Social mobilisation in partisan spaces

Data Appendix

1 Selection of the Spillover Sample

Large numbers of households were excluded from the experiment during the design phase. On the one hand, there were legal considerations, such as individuals on no-contact lists. On the other hand, some households had to be excluded for feasability considerations, i.e., to ensure that the share of households could realistically be targeted within the campaign budget. Those excluded households comprise the spillover sample.

The Labour Party implemented the experiment from our randomization and sample selection rules. Our final selection of restrictions on the experimental sample is a carefully checked combination of the restrictions as we designed them and as they were implemented by Labour. Table 1 contains an overview over the restrictions. It shows whether the restriction was implemented in our design code, in Labour's implementation, and in the final selection for the analysis. Those restrictions are discussed in more detail below.

Table 1: Overview over sample restrictions.

Description	Our Code	Labour's File	**Final Selection**	Explanation
Exclude wards with no work	Yes	Yes	**Yes**	Those individuals are dropped early in the final code.
Moved or died	Yes	Yes	**Yes**	Individuals who died or moved are dropped early in the final code.
Do not contact	Yes	Yes	**Yes**	We follow our code 1:1, not Labour's, here.
Postal voters	Yes	No	**Yes**	Labour's file does not explicitly mention to exclude postal voters, but careful checking of the implementation data revealed that those individuals were actually excluded as specified in our code.
Partisanship	Yes	No	**Yes**	Individuals with certain voter identifications were excluded across all wards. Additionally, Conservative voters were excluded in two of the wards.
People who cannot vote	No	Yes	**No**	Individuals who cannot vote are not excluded from the experiment but we create a variable `unable_to_vote` which flags people who do not have voting rights.

The first two restrictions, **exclude wards with no work going on** and remove individuals who **moved or died**, are elementary. Labour targeted only a selection of nine wards in the constituency. Data from additional wards where no work was being implemented so they were not part of the experiment are excluded at the start. A very small number of individuals in our data were marked as having moved or died before the experiment was implemented. Those individuals are not part of the experiment in any way. Individuals who fall under either category are completely dropped from the data set at the very beginning because they are

not relevant to the present experiment. Therefore, they are not part of the analysis and replication materials.

The three following restrictions split the sample into individuals who were directly targeted in the experiment and individuals who were not directly targeted. Individuals who are on **no contact** lists or were missing phone contact information were not part of the experiment because the original experimental design included a telephone survey as one of the outcomes. Therefore, individuals who could not be contacted for legal or practical reasons were not part of the experiment. In a similar vein, **postal voters** were also excluded from being targeted in the experiment. Finally, individuals with certain pre-treatment **partisanship** identifications were systematically excluded from the experiment. This was necessary because limitations in campaign funds. Therefore, individuals with a range of pre-treatment party identifications were systematically excluded such that Labour was able to target voters who they identified as more likely Labour supporters. Table 2 gives an overview over the number of individuals who were excluded from the experiment for each of the three restrictions.

Table 2: Summary of Restrictions and Final Sample at the Individual Level.

	Yes	No
Do not contact?	31,810	13,705
Postal voter?	8,564	36,951
Partisanhip restriction?	19,263	26,252
	Indirect	Direct
Individual sample	37,534	7,981

These three restrictions split the sample into individuals who were part of the experiment, i.e. directly targeted, and other who were only indirectly part of the experiment. While the restrictions are defined and implemented on the individual level, the actual level of treatment distribution is the household. Therefore, certain individuals who would not have been part of the experiment were directly targeted as well because at least one other member in the same households was targeted. They were thus exposed to the leaflet (or the control condition) because at least one household member was part of the experiment. The individual-level variable denoting the experiment sample has, therefore, to be expanded to account for this detail. Table 3 contains the number of individuals in each sample before and after adjusting for households.

Table 3: Experiment Sample before and after Household-Level corrections.

	Direct	Indirect
Individual sample	7,981	37,534
Final sample	9,460	36,055

Lastly, some individuals are marked as having **no voting rights** for the election under study. Those individuals should have been excluded from the experiment and the campaign as a whole as they cannot be persuaded in any way for obvious reasons. However, this was not consistently implemented. Thus, we do not exclude these individuals from the experimental sample but we create an indicator variable. Table 4 contains the number of individuals without voting rights broken down for the two samples.

Table 4: Overview of Individuals Without the Right to Vote in the Sample.

	Sample		
Voting Rights?	Direct	Indirect	Total
Yes	9,387	34,833	44,220
No	73	1,222	1,295

2 Geolocation and Distance Computation

We use the Google's Geocoding API to locate addresses in our data on the map, i.e. to find their latitude/longitude.

Geolocating addresses is an inherently inexact procedure with several problems. One is that the names of towns and streets are in no way unique even within a region or country and much less worldwide. Google's solution to this problem is what they call "region biasing". We biased the results towards the UK, such that results from the UK are prefered but results are not restricted to be in the town the experiment ran or the UK. Therefore, we put the data from Google through a number of post-processing steps to make sure that results are in the UK and results are in the town

Quality of the data

The Google Maps API gives two indications of the quality of the geolocated address (see here). One is an indicator for whether only a part of the address could be located. The other gives an indication for how precise the geocode is.

A partially matched address means that the complete address in the input does not exist but a subset of the address parts could be matched. Around 20% of the addresses we geolocated are of this type. This is to be expected as many of the address lines contain more than the street address, especially the name or number of the flat, e.g., "Flat 4, 249 Cumberland Road, ...". Although the match for the address is only partial, that is, only "249 Cumberland Road, ..." is identified, the location of the building is still correct.

The precision of the location is returned in four categories. The best quality matches are precise geocodes for the building. If those are unavailable, Google Maps either interpolates the range between two points (e.g. neighbouring street addresses with precise locations), locates the address at the geometric center of a road, or returns an "approximate" location without further specification.

To deal with those two issues, we use two versions of the geolocated data to compute distance-based variables for the analyses. The first one, which is identified by the `all` keyword in the variable name (e.g. `match_top1_all_dist`), uses all available geolocation data as long as it passes the checks described above. This maximizes the amount of available data and thus the power of the analyses.

Identification of Closest Household(s)

The first step is to compute the distance between addresses from the latitude/longitude from the Google Maps API. The distances we compute are "as the crow flies" or "great-circle distances". The algorithm ("Vincenty" ellipsoid) and reference grid (WGS84) results in distances which are exact +/- 1 metre. As treatment assignment was on the street level, we only compute distances and identify closest households within the same street.

For every household which was not targeted directly in the experiment, we identify the closest household(s) which were part of the experiment. This happens in two different modes. The first mode is when there is no household in the experiment at the same address. This can either be because there is only one household in the building or, if there are more than a single household, no household in the given building was part of the experiment. In this case, the closest household is identified by minimizing the distance between the location of the current address and the location of the all addresses with at least one household in the experiment. The second mode is when there are households in the experiment at the same address as the in indirect household to match. This very common for any building, i.e. a single locatable entity, comprising several separate flats. In this case, all households in the same building which were directly part of the experiment are identified as the closest households. If households are matched within buildings, the distance to the match is equal to zero. In this case, we impute the distance with the closest non-zero distance in our data and we generate a dummy variable for whether the matched household(s) are within the same address (e.g. `match_top1_all_within`).

Both modes often lead to ties which is most commonly the case when households are matched within buildings or the address of closest treated household includes additional treated households. In case of a tie, all closest households are averaged over. We then compute the number of individuals (`..._n_obs`) and the number of households (`..._n_hh`) we average over.

3 Data Description

Table 5 contains an overview of the number of units split between treated and control across direct and indirect samples.

Table 5: Number of Individuals, Households, and Streets in each Group.

	Direct		Indirect	
	Treated	Control	Treated	Control
Individuals	6,940	2,520	26,667	9,388
Households	3,447	1,250	13,760	4,944
Streets	447	166	447	165

Table 6 contains an overview over the number of individuals for each pre-treatment voter identification as well as the guide on how we recoded and aggregated these voter identifications into variables in our data.

Table 6: Voter Identification Breakdown and Details of Recoding.

		Recoding Details					
Voter Id.	N	..._labour	..._conservative	..._rivalparty	..._other	..._nonvoter	..._observed
labour	11,930	1	0	0	0	0	1
conservative	2,135	0	1	0	0	0	1
against	6,768	0	0	1	0	0	1
bnp	6	0	0	1	0	0	1
green	177	0	0	1	0	0	1
liberal democrat	460	0	0	1	0	0	1
other party	12	0	0	1	0	0	1
plaid cymru	1	0	0	1	0	0	1
respect/soc lab	2	0	0	1	0	0	1
snp	3	0	0	1	0	0	1
uk independence	301	0	0	1	0	0	1
don't know	6,213	0	0	0	1	0	1
independent	1	0	0	0	1	0	1
won't say	297	0	0	0	1	0	1
non voter	3,375	0	0	0	0	1	1
n/a	13,834	0	0	0	0	0	0