime minister's party in the year leading up to the election for which the

---

[22]We present results from generalized linear models as well as models adjusted with the design-based supervised learning (DSL) method proposed by Egami et al. (2024).

Figure 7. Estimates from logistic regressions analyzing whether sentences that contain group mentions are more likely to contain emotion words. The x-axis reports our estimates of the odds that a sentence contains emotional language when it contains at least one social group mention compared to when it contains no social group mention. Points (line ranges) report the coefficients point estimates (95 per cent confidence intervals) of logistic regression models. The y-axis values differentiate between different emotion dictionary categories.

manifesto was written, and the number of words in a sentence. We use these indicators to fit logistic regression models with the binary emotion indicator as the outcome. All our models include election fixed effects, and we cluster standard errors at the level of parties and elections.

Figure 7 presents the coefficient estimates of our logistic regression models for our binary, sentence-level social group mention indicator as odds.[23] The odds measure how much more likely a sentence is to contain emotional words when it contains at least one social group mention compared to when it contains no social group mention. Figure 7 shows that sentences that contain at least one social group mention are about 1.2 to 1.4 times more likely to contain emotional words. This association exists with positive and negative emotional language use, as we find positive and statistically significant associations when measuring emotional language use only with positive or negative emotion words in the LIWC dictionary.[24]

This analysis underscores that applying our method for automatically detecting social group mentions in political texts enables new empirical insights into the relation between group-based rhetoric and emotional appeals in parties' campaign communication.

Conclusion and discussion

While the extant political science literature offers many hypotheses on how and why politicians relate themselves to social groups in their public communication, studying this facet of politics quantitatively is challenging with existing text-as-data methods. We have proposed a supervised token classification method that enables researchers to automatically identify and extract group mentions in large text corpora based on a small sample of human-annotated documents. After theoretically defining the target concept, human coders first highlight all text passages that mention social groups in a set of documents sampled from the target corpus. These labeled documents then serve as data to train a supervised token classifier that learns to predict labels at the word level while accounting for words' sentence context. Finally, the resulting classifier allows detecting and extracting group mentions in the entire target corpus.

We have illustrated this method in a study of British parties' group-based rhetoric. Trained on less than 7,000 labeled sentences, our token classifiers prove highly reliable in detecting social group mentions – independent of whether they are evaluated at the sentence or group mention

---

[23]All estimates are reported in Table H1 in the Supplementary Material. These results are robust when accounting for the classification error of our group mention detection classifier with the design-based supervised learning (DSL) method proposed by Egami et al. (2024) (see Table H2).

[24]Additional analyses reported in Tables H4 and H5 in the Supplementary Material show that this finding holds when we include minor parties' manifestos in our analysis or focus only on manifestos from the elections of 2015 onwards. Further, our finding holds when we apply the design-based supervised learning error correction approach proposed by Egami et al. (2024) (see Table H2).

level. Further, our cross-party, cross-domain, and cross-lingual transfer experiments show that adapting a pretrained group mention detection classifier to a new context can prove successful with only a few hundred labeled sentences from the target context. Moreover, our approach yields valid measurements. Document-level indicators of social groups' salience in party manifestos resulting from our supervised token classification approach correlate very strongly with those obtained through fully manual content analysis.

We demonstrated the innovative potential of our method with two applications. Applying our approach to all UK party manifestos in our corpus, we have documented that the British Labour and Conservative parties mention social groups to different extents when discussing different policy topics. Further, our inductive analysis of the words that distinguish British parties' social group mentions uncovered patterns familiar to students of party competition and cleavage formation. Second, we have applied our method to study the link between parties' mention of social groups and their use of emotional language, uncovering a positive association between these two rhetorical strategies.

Given these results and our encouraging findings about the data efficiency and generalization potential of our approach (see Supplementary Material E.1), we believe that our method opens up exciting new avenues for further research. For example, our proposed method could enable analyses of political elites' framing and stereotyping of groups, how they relate different groups to each other, how parties' attempts to create new or maintain existing voter linkages manifest in their communication, and how parties' group-based strategies respond to long-term socio-economic transformations.

We recommend three directions for further methodological research to enable these and other applications. First, future research should focus on developing and testing methods for inductively grouping extracted mentions into conceptually coherent categories (cf. Thau 2019, 70) like those applied in existing manual content analysis (for example, working-class people, Stückelberger and Tresch 2022). While our method predicts which parts of a sentence are group mentions, it does not categorize them into types of groups.

Second, we see great potential in our method for closing the gap between the concept of a group *mention* and that of a group *appeal*. To close this gap, researchers will need to measure how politicians relate themselves to the social groups they mention. We believe that existing natural language processing methods, such as aspect-based sentiment analysis, would allow learning from labeled data whether a group mentioned in a text is connoted positively or negatively (cf. Horne et al. 2024).

Third, future research should investigate whether applying in-context learning through Large Language Model (LLM) prompting proves as or even more reliable in social group mention detection as our Transformer encoder fine-tuning approach (cf. Jalali Farahani et al. 2024). Recent advances in so-called open-named entity recognition and information extraction with LLMs promise this to be a fruitful avenue for further methodological research (for example, Zhou et al. 2024).

Competing Interests. None declared.

References

Barberá, P, Boydstun AE, Linn S, McMahon R and Nagler J (2021) Automated Text Classification of News Articles: A Practical Guide. *Political Analysis* **29**(1), 19–42. doi: 10.1017/pan.2020.8.

Benoit K, Conway D, Lauderdale BE, Laver M and Mikhaylov S (2016) Crowd-sourced Text Analysis: Reproducible and Agile Production of Political Data. *American Political Science Review* **110**(2), 278–295. doi: 10.1017/S0003055416000058.

Bornschier S, Häusermann S, Zollinger D and Colombo C (2021) How 'Us' and 'Them' Relates to Voting Behavior-Social Structure, Social Identities, and Electoral Choice. *Comparative Political Studies* **54**(12), 2087–2122. doi: 10.1177/0010414021997504.

Chandra K (ed.) (2012). *Constructivist theories of ethnic politics*. New York: Oxford University Press. 500 pp.

Chatterjee S, Mukhopadhyay A and Bhattacharyya M (2019) A review of judgment analysis algorithms for crowdsourced opinions. *IEEE Transactions on Knowledge and Data Engineering* **32**(7), 1234–1248. doi: 10.1109/TKDE.2019.2904064.

Conneau A, Khandelwal K, Goyal N, Chaudhary V, Wenzek G, Guzmán F, Grave E, Ott M, Zettlemoyer L and Stoyanov V (2020) Unsupervised Cross-lingual Representation Learning at Scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. doi: 10.18653/v1/2020.acl-main.747.

Conover PJ (1988) The Role of Social Groups in Political Thinking. *British Journal of Political Science* **18**(1), 51–76.

Crabtree C, Golder M, Gschwend T and Indriđason IH (2019) It Is Not Only What You Say, It Is Also How You Say It: The Strategic Use of Campaign Sentiment. *The Journal of Politics* **82**(3), 1044–1060. doi: 10.1086/707613.

Devlin J, Chang M-W, Lee K and Toutanova K (2019) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1 (Long and Short Papers). Minneapolis, Minnesota: Association for Computational Linguistics, 4171–4186.

Dolinsky AO (2022) 'Parties' group appeals across time, countries, and communication channels – examining appeals to social groups via the Parties' Group Appeals Dataset'. *Party Politics* **29**(6). doi: 10.1177/13540688221131982.

Dolinsky AO, Horne W and Huber LM (2023) 'Parties' group appeals across space and time: An effort towards an automated, large-scale analysis of parties' election manifestos. Working Paper.

Egami N, Hinck M, Stewart BM and Wei H (2024) Using imperfect surrogates for downstream inference: Design-based supervised learning for social science applications of large language models. *Proceedings of the 37th International Conference on Neural Information Processing Systems*. NIPS '23. Red Hook, NY, USA: Curran Associates Inc.

Enyedi Z (2005) The role of agency in cleavage formation. *European Journal of Political Research* **44**(5), 697–720. doi: 10.1111/j.1475-6765.2005.00244.x.

Gadjanova E (2015) Measuring parties' ethnic appeals in democracies. *Party Politics* **21**(2), 309–327. doi: 10.1177/1354068812472586.

Gennaro G and Ash E (2022) Emotion and Reason in Political Language. *The Economic Journal* **132**(643), 1037–1059. doi: 10.1093/ej/ueab104.

Goodman, R. and S. Bagg (2022). 'Preaching to the Choir? Rhetoric and Identity in a Polarized Age'. In: *The Journal of Politics* **84**(1), 511–524. doi: 10.1086/715171.

Graf SM, Rubin Y, Assilamehou-Kunz M, Bianchi A, Carnaghi F, Fasoli E, Finell M, Sendén G, Shamloo SE and Tocik J (2023) Migrants, asylum seekers, and refugees: Different labels for immigrants influence attitudes through perceived benefits in nine countries. *European Journal of Social Psychology* **53**(5), 970–983.

Grimmer J and Stewart BM (2013) Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis* **21**(3), 267–297. doi: 10.1093/pan/mps028.

He P, Liu X, Gao J and Chen W (2021). *DeBERTa: Decoding-enhanced BERT with Disentangled Attention*. arXiv: 2006.03654[cs].

Hersh ED and Schaffner BF (2013) Targeted campaign appeals and the value of ambiguity. *The Journal of Politics* **75**(2), 520–534. doi: 10.1017/s0022381613000182.

Holman MR, Schneider MC and Pondel K (2015) Gender Targeting in Political Advertisements. *Political Research Quarterly* **68**(4), 816–829. doi: 10.1177/1065912915605182.

Hopkins DJ, Lelkes Y and Wolken S (2024) The rise of and demand for identity oriented media coverage'. *American Journal of Political Science*. doi: 10.1111/ajps.12875.

Horn A, Kevins A, Jensen C and Van Kersbergen K (2021) Political parties and social groups: New perspectives and data on group and policy appeals. *Party Politics* **27**(5), 983–995. doi: 10.1177/1354068820907998.

Horne W, Dolinsky AO and Huber LM (2024) Using LLMs to Detect Group Appeals in Parties' Election Manifestos. Working paper.

Howe PJ, Szöcsik E and Zuber CI (2022) Nationalism, Class, and Status: How Nationalists Use Policy Offers and Group Appeals to Attract a New Electorate. *Comparative Political Studies* **55**(5), 832–868. doi: 10.1177/00104140211036033.

Huber LM (2021) Beyond Policy: The Use of Social Group Appeals in Party Communication. *Political Communication* **39**(3), 293–310. doi: 10.1080/10584609.2021.1998264.

Huber LM, Meyer TM and Wagner M (2024) Social group appeals in party rhetoric: Effects on policy support and polarization. *The Journal of Politics* **86**(4), 1304–1318. doi: 10.1086/729946.

Huddy L (2001) From Social to Political Identity: A Critical Examination of Social Identity Theory. *Political Psychology* **22**(1), 127–156. doi: 10.1111/0162-895X.00230.

Jackson MS (2011) Priming the Sleeping Giant: The Dynamics of Latino Political Identity and Vote Choice. *Political Psychology* **32**(4), 691–716. doi: 10.1111/j.1467-9221.2011.00823.x.

Jalali Farahani F, Hanke S, Dima C, Heiberger RH and Staab S (2024) Who is targeted? Detecting social group mentions in online political discussions. *Companion Publication of the 16th ACM Web Science Conference*. Websci Companion '24. Stuttgart, Germany: Association for Computing Machinery. doi: 10.1145/3630744.3658412.

Jennings W, Bevan S and John P (2011) The Agenda of British Government: The Speech from the Throne, 1911-2008. *Political Studies* **59**(1), 74–98. doi: 10.1111/j.1467-9248.2010.00859.x.

Kam CD, Archer AMN and Geer JG (2017) Courting the Women's Vote: The Emotional, Cognitive, and Persuasive Effects of Gender-Based Appeals in Campaign Advertisements. *Political Behavior* **39**(1), 51–75. doi: 10.1007/s11109-016-9347-7.

King G, Lam P and Roberts ME (2017) Computer-Assisted Keyword and Document Set Discovery from Unstructured Text. *American Journal of Political Science* **61**(4), 971–988. doi: 10.1111/ajps.12291.

Kitschelt H (2000) Linkages between citizens and politicians in democratic polities. *Comparative Political Studies* **33**(6), 845–879. doi: 10.1177/001041400003300607.

Klamm C, Rehbein I and Ponzetto SP (2023). Our kind of people? Detecting populist references in political debates. In Vlachos A and Augenstein I (eds), *Findings of the Association for Computational Linguistics: EACL 2023*. Dubrovnik, Croatia: Association for Computational Linguistics. doi: 10.18653/v1/2023.findings-eacl.91.

Krippendorff K (2004) *Content analysis: An introduction to its methodology*. 2nd ed. Sage.

Lamont M and Molnár V (2002) The Study of Boundaries in the Social Sciences. *Annual Review of Sociology* **28**(1), 167–195. doi: 10.1146/annurev.soc.28.110601.

Lehmann P, Burst T, Matthieß T, Regel S, Volkens A, Weßels B, Zehnter L and **W.B.F.S.** (WZB) (2022). *Manifesto Project Dataset*. Version 2022a.

Licht H (2023) Cross-Lingual Classification of Political Texts Using Multilingual Sentence Embeddings. *Political Analysis* **31**(3), 366–379. doi: 10.1017/pan.2022.29.

Licht H and Ronja S (2025) Replication Data for: Detecting Group Mentions in PoliticalRhetoric: A Supervised Learning Approach, https://doi.org/10.7910/DVN/QCOQ0T, Harvard Dataverse, V1.

Lieberman E and Miller A (2021) Do online newspapers promote or undermine nationbuilding in divided societies? Evidence from Africa. *Nations and Nationalism* **27**(1), 238–259. doi: 10.1111/nana.12661.

Liu Z, Lin W, Shi Y, and Zhao J (2021) A Robustly Optimized BERT Pre-Training Approach with Post-Training. In *Proceedings of the 20th Chinese National Conference on Computational Linguistics*. Huhhot, China: Chinese Information Processing Society of China, 1218–1227.

Majone G (1997) From the positive to the regulatory state: Causes and consequences of changes in the mode of governance. *Journal of public policy* **17**(2), 139–167.

Mierke-Zatwarnicki A (2023) Varieties of identity politics: A macro-historical approach. Working Paper.

Miller AH, Wlezien C and Hildreth A (1991) A Reference Group Theory of Partisan Coalitions. *The Journal of Politics* **53**(4), 1134–1149. doi: 10.2307/2131871.

Monroe BL, Colaresi MP and Quinn KM (2008) Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict. *Political Analysis* **16**(4), 372–403. doi: 10.1093/pan/mpn018.

Muddiman A, McGregor SC and Stroud NJ (2019) (Re)Claiming Our Expertise: Parsing Large Text Corpora With Manually Validated and Organic Dictionaries. *Political Communication* **36**(2), 214–226. doi: 10.1080/10584609.2018.1517843.

Nakayama H (2018) *seqeval: A Python framework for sequence labeling evaluation*. Version 1.2.2.

Nakayama H, Kubo T, Kamura J, Taniguchi Y and Liang X (2018) *doccano: Text Annotation Tool for Human*. Version 1.8.0.

Nteta T and Schaffner B (2013) Substance and Symbolism: Race, Ethnicity, and Campaign Appeals in the United States. *Political Communication* **30**(2), 232–253. doi: 10.1080/10584609.2012.737425.

O'Grady T (2022) *The transformation of British welfare policy: Politics, discourse, and public opinion*. Oxford University Press.

Osnabrügge M, Ash E and Morelli M (2021a) Cross-Domain Topic Classification for Political Texts. *Political Analysis* **31**(1), 59–80. doi: 10.1017/pan.2021.37.

Osnabrügge M, Hobolt SB and Rodon T (2021b) Playing to the Gallery: Emotive Rhetoric in Parliaments. *American Political Science Review* **115**(3), 885–899. doi: 10.1017/S0003055421000356.

Pennebaker JW, Boyd RL, Jordan K and Blackburn K (2015) The Development and Psychometric Properties of LIWC2015. doi: 10.15781/T29G6Z.

Petrocik JR (1996) Issue Ownership in Presidential Elections, with a 1980 Case Study. *American Journal of Political Science* **40**(3), 825–850. doi: 10.2307/2111797.

Pitkin HF (1967) *The concept of representation*. University of California Press.

Quinn KM, Monroe BL, Colaresi M, Crespin MH and Radev DR (2010) How to Analyze Political Attention with Minimal Assumptions and Costs. *American Journal of Political Science* **54**(1), 209–228. doi: 10.1111/j.1540-5907.2009.00427.x.

Ramshaw L and Marcus M (1995) Text Chunking using Transformation-Based Learning. *Third Workshop on Very Large Corpora*.

Robison J, Stubager R, Thau M and Tilley J (2021) Does Class-Based Campaigning Work? How Working Class Appeals Attract and Polarize Voters. *Comparative Political Studies* **54**(5), 723–752. doi: 10.1177/0010414020957684.

Saward M (2006) The representative claim. *Contemporary Political Theory* **5**(3), 297–318. doi: 10.1057/palgrave.cpt.9300234.

Sczepanski R (2023) Who are the Cosmopolitans? How Perceived Social Sorting and Social Identities Relate to European and National Identities. *Comparative Political Studies* **57**(7), 1210–1239. doi: 10.1177/00104140231194054.

Simpson E and Gurevych I (2019) A Bayesian Approach for Sequence Tagging with Crowds. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China: Association for Computational Linguistics. doi: 10.18653/v1/D19-1101.

Skorupa Parolin E, Hosseini MS, Hu Y, Khan L, Brandt PT, Osorio J and D'Orazio V (2022) Multi-CoPED: A Multilingual Multi-Task Approach for Coding Political Event Data on Conflict and Mediation Domain. Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society. AIES '22. New York, NY, USA: Association for Computing Machinery. doi: 10.1145/3514094.3534178.

Slothuus R (2007) Framing deservingness to win support for welfare state retrenchment. *Scandinavian Political Studies* **30**(3), 323–344.

Stückelberger S and Tresch A (2022) Group Appeals of Parties in Times of Economic and Identity Conflicts and Realignment. *Political Studies* **72**(2), 463–485. doi: 10.1177/00323217221123147.

Thau M (2019) How Political Parties Use Group-Based Appeals: Evidence from Britain 1964–2015. In: *Political Studies* **67**(1), 63–82. doi: 10.1177/0032321717744495.

Thau M (2021) The Social Divisions of Politics: How Parties' Group-Based Appeals Influence Social Group Differences in Vote Choice. *The Journal of Politics* **83**(2), 675–688. doi: 10.1086/710018.

Thau M (2023) The Group Appeal Strategy: Beyond the Policy Perspective on Party Electoral Success. *Political Studies*, p. 00323217231220127. doi: 10.1177/00323217231220127.

Timoneda JC and Vallejo Vera S (2025) BERT, RoBERTa or DeBERTa? Comparing Performance Across Transformer Models in Political Science Text. *The Journal of Politics* **78**(1), 347–364. doi: 10.1086/730737.

Valenzuela AA and Michelson MR (2016) Turnout, status, and identity: Mobilizing Latinos to vote with group appeals. *American Political Science Review* **110**(4), 615–630. doi: 10.1017/S000305541600040X.

Weber C and Thornton M (2012) Courting Christians: How Political Candidates Prime Religious Considerations in Campaign Ads. *The Journal of Politics* **74**(2), 400–413. doi: 10.1017/S0022381611001617.

White IK (2007) When Race Matters and When It Doesn't: Racial Group Differences in Response to Racial Cues. *American Political Science Review* **101**(2), 339–354. doi: 10.1017/S0003055407070177.

Wolkenstein F (2021) Revisiting the constructivist turn in political representation. *European Journal of Political Theory* **23**(2), 277–287. doi: 10.1177/14748851211055951.

Wolkenstein F and Wratil C (2021) Multidimensional Representation. *American Journal of Political Science* **65**(4), 862–876. doi: 10.1111/ajps.12563.

Zhou W, Zhang S, Gu Y, Chen M and Poon H (2024) *UniversalNER: Targeted Distillation from Large Language Models for Open Named Entity Recognition.* arXiv: 2308.03279 [cs.CL].

Zollinger D (2022) Cleavage Identities in Voters' Own Words: Harnessing Open-Ended Survey Responses. *American Journal of Political Science* **68**, 139–159. doi: 10.1111/ajps.12743.

Zollinger D (2024) Place-based identities and cleavage formation in the knowledge society. *Electoral Studies* **88**, p. 102768. doi: 10.1016/j.electstud.2024.102768.