

Supplementary Appendix

This memo contains the supplementary analyses that are referenced in the paper “The face of the party? Leader personalisation in British campaigns.”

Data gathering and coding

Gathering and parsing 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 focuses on the traits of an individual candidate and if the leaflet includes a photograph of the candidate or leader; if the leaflet and the extent to which the publishing party 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 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)

¹ This script is available upon request.

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

- 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, 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.

This content-coding scheme allows us to perform a variety of detailed statistical analyses: not just where – and to what extent – the different parties focus their limited messaging budgets, but also how a single party may tailor the type and focus of their message to the localized concerns of different constituencies.

Further information on the distribution of photographs

The primary focus of the paper is decision to include photographs of party leaders, however, we acknowledge that candidate photographs feature prominently in election leaflets. As we state in the text, more than 80 per cent of all of the election communications we observed include an image of the party’s local candidate, while less than one in five leaflets contains an image of the party’s leader.

[Table S1]

Evaluating the potential bias associated with self-selected data

While comparisons with other available sources of campaign data suggest that our sample is a reasonable proxy for patterns of leaflet distribution by constituency, we acknowledge the potential for bias in the sample. Our dataset covers a wide range of seats in England, Scotland, and Wales, but there are 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.

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 S2 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.

[Table S2 here]

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.³ 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.⁴ 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 constituencies 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 general election. However, there are some other political differences, namely our sample over-represents constituencies held by the Liberal Democrats after 2010.⁵

The constituencies included in our sample are similar to those omitted in terms of their political characteristics, but there are still important socio-demographic and economic differences. Therefore, we may need to account for these factors when using the data to explore variation in the messages of the leaflets. We return to this issue below.

Leaflet publication and campaign contact

As an additional robustness check we use data from the 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

³ 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.

⁴ 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.

⁵ We also tested for political differences using the results from the 2015 general election, but we found no meaningful differences on any dimension.

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 $p < 0.01$. It is worth noting that the correlations between the BES and our self-reported data are lowest (but still significant) for the Conservatives (0.18) and UKIP (0.21). This may be in line with the potential sources of bias we note earlier. The demographic profile of UKIP supporters is consistent with those less likely to upload leaflets they receive. Though Conservative voters tend to be more affluent, the party tends to dominate in rural areas, which also feature less prominently in our dataset.

Supplementary analyses 1: Accounting for the potential bias associated with self-selected data

Our preliminary analyses above indicated that certain types of constituencies were more likely to ‘select’ into our dataset. Given the systematic nature of this selection process, ignoring the constituencies for which we have no leaflets introduces estimation bias that could lead us to make erroneous inferences from the data (Heckman, 1976). Therefore, we employ a two-stage Heckman selection model (Heckman, 1979), which allow us to model the selection into the dataset (stage 1) and the presence of an image of the leader (stage 2).

The first stage of the Heckman model estimates a binary outcome indicating whether any leaflets from a given constituency appear in our data set. In the first stage of the model, the dependent variable is coded ‘1’ if our dataset includes leaflets from a given constituency is included in our dataset, and ‘0’ if the constituency is missing. We model a constituency’s inclusion as a function of a number of key characteristics, including the affluence of the constituency, the winner and the marginality of the seat after the 2010 general election, population density, and the level of internet penetration in the constituency.⁶ Using data from the BES, we also control for

⁶ We conduct a principal factors analysis of the median income of the constituency, the percentage of the constituency who have advanced educational qualifications, no educational qualifications, are employed in routine or manual occupations, or are employed in professional occupations. All five variables have high loadings on a single dimension, which we label ‘constituency affluence’ The use of principal factors factor analysis allows us to create a single

the average level of attention to politics and interest in the election in the constituency. Finally, in an effort to gain some traction of the supply of leaflets in the seat, we use data from BES to estimate the average rate of party contact via leaflet per constituency. We expect that constituencies where, on average, respondents were contacted by more parties would be more likely to be represented in our dataset.

In the second stage of the analysis, we replicate the multivariate analyses presented in Table 2 of the paper. Here, the dependent variable is coded ‘1’ if the leaflet includes at least one image of the party’s leader, and ‘0’ otherwise. The results of the multivariate analysis, presented in Table S4, are substantively similar to the logistic regression models presented in the text: factors that were identified as statistically and substantively significant in the logistic regression model are identified with the Heckman as well.

[Table S4 here]

Supplementary analyses 2: Accounting for variation in the number of leaflets across constituencies

The simplest way to examine the relative frequency by which leaders feature in the electoral communications is to take the percentage of each party’s leaflets in that constituency that include at least one image of the party’s leader. However, given the wide variance in how many leaflets are distributed/reported over these constituencies, we have concerns that merely reporting percentages may be misleading. For example, consider a constituency where we have two Conservative leaflets reported, one of which contained an image of David Cameron. In this case, we would record that 50 per cent of Conservative leaflets in this constituency contained an image of the party’s leader. However, given that we only observe two Conservative leaflets in this constituency, our level of confidence as to the ‘true’ use of leader imaging is very low.

To account for the variation in the number of leaflets reported by constituency, we weight leaflet counts and the number of leaflets containing an image of the party leader by the logged number of total leaflets reported (weights = $\log(N+1)$ for all constituencies where $N \geq 1$) because we have more confidence about constituencies for which we have more information, giving us good reason to weight patterns of political communication in these locations more heavily. This means that, for example, a constituency where 10 out of 20 leaflets include an image of the party

variable, which preserves degrees of freedom and reduces the measurement error associated with any one social characteristic and the overall affluence of the constituency. The factor loadings and eigenvalues are shown in Table S3.

leader will have a greater impact on our overall figures than a constituency where one out of two leaflets includes an image. We use logged leaflet counts rather than raw counts because the distribution of leaflet reporting is quite skewed. By logging leaflet count, we are able to assure that high-volume constituencies exert a greater weight on our estimates that properly reflect the additional information available in these areas, while not allowing outliers to exert undue influence on our results.

Table S5 presents a logistic regression model in which observations are weighted by the (logged) number of leaflets reported in a given constituency. Leaflets from constituencies in which many leaflets are reported are given higher weight, and therefore greater influence on the estimated coefficients. This robustness check is implemented because we have fewer degrees of freedom in constituencies where only one or a few leaflets are reported, potentially giving those leaflets outsized weight in estimating relationships between constituency-level inputs and leaflet-level outputs. Here, too, we find substantively identical results in the statistical and substantive impact of our key independent variables.

[Table S5 here]

References

- Brady, Henry E., Sidney Verba, and Kay Lehman Schlozman. 1995. "Beyond SES: A Resource Model of Political Participation." *American Political Science Review* 89(2): 271-294.
- Heckman, James J., 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for such Models." *Annals of Economic and Social Measurement* 5: 475-492.
- Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47: 153-162.

Table S1. Distribution of photograph type by party

Party	Leader Photograph (%)	Candidate Photograph (%)	Both (%)
Conservative	41.1	86.3	37.1
Green	32.7	88.2	30.6
Labour	7.8	74.2	3.5
Liberal Democrat	3.9	81.0	3.9
SNP	64.6	63.5	36.5
UKIP	11.3	79.4	4.5
All	19.4	80.4	15.3

Table S2. Constituency Sample Representativeness

Variables	Not included	Included	p-value
<i>Socio-demographic characteristics</i>			
Population density (number per square km)	14.84	22.77	0.00
Internet take-up (% of households)	80.67	81.10	0.78
Superfast internet take-up (% of households)	31.30	31.95	0.58
Median income (1,000s)	20.53	21.68	0.00
Professional occupations (%)	28.82	31.65	0.00
Routine/manual occupations (%)	35.67	31.70	0.00
Level 4 qualifications or above (%)	24.02	28.04	0.00
No qualifications (%)	25.05	22.48	0.00
<i>Political engagement</i>			
Interest in the election (mean)	3.41	3.44	0.03
Attention to politics (mean)	6.86	6.95	0.03
<i>2010 election results</i>			
Margin of victory (%)	19.09	18.07	0.32
Conservative Winner	0.46	0.50	0.37
Labour Winner	0.46	0.38	0.08
Lib Dem Winner	0.05	0.11	0.03
Other Party Winner	0.03	0.01	0.06
N	203	429	

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

Table S3. Factor Analysis Results - Constituency Affluence

Variables	Factor loading
Median income (1,000s)	0.88
Professional occupations (%)	0.91
Routine/manual occupations (%)	-0.95
Level 4 qualifications or above (%)	0.95
No qualifications (%)	-0.92
Eigenvalues = 4.27,-0.18,- 0.01, -0.03, -0.07,	

Notes. This table presents the results of a factor analysis we use to construct a variable that measures the degree of urbanization in constituencies. One clear factor emerges and all of the variables of interest load at levels greater than 0.8. We use the predicted values of this first factor to measure constituency affluence.

Table S4. Two-stage probit selection model of leader personalisation

Variables	Party only	Local context only	Candidate traits only		Full model
<i>2nd stage: Leaflet contains image of party leader</i>					
Party (ref: Labour)					
Conservative	1.106** (0.077)				1.291** (0.100)
Liberal Democrat	-0.360** (0.094)				-0.061 (0.177)
SNP	1.727** (0.140)				1.627** (0.190)
UKIP	0.230* (0.087)				0.259* (0.129)
Green	0.884** (0.089)				1.238** (0.118)
Local popularity		0.350** (0.054)			0.179** (0.069)
2010 Margin of victory		-0.005 (0.011)			0.002 (0.012)
Local popularity*Margin of victory		-0.000 (0.003)			-0.002 (0.003)
Incumbent			0.303** (0.062)		0.132 (0.079)
MP more than 10 years				-0.085 (0.105)	
Local ties			0.460** (0.058)	0.417** (0.053)	0.403** (0.064)
Constant	-1.250** (0.066)	-2.137** (0.220)	-1.088** (0.058)	-0.987** (0.052)	-2.323** (0.296)
<i>1st stage: Leaflet from constituency included</i>					
2010 constituency status (ref=Other)					
Conservative-hold	0.990** (0.323)	0.888* (0.337)	1.060** (0.333)	1.044** (0.345)	1.190** (0.359)
Labour-hold	0.899** (0.328)	0.853* (0.343)	0.994** (0.354)	0.972** (0.359)	1.111** (0.363)
Lib Dem-hold	1.034** (0.347)	1.023** (0.366)	1.128** (0.379)	1.111** (0.380)	1.286** (0.383)
Margin of victory	-0.007* (0.003)	-0.004 (0.003)	-0.005 (0.004)	-0.005 (0.004)	-0.005 (0.004)
Constituency affluence	0.354** (0.052)	0.337** (0.056)	0.362** (0.055)	0.363** (0.055)	0.370** (0.054)
Population density	0.010** (0.003)	0.011** (0.003)	0.012** (0.003)	0.012** (0.003)	0.012** (0.003)
Internet take-up	-0.608* (0.266)	-0.716** (0.275)	-0.808* (0.283)	-0.803** (0.284)	-0.767** (0.283)

Interest in the election	-0.238 (0.285)	-0.162 (0.307)	-0.146 (0.303)	-0.151 (0.304)	-0.255 (0.301)
Attention to politics	0.141 (0.101)	0.141 (0.109)	0.174 (0.110)	0.175 (0.111)	0.187 (0.109)
Party leaflet contact	0.346** (0.097)	0.349** (0.102)	0.278** (0.104)	0.274** (0.104)	0.268** (0.102)
Constant	0.528 (0.841)	0.338 (0.879)	0.018 (0.885)	0.045 (0.887)	0.185 (0.882)
N (Leaflets)	3,300	3,300	2,720	2,720	2,720
Censored Constituencies	199	199	199	199	199
ρ					
Wald χ^2	-0.781	-0.040	-0.479	-0.454	-0.636
[$p > \chi^2$]	20.98 [0.000]	0.02 [0.889]	4.99 [0.025]	4.50 [0.034]	10.63 [0.001]

Notes. * $p < 0.05$ ** $p < 0.01$.

Table S5. Logistic regression model of leader personalisation (weighted data)

Variables	Party only	Local context only	Candidate traits only		Full model
Party (ref: Labour)					
Conservative	1.983** (0.082)				2.297** (0.116)
Liberal Democrat	-0.908** (0.128)				-0.233 (0.153)
SNP	2.983** (0.153)				2.776** (0.213)
UKIP	0.279* (0.114)				0.285 (0.180)
Green	1.602** (0.095)				2.367** (0.131)
Local popularity		0.741** (0.059)			0.479** (0.078)
2010 Margin of victory		0.008 (0.012)			0.006 (0.014)
Local popularity*Margin of victory		-0.004 (0.003)			-0.004 (0.003)
Incumbent			0.491** (0.067)		0.233** (0.087)
MP more than 10 years				-0.204 (0.124)	
Local ties			0.861** (0.062)	0.781** (0.061)	0.871** (0.071)
Constant	-2.311** (0.069)	-4.262** (0.247)	-1.974** (0.048)	-1.788** (0.041)	-4.959** (0.338)
<i>N</i>	3,304	3,304	3,304	2,723	2,723
McFadden <i>R</i> ²	0.179	0.046	0.030	0.023	0.246

Notes. Robust standard errors are given in parentheses. * $p < 0.05$ ** $p < 0.01$.