British Election Leaflet Project - Data overview

Gathering data on electoral leaflets from a large number of constituencies would be prohibitively difficult – at least, without major outside funding – without the resources of the crowdsourced record-keeping website Electionleaflets.org (http://www.electionleaflets.org). The gathering and parsing of the contents of the Electionleaflets.org database was done automatically via a script written in the Python programming language.[1] The script automatically downloads scanned leaflet images from the Electionleaflets repository, and scrapes and stores meta-information such as constituency, date of receipt, and any keywords or content flags stored along with the leaflet images. Within this dataset, we subset along content and party lines. For example, some leaflets in the repository were issued regarding local elections. As we are only interested in political communication during the general election, these leaflets were discarded. We also discarded leaflets that were incorrectly or incompletely scanned.

The Python scraper downloaded the images of the leaflets, so we were also able to manually code additional information about each leaflet's contents. We evaluate each leaflet on twenty-one policy dimensions: the economy, civil rights, immigration, health, agriculture, labour, education, environment, energy, transportation, law/crime/family issues, social welfare, community, banking, defence, international affairs, science, trade, government, territorial issues, and Europe. With exception of Europe and immigration, all issues dimensions are coded using the topics and subtopics used by the UK Policy Agendas Project.[2] In addition to policy issues, we also identified whether the leaflet:

- Highlights the traits or local ties of an individual candidate

[1] This script is available upon request.

[2] The full codebook is available at: https://policyagendasuk.files.wordpress.com [accessed 30 November 2015].

Includes a photograph of the candidate or leader, and whether those photos are professional headshots or 'action' shots (or both)

- Directly criticises other parties or candidates

Due to the flexible size and content of electoral leaflets, a single leaflet may include multiple policy statements, critical messages, or other forms of potentially interesting content. For example, a leaflet distributed by the Liberal Democrats in Bath in 2015 contains the following statements that can be coded in specific ways:

- "Only the Liberal Democrats are fighting to get a better deal for our local NHS services" (issue - health)
- A promise to "Cut Income Tax by an additional 400[pounds] for low and middle income workers" (issue - economy)
- "Liberal Democrats have spent our time in Government and on the Council standing in the way of unfair policies the Tories wanted to put through" (criticism of Conservative policy stance)

This leaflet also includes multiple pictures, mentions, and quotes of the Lib-Dem candidate, Steven Bradley. Similarly, in a leaflet distributed in Broxtowe during the 2015 general election, a Labour leaflet states that the party will:

- "Guarantee cancer tests and results within one week (issues – health)
- "Recruit 20,000 more nurses and 8,000 more GPs" (issues – health)
- "Fund this through a mansion tax on properties worth over £2 million, a clampdown on tax avoidance and a levy on tobacco firms (issues – economy)

The party also references its opponents in the following statements:

- "Under the Tories there's a crisis in A&E, waiting lists are at their highest for six years and it's harder to see a GP" (criticism of Conservative policy stance)

- "Nigel Farage wants an American-style healthcare insurance system, not our free NHS" (criticism of UKIP policy stance)

However, this leaflet makes no mention of the party's local candidate, Nick Palmer.

The distribution of leaflets by party, 2010-2017

We limit our data collection to include only general election leaflets published by Britain's most competitive parties, defined as the Conservative Party, the Labour Party, the Liberal Democrats, the UK Independence Party (UKIP), the Green Party, and the Scottish National Party (SNP).[3] These parties are unique in that they (1) have the resources available to engage in campaigning across a wide range of constituencies, and (2) are mainstream political parties with broad policy platforms rather than single-issue parties relying on niche appeals. These parties have both the resources and the motivation to tailor their broader message differently to different local audiences. As such, they are a logical choice for our analysis of local communication strategies.

The distribution of leaflets across parties for the 2010, 2015, and 2017 general elections is summarised in Table 1.

Table 1. Distribution of election leaflets by party, 2010-2017

	2010		2015		2017	
Party	Count	Per cent	Count	Per cent	Count	Per cent
Conservative	985	26.6	750	22.7	364	27.7
Green	214	5.8	373	11.3	107	8.1
Labour	997	26.9	891	27.0	406	30.9
Lib Dem	1,154	31.1	727	22.0	361	27.5
SNP	82	2.2	96	2.9	24	1.8
UKIP	277	7.5	467	14.1	53	4.0
Total	3,709	100.0	3,304	100.0	1,315	100.0

[3] We did not acquire sufficient leaflets from the Plaid Cymru to perform a reliable analysis.

Table 2 summarises the distribution of leaflets across constituencies, as well as average number of leaflets per constituency and the range.

Table 2. Distribution of election leaflets across constituencies, 2010-2017

Election	No. of constituencies included	Per cent constituencies included	Mean no. of leaflets per constituency	Range
2010	488	77.2	7.6	[1, 71]
2015	429	67.9	8.0	[1, 133]
2017	242	38.3	5.8	[1, 89]

Comparing our data to campaign spending – 2015 general election

Though parties and candidates are not obliged to provide information on the leaflets they produced, they are required to report the amount of money that they spent on unsolicited materials during the campaign. We note that the distribution of leaflets across parties broadly mirrors the patterns of party spending on unsolicited materials.[4] The largest share of the leaflets in our dataset were authored by the Labour party, which spent more than £7,000,000 on unsolicited materials – a figure that outweighed the spending of any other party in this area. The Conservative Party outspent Labour by more than £3 million overall, but Labour spent more on unsolicited materials. More than 60 per cent of Labour's spending went to unsolicited materials, compared to 28 per cent for the Conservative Party. We have fewer leaflets for the Conservatives and Liberal Democrats, both of which spent less on leaflets and other unsolicited materials. Our dataset contains far fewer leaflets for the minor parties, but these parties spent considerably less on unsolicited materials.

When we compare candidate spending on unsolicited materials with the distribution of leaflets by party across constituencies, we see that for all parties except the SNP – for

[4] Party spending on unsolicited materials can only be considered a rough proxy for the number of leaflets distributed. Some parties may have spent more money on design, while others may have favoured a wider distribution.

which we have relatively few leaflets – there is a positive correlation between candidate spending on unsolicited materials and the total number of leaflets we have for the candidate's party in the seat.[5] In other words, we have more communications from seats where candidates devoted a larger portion of their budget to unsolicited materials.

As a final robustness check, we use data from the 2015 post-election wave of the 2014-2017 British Election Study (BES) Internet Panel to calculate leaflet contact rates for each party by constituency – i.e., we estimate the percentage of BES respondents in the constituency who reported they received a leaflet from the given party in the previous 4 weeks. When we compare these figures to the total number of the total number of leaflets we have for the party in the constituency, we find a positive and statistically significant correlation for all parties included in our dataset. For each party, we calculate the percentage of BES respondents in each constituency who reported they received a leaflet from the given party in the previous 4 weeks. We then compare these estimates to the total number of leaflets we have for the party in that constituency. The correlations for the Conservative Party, the Labour Party, Liberal Democrats, the Green Party, and UKIP are 0.18, 0.28, 0.47, 0.56, 0.46, and 0.21, respectively. In all cases, the correlations are statistically significant at $p < 0.01$.

[5] For each candidate we add the total spending on solicited materials during the long and short campaign and we compare this figure to the total number of leaflets we have for the candidate's party in the constituency. The correlations for the Conservative Party, the Labour Party, Liberal Democrats, the Green Party, and UKIP are 0.32, 0.38, 0.50, 0.45, and 0.30, respectively. In all cases, the correlations are statistically significant at $p < 0.01$. For the SNP the correlation is -0.10, but $p > 0.05$.

Acknowledging the potential for bias in the data-generating process

While our dataset represents the largest collection of election leaflets to date, we acknowledge that it is a sample of convenience. These are self-reported data; there are no incentives or institutions encouraging citizens to upload their leaflets to the Electionleaflets repository, nor are parties required to report how many leaflets they disseminated. This means that we are unable to determine whether our sample is representative of larger population of leaflets distributed by parties in the run-up to the 2015 general election. That being said, we have no reason to believe that there is bias associated with the types of leaflets that individuals chose to upload. As we have stated previously, Electionleafets is run by a non-partisan organisation. On the website, individuals are encouraged to upload any – and all – leaflets they receive, and we have no reason to believe that those who uploaded leaflets did so strategically.

There are, however, a number of reasons to expect certain types of constituencies may be overrepresented in our sample. First, it has long been argued that individuals with more resources or interest in political matters are more likely to participate in politics (Brady et al., 1995). If such individuals are more inclined to upload their leaflets to the website, then our dataset may over-represented constituencies with more affluent and educated populations or constituencies where there is a higher level of political engagement. Second, constituencies in more urban areas may be more likely to have the have high-speed internet access that would allow residents to upload their leaflets with ease. Internet access is widespread in UK – as of early 2015, 86 per cent of the population reported using the Internet in the last 3 months (ONS, 2015) – but it is not universal. Providers may be more inclined to invest funds to upgrade lines to allow for high-speed access in urban areas, and therefore, we would expect more urban constituencies to be over-represented in our sample.

Finally, we expect a higher number of leaflets reported in marginal constituencies. This is not a source of bias in the same way as the demographic and infrastructural factors we note above, but is still a source of geographic variance that needs to be taken into account. Voters in marginal seats may not be any more likely to upload leaflets than voters in safe seats, but these constituencies are more likely to receive a higher number of leaflets overall, as British parties tend to spend more money on elections in marginal seats. Given a higher number of leaflets received in these constituencies, we would expect more leaflets to be reported in marginal seats.

Table 3. Constituency Sample Representativeness, 2015

Variables	Not included	Included	p-value	Not included	Included	p-value
Socio-demographic characteristics						
Population density (number per square km)	14.8	22.8	0.00	16.5	25.9	0.00
Internet take-up (% of households)	80.7	81.1	0.78	80.0	82.4	0.11
Internet population penetration	31.3	32.0	0.58	37.0	38.1	0.09
Professional occupations (%)	28.8	31.7	0.00	29.3	32.9	0.00
Routine/manual occupations (%)	35.7	31.7	0.00	27.2	23.7	0.00
Level 4 qualifications or above (%)	24.0	28.0	0.00	24.6	30.0	0.00
No qualifications (%)	25.1	22.5	0.00	24.7	21.1	0.00
Political engagement						
Interest in the election (mean)	3.4	3.4	0.03	3.4	3.4	0.00
Attention to politics (mean)	6.9	7.0	0.03	6.7	6.8	0.00
2010 election results						
Margin of victory (%)	19.1	18.1	0.32	24.4	23.4	0.44
Conservative Winner	0.5	0.5	0.37	0.50	0.51	0.72
Labour Winner	0.5	0.4	0.08	0.42	0.41	0.79
Lib Dem Winner	0.1	0.1	0.03	0.01	0.03	0.18
Other Party Winner	0.0	0.0	0.06	0.07	0.52	0.35
N	203	429		382	250	

Notes. Unless otherwise noted, figures represent the proportion of constituencies that belong to the given group.

Fortunately, there are ample data available that allow us to compare the constituencies in our sample to those that have been omitted. As an exploratory step, we conduct a series of

t-tests to identify systematic differences in constituencies that report leaflets versus those that do not. Table 3 compares the mean values for the constituencies that are included in our dataset versus those constituencies where we have no leaflets on a number of dimensions, as well as the p-value associated with the difference between the two groups. A p-value below 0.05 suggests that constituencies for which we have data differ meaningfully from the constituencies that are absent from the dataset on that trait. When we compare our sample and the constituencies that are missing from the dataset in terms of socio-demographic and economic characterises of the seats, we find that our sample of constituencies tends to be more affluent, educated, and more densely populated. Interestingly, we find no differences between our sample and the omitted constituencies in terms of household internet take-up or superfast household internet access.[6] Similarly, when we compare the two sets of constituencies in terms of political engagement we find that, on average, respondents surveyed in the constituencies for which we have data tend to report having more interest in the election and paying greater attention to politics.[7] All of these differences are consistent with our theoretical expectations above, but taken together, the tests suggest that the differences we observe are due more to factors that relate to the political engagement or socio-economic resources of a constituency's population, rather than to an ability to upload the communications quickly. In addition to these key characteristics, we also compare our sample of consistencies with those omitted in terms of the political differences. Counter to the arguments above, there is no difference in terms of margin of victory after the 2010

[6] There are no publically available data on internet access at the constituency level. We rely on estimates provided by Point Topic, an independent company that forecasts internet access by postcode. We aggregate these estimates up to the level of parliamentary constituency.

[7] The two measures of political engagement, *Interest in the election* and *Attention to politics*, are taken from the post-election wave of the 2014-2017 British Election Study Internet Panel. For each variable, we calculate the weighted mean of all respondents in the constituency.

general election. However, there are some other political differences, namely our sample over-represents constituencies held by the Liberal Democrats after 2010.[8]

In summary, we acknowledge the limitations associated with the data. First, as discussed previously, these are self-reported data, meaning that we have no control over who uploads leaflets or where they are uploaded. However, our results remain robust after controlling for a wide variety of potentially biasing factors. Second, while our dataset includes a large number – and a wide range – of electoral communications, this is nowhere near a complete count of leaflets distributed during the campaign. While there are no official figures of election communication distribution, based on a survey of election agents, Fisher et al. (2012) estimate that the main parties distributed 27-35 million leaflets and communications prior to the 2010 general election. Many of these leaflets would have been of the same design, distributed to households across the country, but we acknowledge that our data captures only a small portion of the leaflets distributed. The nature of our sample of convenience means that we must be cautious about drawing deterministic conclusions about the larger population of leaflets and parties' behaviour more generally.

Despite these limitations, we also stress that despite the fact that parties devote much of their campaign spending to disseminating electoral communications, *there are currently no other data available that allow us to explore messaging contained in these materials across a large number of constituencies in recent elections*. Parties and candidates are not required to make this information available, and they are unlikely to do so unasked. A representative sample would certainly be preferable, but nevertheless, we still contend that our sample of convenience provides better insights than no sample at all. Election communications are such a key point of interaction between voters and political elites during an election, that failure to

[8] In supplementary analyses, we also tested for political differences using the results from the 2015 general election, but we found no meaningful differences on any dimension.

gain insight into these messages leaves a significant gap in our understanding of how campaigns are conducted. Thus, our data provides a unique avenue to explore the types of issues parties emphasise, how frequently and where they talk about their opponents, and even how they use images in their communications.