Differential inequalities in protest participation in Britain: Comparing non protesters to novices, the middle group and stalwarts

Abstract

Inequalities in protest participation are shrinking as protesters become increasingly similar to the non-protesting population. However, extant evidence only looks at a binomial (yes/no) response to protest participation. This paper considers the extent to which those who protest to different degrees represent non-protesters. Selection bias in protest survey data is compensated for by combining data with random samples from the European Social Survey. Hypotheses on the normalization of protesters are tested using statistical analysis and methods for combining samples are compared. A propensity score stratification approach provides a fairly representative combined sample of protesters to compare to non-protesters. We find that protesters become increasingly differentiated from non-protesters as their extent of protest participation increases.

Key words: normalization of protest, political participation, protest inequalities, protest survey, propensity matching

Introduction

The days of automatically viewing all protest as a minority activity undertaken by extreme radicals appear to be over. It is now commonly argued that protest is a popular and ever-expanding means of engaging with politics (see, for example, Norris, 2003). Scholarly articles on protest often begin with a grand statement that 'protest is booming' (e.g. Walgrave, Wouters & Ketelaars, 2016). In tandem, it is frequently asserted that a decline in electoral participation is compensated by a rise in alternative forms of protest (Norris, 2003). However, claims about rising rates of protest in Britain might be questioned since they are based on data from a survey question that asks 'have you ever' participated in a demonstration rather than 'have you in the past 12 months …'. Nevertheless, it is certainly the case that protesters are no longer routinely dismissed as being extremists. Instead they are frequently represented as quite ordinary. As protest has allegedly become a more routine way to influence politics, the characteristics of the protesters and the general public have apparently become less distinguishable (Van Aelst and Walgrave, 2001). However, the evidence on the *extent* to which different sub-samples of protesters have become normalized is in the balance, spurring the need for reappraisal with fresh evidence.

The majority of existing studies on the normalization of protest(ers) has been reliant on established cross-national random surveys such as the European Social Survey (ESS) and the European/World Values Survey (EVS, WVS). These survey instruments have allowed scholars to examine trends in protest participation and assess the differences between non-protesters and protesters. However, these survey instruments have only a 'yes', 'no' answer option to questions asking about participation in various political acts, including in demonstrations. This makes the assumption that protesters are a homogenised group – one is either in or out. There is no differentiation between those who protest more or less. Thus,

extant cross-national surveys like the ESS, EVS and WVS preclude scholars from making careful comparisons between non-protesters and those who engage in protest to differential degrees according to their intensity and frequency of protest.

In this paper, we make a novel contribution to the literature in two ways: substantively and methodologically. Substantively, we make fresh comparisons between non-protesters and sub-groups of protesters differentiated by their intensity and frequency of protest: novices, stalwarts and those in-between, who we label 'the middle group'. Although we do not use longitudinal data to test the normalization of protest as a *process*, we follow Van Aelst and Walgrave (2001) in analysing the extent to which protesters who protest to differential degrees are normalised *during a snap-shot in time* (2009-2013). This is an important time period to study because it witnessed the rise of an anti-austerity cycle of contention (Bailey, 2014). Methodologically, we illustrate the use of propensity score stratification sample matching with benchmarking to integrate the non-random sample of protesters derived from the Caught in the Act of Protest Contextualising Contestation project (CCC) (Klandermans et al 2009) with the random sample of ESS. Although the CCC data is not strictly randomized, the survey instrument has the significant advantage over the 'yes' or 'no' questions in existing surveys. This is because they ask participants *how many times* they have participated in protest in the past 12 months *and* in their lifetimes. Note that the EVS/ WVS asks respondents only whether they have 'ever' participated in demonstrations, whereas ESS/WSS asks whether they have participated 'in the past 12 months'. Both only allow a binary (yes/no) response.

The paper proceeds as follows. We begin with a synopsis of existing literature on the normalization of protesters. From this literature, we develop hypotheses that address the

extent to which novices, stalwarts and the 'middle group' might be expected to resemble or not resemble non-protesters. Given that protesters, who we define here as participants in street demonstrations, are a small minority of the public (only around 3.8% in the UK) we take as our starting point the assumption that non-protesters are not very different from the general public since they are a significantly larger group (over 96% of the population). With this reasonable assumption in mind, we will be able to assess the extent to which different sub-groups of protesters have become normalized on the basis of the extent of differences in characteristics of non-protesters compared to protesters. We expect the extent of protest inequalities to differ across subgroups of protesters. Rather than imposing a straightforward binary measure of protest (i.e protester or not), we hypothesize in the next section that protest inequalities are smaller within some groups of protesters than others.

After presenting our hypotheses, we introduce the methodology. First, we describe the sampling protocols used by the CCC team to ensure randomization and gauge response rate bias (Van Stekelenberg *et al*, 2012). The CCC surveying protocols are to be commended for representing the state-of-the-art of protest surveying. However, despite the best efforts of the research team, there remains evidence of sampling bias through selection and response, as well as fieldwork issues that made random selection of respondents almost impossible (Walgrave, Wouters and Ketelaars, 2016). Second, we describe two methods that we have developed for integrating the CCC protesters into the ESS sample by inducing randomization for the non-random CCC sample. These two methods are a proportional weighting with a population benchmarking approach (from here on, proportional weighting) and a propensity score stratification sample matching also with population benchmarking (from here on sample matching). We use these two methods of sample adjustment in order to appropriately combine CCC protesters with ESS non-protesters.[1] These two methods are selected for

comparative purposes, but also because they represent the state-of-the-art in compensating for non-random samples. One of our objectives is to discuss how effective they are at obtaining a representative sample from which we can test hypotheses and carry out statistical inference. Third, we introduce the dependent and independent variables in our model of interest.

In our results section, we address the fit of our two samples derived from proportional weighting and sample matching by comparing the frequencies of key variables of interest with the complete ESS data-file (consisting of protesters and non-protesters). We also assess our two samples by comparing the characteristics of the 'ESS protesters only' to the raw CCC data (which by nature consists of protesters only). We subsequently present descriptive statistics under the two samples. We next present our hypotheses and present the results of the substantive model of interest. Our conclusions address both our substantive and methodological contribution. In short, our methodological conclusion is that the sample matching approach provides a dataset that more closely resembles the random probability ESS sample of the UK protesters compared to the raw data files and the proportionally weighted sample adjustments. It also goes some way towards mitigating the effects of demonstration specific contexts. From the substantive analysis on this dataset, we note that protesters become increasingly differentiated from non-protesters as they increase their degree of protest participation.

Normalization of the protester

Although the gap between protesters and the general public has generally shrunk over time, the literature shows that there are some persistent inequalities in protest participation, particularly in relation to age, gender, left-right self-placement and education. Our analysis thus focuses primary on variables that are unequivocally indicators of inequalities in protest inequalities. This means that we do not include class or income in our model

because these variables are often not important at explaining protest inequalities. Little has been written specifically about class in recent studies on protest inequalities. Instead, we take income as a rough proxy. Quaranta (2014, p.37), found that "Citizens belonging to the three income groups do not show different probabilities of not engaging in any protest action". Moseley and Morena (2010) similarly found that income is not a significant predictor of activism. In Christensen's (2014) study, "feelings about income" was a significant predictor of protest participation, but only marginally so. In Verhulst and Walgrave's (2008) study that predicts what it is that makes someone a novice, "not working" is *not* a significant predictor in their best fitting model (the effect is cancelled out by age). Moreover, several studies on protest inequalities do not include class or income in their models at all (e.g. Van Aelst & Walgrave, 2001; Stolle & Hooghe, 2011).

Even though class and income are not important predictors of protest participation, we know that protest inequalities remain, but have been reducing. At the end of the 1990s, Meyer and Tarrow (1998, p.11) put it like this: "In summary … the last thirty years have seen a generalization of the repertoire of contention across age groups, from men to women, from left to right and from workers and students to other social groups." Nonetheless, those who are younger, male, left-wing and more highly educated remain disproportionately present at protests.

Dalton's (1996) evidence also shows that a broader range of age groups are protesting as time passes, that the male-biased gender gap appears to be closing, and that the left-wing bias of protests is lessening as Conservative-led protests such as anti-abortion and mobilizations against tax occur with increasing frequency (although right-wing demonstrations are considerably more prominent in the US than in the UK). More recent

evidence (2008) also suggests that protesters are relatively normal. Gallego (2008) analysed European Social Survey data and found that men are more likely to attend demonstrations than women, but that the effect is small when other variables are controlled for. Young people demonstrate more than older people, as do the more highly educated compared to the less well educated. According to Gallego's (2008) findings, income, social class, ethnic minority and citizen status have no significant impact on participation in lawful demonstrations.

Quaranta's (2014) work also mostly concurs with Dalton's (1996) classic study, although the left-right gap – at least in Italy – appears to be increasing, not shrinking. His dependent variable is participation in two acts of protest, which are assumed, but are not guaranteed, to be signing petitions and attending demonstrations. He uses an ordinal scale of protest and Mokken Scale analysis, assuming that people participate in combinations of lower risk activities before moving to more complex or risky forms of political participation (thus, a combination of petition-signing and street demonstrating is assumed to be more likely than a combination of petition signing and occupying space as an act of protest). Comparing 1981 and 2009, he finds, notwithstanding potential issues with the assumptions of his Mokken Scale analysis, that those who participate in demonstrations (combined with petition signing) continue to be more highly educated. The gap between those who demonstrate and those who do not is shrinking for age (demonstrators are younger) and partisanship (demonstrators are more likely to be party members), but the gap is widening for self-placement on the left-right scale (demonstrators are increasingly left-wing). Torcal, Rodon and Hierro (2016) comment on the persistence of left-wing demonstrators. They find that, although those on the left protest more when a right-wing government is in power, they protest more than those on the left even under a left-wing government.

Stolle and Hooghe (2011) use evidence from the ESS, United States Citizenship, Involved and Democracy Survey (USCID) and the Barnes and Kaase *Political Action* study (1979) and additionally find that inequalities in political participation are reducing. In 1974 (drawn from Barnes and Kaase's data, using 'have you ever…'), 60% fewer women than men protested, but this had reduced to 13% in 2002 (using 'have you in the past 12 months…'). Studies that lump together different types of non-institutional political acts (e.g. signing petitions, boycotting and participating in demonstrations) find that women on aggregate participate more than men (Kern, Marien & Hooghe, 2015). It is important to contrast this with studies that look only at participation in demonstrations, which continue find that more men still demonstrate compared to women (Marien, Hooghe & Quintelier, 2010). According to Stolle and Hooghe (2011) education continues to differentiate participants from non-participants, although the gap appears to have evened out. Young people remain disproportionate, but some members of the older cohorts who protested in the 1970s continued to do so. Stolle and Hooghe's work might be criticised for drawing on different samples and for comparing questions with very different wording (see Saunders 2014) and presents yet another reason for reassessing the normalization of protest thesis based on datasets with comparable question wording. In our work, we carefully combine samples and only combine data items from different surveys when the questions have very similar wordings.

Hypotheses

Existing literature suggests that normalization of the protester might be increasing, but that it is in no way complete. There are two ways to infer hypotheses from this. The first is to view the 'normality' of protesters as a matter of degree: that is that it decreases as individuals

engage in more protest. The second is to view it as a matter of non-linear differentiation, viewing those who protest the most and novices as less special or different (more on this, below). Should we take the first approach, we would anticipate that as protesters become more engaged in protest participation they become increasingly distinct from non-protesters (Saunders, Grasso, Olcese, Rainsford & Rootes, 2012). Thus, our first hypothesis is:

H1: Under the period examined, novices are the most normalized, stalwarts are the least normalized and middle group fall between the two extremes.

Taking a contrasting non-linear approach, we might expect that the middle group are the most normal. This is because they are by far the largest sub-group (Saunders, Grasso, Olcese, Rainsford & Rootes, 2012) and are – at least theoretically – considered the least remarkable. Novices (by which we mean first-timers) face high barriers to first-time protest participation (Verlhulst & Walgrave 2009). Consensus mobilization, which involves agreeing with the aims of a demonstration, is relatively easy. In contrast, action mobilization – which involves actually going to a demonstration – is a much more difficult step, requiring considerably more commitment and dedication. Researchers have long established that, in practice, a huge majority of movement sympathizers fail to make it along to a protest (Klandermans & Oegema, 1992). This implies that there might be something special about novices. In addition, there is evidence from a survey of Dutch protesters that novices tend to mostly be younger (Verlhulst & Walgrave 2009, p.243).

By contrast, stalwarts can legitimately be thought of as highly committed to protest. The highly committed have been said to require special life circumstances to sustain their involvement (Downton & Wehr 1998). Persistent activists are known to be different in a variety of ways from those who withdraw from activism. They tend not to have married

spouses or children (Corrigall-Brown 2012) and are more likely to be male. They are better socialized into networks of activists and organizations (McAdam and Paulsen 1993; McAdam 1986; Oberschall 1973; Snow 2010:122), including in their friendship and familial networks (Stryker 1968) and they are more likely to make disproportionate use of protest vis-à-vis other forms of political participation (Dalton 2002). In their study of novices, returners, repeaters and stalwarts, Saunders et al. (2012) found that the stalwarts were much more left-wing than novices and the middle group (which is co-constituted of their "returners" and "repeaters"). Of these three groups – novices, stalwarts, and everyone in between – we might, for these reasons, realistically posit that those in between the two extremes of protests (i.e, the middle group) are least exceptional. Thus, our second and contrasting hypothesis, which draws on Verhulst and Walgrave (2009) is:

H2. Under the period examined, the middle group are the most normalized. Novices and stalwarts have the most distinct characteristics compared to non-protesters.

Data and Methods

CCC dataset

The CCC project involved collecting survey responses in the field at randomly selected demonstrations. The data for the UK were collected at thirteen demonstrations in London between 2009 and 2013. The demonstrations surveyed are shown in Table 1 along with the net sample size (n) in each demonstration. Note that although the demonstrations took place in London, protesters travelled from multiple cities across the UK to attend the demonstrations. The largest (and more effective) protests brought in protesters by coach sometimes from many miles away.

<Table 1 here>

Around 1,000 postal survey questionnaire booklets along with a pre-addressed stamped envelope were handed out at each demonstration. In the UK sample, postal survey response rates vary across demonstrations from around 9% to 35%. Aware of refusal bias from earlier protest survey projects, the CCC team devised a method to test its extent. One in every five questionnaire booklets is accompanied by a very short matched-numbered face-to-face interview. Those receiving interviews and/or questionnaire booklets are randomly selected by a team leader, known as a 'pointer', every nth row, depending on the size of the demonstration. Those who are interviewed *and* who return a face-to-face survey are considered representative of those who respond to the survey. Since face-to-face interview refusal rates are considered low, those answering only the face-to-face survey approximate a random and representative sample of those who take but do not return a questionnaire. Comparing these two sub-samples allows for an approximation of response bias (Van de Stekelenberg et al 2012).

The CCC project sampling methodology represents the state-of-the-art in protest surveying. However, Walgrave et al.'s (2016) analysis of selection and response bias in the CCC data on 51 demonstrations across Europe finds (among other things) that around one-fifth of the demonstrations surveyed were too chaotic to effectively sample randomly. Over half of the time, the field supervisor wrongly estimated the size of the demonstration, which means that the one in every n rows randomisation technique did not always ensure that everyone on the protest had a random chance of being selected. In four out of every five demonstrations, fieldwork interviewers reported losing their 'pointer'. In such a situation, the instruction to field workers was to continue on their own, potentially leading to selection bias. Earlier project revealed that this effect led to interviewers selecting contrasting sub-samples of

protesters to interview (Saunders & Rootes 2009). The refusal rate for face-to-face interviews was substantially higher than in field trials (at 13%) making it more difficult to accurately measure and account for response bias than the team had anticipated. Overall, it is known that respondents to the mail back survey are more likely to be female, that they are more interested in politics and are better educated than those who responded only to the interview.

In addition to selection and response bias, there is also bias in the selection of demonstrations to survey. For it to be feasible, the protest survey technique requires that the research team attend only demonstrations anticipated to attract over 2,000 people so that the survey team do not become overly conspicuous. Moreover, all demonstrations must be relatively 'safe' for a research team to avoid breaching research ethics agreements or jeopardising the safety of fieldworkers. To be noticed by the research team, protests also needed to be clearly advertised in advance, since researchers needed time to prepare the necessary paperwork and brief field workers. This resulted in an over-sampling of left-wing demonstrations that were both larger and predicted to be less violent.

The final UK CCC dataset had 2,533 records after deleting 29 records with missing values in the covariates of our substantive model. There were initially 303 missing values in the dependent variable that measures frequency and intensity of protest participation (novice, stalwart and middle group, see definitions below). These were imputed by nearest neighbour hot-deck imputation where the imputation classes were composed of gender, age, having a university degree, wearing a badge, signing a petition, and boycotting products (see Appendix for a description of the variables).

ESS dataset

In addition to the UK CCC dataset (2009-13) we used the UK ESS dataset combined from the years 2008, 2010, 2012 and 2014. These years more-or-less approximate the period in which demonstrations were sampled for the CCC data (2009-2013).

The UK ESS is a representative survey of all persons resident within private households in the UK aged 15 and over. The sample design is a clustered, stratified, 3-stage random probability design with unequal probabilities of selection within the household, i.e. one adult was sampled in each selected household. Given unequal probabilities of selection, the provided design weights are used in all analysis of ESS data. The final UK ESS dataset has 9,222 individuals after deleting 51 individuals due to missing values for some of the covariates of our substantive model. One of the questions in the ESS survey is: 'Have you participated in a legal public demonstration in the past 12 months?'. This question is used to differentiate protesters from non-protesters. There were 351 protesters and 8,871 non-protesters in the combined UK ESS dataset for 2008, 2010, 2012 and 2014.

To prepare the dataset used in the substantive analysis, we sought to induce randomization in the non-random CCC dataset by integrating it with the random sample of the ESS protesters (those who said "yes" to having participated in a demonstration in the past 12 months). We take into account the ESS survey design by using design weights. An overview of methods to deal with non-random samples is provided by Baker et al. (2013, see references therein). Baker et al (2013:23) discuss two approaches for correction procedures on non-random samples. They are: (1) post-stratification, which is a procedure to benchmark to known population totals (often estimated from a larger survey) and compensate for non-response and selection bias in non-random samples; (2) and sample matching to a random survey on a set of characteristics based on standard demographics as well as measures related to the variables

of interest (Rivers and Bailey 2009), in this case, variables that explain the propensity to participate in a protest. This is carried out on the basis of propensity score modelling (Lee and Valliant 2009). A further post-survey adjustment is then carried out by benchmarking to known population totals to compensate for non-response in the random sample. Whilst the first approach of using benchmarking weights to adjust non-random samples may be more common in practice when random probability samples may not be available, we expect that the second approach will perform better since it explicitly accounts for the selection bias by combining with a random sample having relevant overlapping covariates (Elliot, 2009). In many settings, estimates of interest may only be obtainable from a non- probability sample; this method can only be used to make such a sample more representative if a probability sample with overlapping covariates is available. We describe next the two approaches used in this study.

Proportional weighting

The ESS protesters are replaced with protesters obtained from the CCC dataset. We set the design weight $d_i^* = 1$ for individual *i* in the CCC dataset whilst retaining the original design weights of the ESS for the non-protesters: $d_i^* = d_i^{orig}$. An initial correction factor *C* is applied to the design weights d_i^* to correct for the proportion of protesters in the new dataset to be equal to the proportion estimated from the original ESS dataset. The average proportion of protesters across the four waves in the original ESS dataset was 3.9%. In the final step, we use post-stratification benchmarking to adjust the design weights Cd_i^* for individual *i* based on the UK Census 2011 population distribution of age group x sex. The final weights are calculated as follows: $w_i^* = [N / \sum_i Cd_i^*] \times Cd_i^*$ where *N* is the size of the post-stratum group.

Sample matching

The sample matching approach is modified from the matched sampling approach in quasi-experimental or observational designs (Holmes 2014). In these types of research designs, groups are adjusted, matched, stratified or weighted where randomness cannot be guaranteed, so that differences on confounding variables are minimised or eliminated and the samples more balanced (Rubin 1979).

Step 1: The CCC dataset is added to the ESS protesters. Let $R_i = 1$ if *i* in ESS, otherwise $R_i = 0$. Using a logistic regression model, we estimate:

$$\Pr(ESS = 1) = \frac{e^{\boldsymbol{\beta}\mathbf{x}}}{1 + e^{\boldsymbol{\beta}\mathbf{x}}} = p_i$$

where the covariates are those in our substantive model (see Table 4) in addition to other known predictors of protest participation such as voting in an election, wearing a badge, signing a petition, contacting a politician and boycotting products.

Step 2: Based on the sub-group of ESS protesters only, we sort their estimated propensities $\hat{p}_i$ and divide the ESS protesters into quintiles and determine the cut-off thresholds $\hat{p}_i^q$. In each quintile *q*, we aggregate the original ESS design weights: $\sum_{i \in q} d_i^{orig}$. Note that Cochran (1968) showed that 90 per cent of the differences between control and treatment groups in quasi-experimental studies could be reduced by stratifying on five groups. Rosenbaum and Rubin (1984) estimated that the figure was as high as 95%. In addition, the propensities estimated from a logistic regression model are used to form stratification groups (quintiles) within which sample adjustments are made as opposed to a direct linkage, and hence protects against possible misspecification of the model. Sekhon (2011) also mentions potential

pitfalls of relying on a parametric model and introduces a non-parametric approach to sample matching.

Step 3: Turning to the CCC dataset, we define the five stratification groups according to the same $\hat{p}_i^q$ from Step 2.

Each protester *i* in strata *q* in the CCC dataset is given a design weight as follows:

$d_i^{*q} = [\sum_{i \in q} d_i^{orig} / n_{ccc}^q]$ where the numerator is the sum of the original d_i^{orig} from the ESS protesters in strata *q* as calculated in Step 2 and the denominator n_{ccc}^q is the number of CCC protesters in strata *q*.

Step 4: Next, we combine CCC protesters having design weights d_i^{*q} from Step 3 with the ESS non-protesters where $d_i^* = d_i^{orig}$ refers to their original design weights. Dropping the index for *q*, the resulting design weights for the combined dataset are then post-stratified according to the UK Census 2011 population distribution of age group x sex. The final weight for individual *i* is calculated as follows: $w_i^* = [N / \sum_i d_i^*] \times d_i^*$ where *N* is the size of the post-stratum group.

In both approaches, final weights are normalised to the sample size for convenience (n=11,404).

Dependent and independent variables

Drawing on but modifying Saunders et al (2012) and Saunders (2014)[2], the dependent variable is protester status, which is a 4-category multinomial variable and includes a category for non-protesters:

0 = *Non-protester*: has not participated in a protest in the past 12 months (ESS).[3]

1= *Novice*: surveyed at their first protest ever (CCC).

2= *Middle group*: more than 1 previous demonstration in the past 12 months, but less than 6 ever (CCC).

3= *Stalwarts*: more than 6 demonstrations in the past 12 months, and more than 6 ever (CCC).

Following the literature on inequalities in protest participation, which we introduced above, the independent variables are age, self-placement on the left-right scale (recoded into an ordinal variable, with zero for 'don't know'[4]), has a university degree and gender, with the additional control of political interest, which is known to be a strong predictor of political participation. It is important to point out that the demonstrations surveyed were on issues usually associated with left-wing politics, but that sample adjustment accounts for protesters' left-wing bias. See Appendix for detail of the coding of independent and auxiliary variables. It is also crucial to note that age does not determine the sub-group of protesters to the degree one might anticipate (see endnote 2).

For the modelling we use the svy command in STATA on the weighted samples, and use the multinomial logistic regression command (mlogit).

Findings

In the first set of findings (Table 2 and Table 3) we assess the representativeness of the two sample adjustment approaches of proportional weighting to the random sample of the UK ESS. In Table 2 we compare the frequencies of key variables. A close match is indicative that the integration of the CCC protesters into the dataset after adjustment provides a reasonable

randomization of the initially non-random CCC sample. The proportional weighting approach increased the percentage of protesters slightly compared to the original UK ESS dataset but all other frequencies are similar across the samples and none of the differences are statistically significant.

<Table 2 here>

Table 3 presents a comparison of the *protesters* only in the original UK ESS with the CCC protesters after sample adjustment based on the proportional weighting and sample matching approaches. We have also included the column entitled 'CCC (raw)' to illustrate how the adjustment approaches – especially for the sample matching approach – provide a closer fit to the UK ESS sample of protesters compared to the original (raw) CCC sample. It is particularly noteworthy that the CCC respondents were very highly educated (82.5% claimed to have a degree). The sample matching approach corrects this, reducing it to 38.1%, which is more comparable to the UK ESS protesters (35.7%). This adjustment is reasonable given that the CCC team are already aware of a response bias in favour of the more highly educated (Walgrave et al., 2016). It also appears that there is a response bias *against* centrists and right-wing protesters, which have also been corrected by the sample matching approach. Table 3 shows how different the raw CCC data is from a) the UK ESS and b) the sample derived from the sample matching approach. Moreover, it illustrates that the proportional weighting is a rather naïve approach, since the adjusted sample does not differ much from the CCC raw dataset.

<Table 3 here>

The evidence shown in Tables 2 and 3 show that the proposed matching approach balances the non-probability sample and introduces randomness comparable to the ESS and allows

statistical inference to address the research questions. We turn now to our substantive model of interest. In Table 4 we present descriptive statistics comparing the distribution of the variables in our model of interest, across the four sub-populations (non-protesters, novices, the middle group and stalwarts). We compare three samples: a) the raw unadjusted combined CCC and UK ESS dataset; b) the integrated dataset after applying the proportionally weighted adjustment approach; and c) the integrated dataset after adjusting the sample using sample matching. The significant difference in gender across the four sub-samples (non-protesters, novices, the middle group and stalwarts) becomes non-significant in the sample adjusted by sample matching, which substantially reduces the female bias in the UK ESS (this is likely an artefact of the ESS sampling strategy which samples first a household and then an individual in the household). Compared to the dataset under the proportional weighting approach, the sample matching approach makes large corrections for having a university degree. Once corrected, education appears to have a linear relationship with protester status. The highly educated appear disproportionately among the more frequent activists (the stalwarts) compared to novices and the middle group. The mean age of non-protesters is most similar, in all samples, to the mean age of stalwarts. By all accounts, novices are younger and the middle group are, on average, older than novices but younger than stalwarts.

Political interest also has a linear relationship with respondents' extent of participation in protests. However, the effect is slightly tempered using the sample matching approach. Only after sample matching do stalwarts appear slightly less politically interested than the middle group. Thus, the sample matching approach appears to correct the bias that we know exists in relation to survey response in the CCC data set – generally, the politically interested have a greater proclivity to respond.

Non-protesters and novices are the two sub-groups most likely to answer ‘don’t know’ in response to a question about their location on the left-right scale. Non-protesters are markedly more likely to place themselves in the centre of the political scale, followed by novices. Whichever sample adjustment is used, extreme left positions are most common among the stalwarts who also entirely avoid placing themselves at the extreme right of the scale. The sample matching approach makes some adjustments on the left-right scale, which seems indicative of the under-sampling of right-wing demonstrators, which is hardly surprising given the left-wing nature of most of the demonstrations surveyed (Table 1). It is important to note, however, that the demonstrations involving students and ‘taking back parliament’ were among the demonstrations to attract lower numbers of left-wing protesters compared to demonstrations on other issues.

<Table 4 here>

Table 5 presents the multinomial logistic regression results with three variations of datasets. These are a) the raw unadjusted combined CCC and UK ESS dataset; b) the integrated dataset after adjustments to CCC using the proportionally weighted adjustment approach; and c) the integrated dataset after sample adjustments to CCC using the sample matching approach. The raw unadjusted combined ESS and CCC dataset is included to allow the reader to make an interpretation of model fit, which is not possible on adjusted samples using the svy command of STATA. The regression model for the raw data is also included to make it possible to observe the way in which coefficients change as we make necessary sample adjustments to induce randomness into the CCC data. Perhaps the most important thing to

notice from modelling the raw combined dataset is the relatively high R^2 score, which is 0.41. This suggests that the predictor variables do a satisfactory job of predicting protesters' status.

Using the proportionally weighted dataset, we find novices are younger than non-protesters, more politically interested, less likely to be centrist or moderate right and more likely to have a degree. The middle group are also younger, more politically interested and more highly educated than non-protesters. When it comes to left-right self-placement, they are more likely to be left-wing and less likely to be centrist or right-wing. Stalwarts are also younger and more highly educated than non-protesters. They have by far the highest positive co-efficient for political interest. In addition, stalwarts are more likely to be male. Stalwarts are the most likely of all to place themselves at the extreme left of the left right scale, and to *not* be centrist or right-wing. All of the protesters are less likely than the non-protesters to say that they 'don't know' their left-right self-placement. These results are not substantively different from the results of analysis on the raw combined file.

The multinomial regression results on the dataset under the sample matching approach tell a slightly different story. Compared to non-protesters, novices are more likely to be extreme left and have a degree. The middle group are more politically interested and tend to have a left-wing or centrist left-right self-placement. They, too, are better educated than novices. Stalwarts are the most likely to be extreme left, and if they are not extreme left, they tend to be moderate left. They are also significantly less likely to be extreme right. They are the most likely group to have a degree. Using this sample, the model does not distinguish stalwarts from non-protesters with regards to age.

<Table 5 here>

Finally, it is important to note that protests are driven by specific grievances (van Stekelenburg and Klandermans 2013), and therefore that the population of protesters in the CCC data-set may be context dependent. The same is true, however, for the ESS protesters, which are also aggregated without regard for the protest issue. To address this concern, we re-ran our model (using the raw data, proportional weighting *and* sample matching) only on demonstrators, including fixed effects for the demonstrations they attended. Table 6 illustrates that the effects of demonstration specific contexts are somewhat evened out by our sample matching approach. A student demonstration (Fund Our Future) and a vigil against hate crime (No to Hate Crime) had more demonstrators in middle group using the raw and proportionally weighted samples compared to novices. These effects disappear in the sample-matched sample. A minority of demonstrations continue, however, to predict stalwarts even in analysis based on the sample-matching. Stalwarts were more likely than novices to be in attendance in a demonstration about democracy (Take Back Parliament), a student demonstration (Fund Our Future), a women's demonstration (Million Women Rise) and an LGBT parade (Gay Pride). However, the effect and/or significance is lower, in half the cases in the sample-matched sample.

<Table 6 here>

Discussion and conclusions

This paper had two key aims. Substantively, we sought to test which groups of protesters had most comprehensively normalized. Methodologically, we sought to solve two problems inherent in protest surveying: these are non-random sampling and sampling on the dependent variable. We solved these problems by using propensity score stratification (sample

matching) to integrate the non-random CCC sample with a random sample (of the ESS) and then benchmarking to known population totals.

Instead of comparing protesters to a homogenous category of non-protesters, we defined sub-groups of protesters – namely, novices, the middle group and stalwarts. Note that, unlike other studies on the normalization of protest, we are not looking at the process of normalization *over-time*. Instead, our focus, similar to Van Aelst and Walgrave (2001), has involved looking at differentiated groups of protesters at a *snap-shot in time*. We posed two contrasting hypotheses under the period examined: Are protesters less like the general public as they participate in protest to a greater extent (H1)? Or, is the middle group the most similar to the general public given the special characteristics required in order to be a novice or stalwart (H2)? We find more support for H1 than H2.

Using the proportionally weighted sample *and* sample matching datasets, it is the novices who appear the most similar to non-protesters. The analysis on the data adjusted by the sample matching approach shows that novices are significantly more likely to be extreme left than non-protesters, but the co-efficient is markedly lower than for the middle group and stalwarts, with a much lower probability ($p=0.03$, compared to $p=0.00$). They are also more likely to hold a degree compared to non-protesters, but the p-value ($p=.001$) and the coefficient (0.76) are less highly significant and lower, respectively, than for the other two sub-samples of protesters (see Table 5). The middle group are more politically interested than novices, markedly more left-wing and better educated. Stalwarts are yet more markedly left-wing, apparently doggedly *not* right-wing and better educated still. These sample adjustments lend support to H1, which posited that: *H1: Under the period examined, novices are the most*

normalized, stalwarts are the least normalised and middle group fall between the two extremes.

But which sample adjustment works best and which do we recommend? Our descriptive analysis in Tables 2 and 3 suggests that the sample matching approach to induce randomization of the CCC dataset by integrating with the ESS has the highest standard of replication of the distribution of variables in the original ESS dataset. It appears to have corrected for under-selection of centrists and over-selection of left-wingers, especially among the stalwarts (Table 4). We are also aware of a left-wing bias in terms of the selection of demonstrations surveyed. Explicitly right-wing demonstrations were not surveyed, since those that occurred (of those that the CCC team were aware of) were small, consisting of no more than 300 demonstrators, or were billed as likely to attract violence and therefore unsuitable to survey for safety and ethical reasons. It is important to reiterate that our sample adjustments correct for our inability to survey small or violent demonstrations.

Although the CCC project has systematic protocols for avoiding survey bias, extant evidence suggests that the sampling cannot be said to be random. Here we have tested and applied novel methodologies for inducing randomization into the CCC survey to allow us to compare protesters to non-protesters. However, our sample matching technique has broad applicability beyond the CCC project data with which we illustrated it here. It might, for example, be used to combine conventional random survey samples with non-random internet survey responses. It also has applicability to observational studies in which the condition of randomisation I experimental design is not fully met and to other rare populations not well captured by cross-national random surveys. The method is highly applicable to other on-site studies, which face more pronounced problems of non-

randomness compared to protest surveys.

Our work is an interesting first-step in reassessing the extent to which different groups of protesters have become normalized. We would like to suggest ways to build further upon our study. At the moment, the results present a snap-shot in time (the period 2008-2014, aggregated). It would be interesting to present this as a longitudinal analysis. However, the number of protesters in the combined data-set for the ESS 2008-2014 is too small to take this step. The number of protesters is just 356 across the aggregated years (92 in 2008, 57 in 2010, 84 in 2012 and 123 in 2014). To take this next important step in our research requires generating panel protest surveying over future 6-year periods, and, again, creatively combining it with the ESS to ensure robust comparison of samples over time. This is certainly an interesting exercise to consider as more protest survey data becomes available over time. As it stands, we have assessed the extent to which different classes of protester have become normalized using a relatively robust methodology for dealing with integrating samples. In this regard, we make a significant methodological contribution to the literature, even though current data limitations make it impossible for us to assess this in a longitudinal fashion.

References

Bailey, D.J. (2013). Contending the Crisis: What Role for Extra-Parliamentary British Politics? *British Politics*, 9(1): 68-92.

Barnes, S.H. & Kaase, M. et al. (1979). *Political Action*. London: Sage.

Baker, R.J., Brick, M., Bates, N.A., Battaglia, M., Couper, M.P., Dever, J.A., Gile, K.J & Tourangeau, R. (2013). *Report of the AAPOR Task Force on Non-Probability Sampling* https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/NPS_TF_Report_Final_7_revised_FNL_6_22_13.pdf [accessed 16 August 2016].

Cochran, W.G. (1968). The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies. *Biometrics* 20:205-213.

Corrigall-Brown, C. (2012). *Patterns of Protest: Trajectories of Participation in Social Movements.* Stanford University Press.

Dalton, R.J. (1996). *Citizen Politics: Public Opinion and Political Parties in Advanced Industrial Democracies*. Second Edition. Chatham House Publisher.

Dalton, R.J. (2002). *Citizen Politics: Public Opinion and Political Parties in Advanced Industrial Democracies*. Third Edition. Chatham House Publisher.

Downton, J. Jr & Wehr, P. (1998). Persistent Pacifism: How Activist Commitment is Developed and Sustained. *Journal of Peace Research* 35(5): 531-550.

Elliott, M. R. (2009). Combining Data from Probability and Non-Probability Samples Using Pseudo-Weights. *Survey Practice*, Vol 2(6) https://surveypractice.scholasticahq.com/article/2982-combining-data-from-probability-and-non-probability-samples-using-pseudo-weights [accessed 09 July 2018].

Gallego, A. (2008). Unequal Political Participation in Europe. *International Journal of Sociology* 37(4): 10-25.

Holmes, W.M. (2014). *Using Propensity Scores in Quasi-Experimental Designs*. Thousand Oaks, Sage.

Kern, A., Marien, S. & Hooghe, M. (2015) Economic Crisis and Levels of Political Participation in Europe (2002-2010): The Role of Resources and Grievances. *West European Politics* 38(3): 465-490.

Klandermans, B. (1992). A Theoretical Framework for Comparisons of Social Movement Participation. *Sociological Forum* 8(3): 383-402.

Klandermans, B., della Porta, D., Giugni, M., Peterson, A., Sabucedo, J., Saunders, C. & Walgrave, S. (2009). Caught in the Act of Protest: Contextualising Contestation Survey Project. ECRP Eurocores Project Funded by the ESF, www.protestsurvey.eu.

Klandermans, B. & Oegema D. (1987). Potentials, Networks, Motivations and Barriers: Steps Towards Participation in Social Movements. *American Sociological Review* 8(1): 519-531.

Lee, S., & Valliant, R. (2009). Estimation for Volunteer Panel Web Surveys using Propensity Score Adjustment and Calibration Adjustment. *Sociological Methods and Research* 37(3):319-43.

Marien, S., Hooghe, M. & Quintelier, E. (2010). Inequalities in Non-Institutionalised Forms

of Political Participation: A Multi-Level Analysis of 25 Countries. *Political Studies*, 58(1): 187-213.

McAdam, D. (1986). Recruitment to High-Risk Activism: The Case of Freedom Summer. *American Journal of Sociology* 92 (1): 64-90.

McAdam, D. & Paulsen, R. (1993). Specifying the Relationship Between Social Ties and Activism. *American Journal of Sociology* 99: 640-67.

Meyer, D. and Tarrow, S. (1998). *The Social Movement Society: Contentious Politics for a New Century*. Boulder, CO: Rowman & Littlefield.

Norris, P. (2003). *Democratic Phoenix: Reinventing Political Activism*, Cambridge: Cambridge University Press.

Oberschall, A. (1973). *Social Conflict and Social Movements*, Englewood Cliffs, N.J: Prentice-Hall.

Quaranta, M. (2014). The 'Normalisation' of the Protester: Changes in Political Activism in Italy (1981-2009). *South European Society and Politics* 19(1): 25-50.

Rivers, D. & Bailey, D. (2009). Inference from Matched Samples in the 2008 U.S. National Elections. Paper presented at the 64th Annual Conference of the American Association for Public Opinion Research, Hollywood, Florida, May.

Rosenbaum, P.R. & Rubin, D.B. (1984). Reducing Bias in Observational Studies using Sub-Classification on the Propensity Score. *Journal of the American Statistician* 39: 33-8.

Rubin, D.B. (1979). Using Multi-Variate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies. *Journal of the American Statistical Association* 74: 318-328.

Saunders, C. & Rootes, C. (2009). Changing Biographies of Climate Change Protesters?. paper presented to the ECPR general conference, Potsdam, September.

Saunders, C, Grasso, M., Olcese, C. Rainsford, E. & Rootes, C. (2012). Explaining Differential Protest Participation: Novices, Returners, Repeaters and Stalwarts. *Mobilization*, 17(3): 263-80.

Saunders, C. (2014). Anti-politics in Action? Measurement Dilemmas in the Study of Unconventional Political Participation. *Political Research Quarterly* 67(3): 574-588.

Snow, D. (2010). *A Primer on Social Movements*. New York: W.W. Norton.

Sekhon, J. S. (2011). Multivariate and propensity score matching software with automated balance optimization: the matching package for R.

Stolle, D. & Hooghe, M. (2011). Shifting Inequalities: Patterns of Exclusion and Inclusion in Emerging Forms of Political Participation. *European Societies* 13(1): 119-142.

Stryker, S. (1968). Identity Salience and Role Performance. *Journal of Marriage and the Family*, 4 (4): 558–64.

Torcal, M., Rodon, T. and Hierro, M.J., (2016). Word on the Street: The Persistence of Leftist-Dominated protest in Europe. *West European Politics*, 39(2), pp.326-350.

Van Aelst, P. &Walgrave, S. (2001). Who Is that (Wo)man in the Street? From the Normalisation of Protest to the Normalisation of the Protester. *European Journal of Political Research* 39(4): 461-486.

Van Stekelenberg , J., Walgrave, S., Klandermans, B. & Verhulst, J. (2012). Contextualising Contestation: Framework, Design and Data. *Mobilization: An International Quarterly* 17(3): 249-262.

Van Stekelenburg, J., & Klandermans, B. (2013). The social psychology of protest. *Current Sociology*, 61(5-6), 886-905.

Verhulst, J. & Walgrave, S. (2009). The First Time is the Hardest? A Cross-National and Cross-Issue Comparison of First-Time Protest Participants. *Political Behavior*, 31 (2): 455-484.

Walgrave, S. & Verhulst, J. (2011). Selection and Response Bias in Protest Surveys. *Mobilization: An International Quarterly* 16(2): 203-222.

Walgrave, S.,Wouters, R. & Ketelaars, P. (2016). Response Problems in the Protest Survey Design: Evidence from Fifty-One Protest Events in Seven Countries. *Mobilization: An International Quarterly* 21(1): 83-104.

Table 1: UK Demonstrations surveyed

Demonstration	**Date**	**n**	**%**
National Climate March (2009)	5 December 2009	227	9.0
May Day Labour March	1 May 2010	166	6.6
Take Back Parliament	15 May 2010	160	3.2
Not to Hate Crime Vigil	23 October 2010	344	13.6
Unite Against Fascism	6 November 2010	184	7.3
Fund Our Future: Stop Education Cuts	10 November 2010	146	5.8
National Climate March (2010)	4 December 2010	351	13.9
Second Student National Demonstration	9 December 2010	96	3.8
Million Women Rise	5 March 2011	165	6.5
TUC's March for the Alternative	26 March 2011	197	7.8
Occupy London	12 November 2011	135	5.3
London Pride Parade	7 July 2012	185	7.3
The Big If	8 June 2013	177	7.0
Total		2533	100

Table 2: Protesters and non-protesters combined, comparing proportions / means in the UK ESS dataset, with the combined ESS and CCC dataset after proportional weighting and sample matching

Variable	Original UK ESS (with post-stratified weights)	Proportional Weighting	Sample Matching
Protester (%)	3.8	4.3	3.9
Auxiliary variables			
Voted in last election (%)	63.0	64.5	64.4
Signed petition (%)	32.0	34.4	34.3
Worn badge (%)	5.3	7.7	6.9
Participated in boycott (%)	19.3	21.6	21.0
Variables modelled			
Is male (%)	48.1	48.6	48.6
Has degree (%)	15.0	17.9	16.3
Age (mean)	46.5	46.6	46.6
Political interest (mean)	2.4	2.5	2.5
Left-right			
Don't know	14.3	13.5	13.4
Extreme left	3.7	4.4	3.9
Moderate left	11.4	12.7	12.1
Centrist	54.2	53.0	53.9
Moderate right	14.0	13.9	14.1
Extreme right	2.5	2.6	2.7

Notes: There are no statistically significant differences between the samples on these key variables. [5]

Table 3: Protesters only in CCC and UK ESS, comparing design weighted datasets and the raw CCC dataset

Variable	Original UK ESS (with design weights)	CCC (raw)	CCC (after proportional weighting)	CCC (after sample matching)
Auxiliary variables				
Voted in last election (%)	79.0	81.2	80.3	80.7
Signed petition (%)	86.5	88.6	88.7	86.6
Worn badge (%)	34.8	62.2	62.1	40.2
Participated in boycott (%)	53.4	67.7	67.1	53.7
Variables modelled				
Is male (%)	46.3	49.9	46.0	49.5
Has degree (%)	35.7	82.5	82.5	38.1
Age (mean)	46.4	43.1	41.2	47.2
Political interest (mean)	3.1	3.4	3.4	3.1
Left-right				
Don't know	3.4	7.7	7.8	4.0
Extreme left	9.9	23.4	23.2	10.1
Moderate left	27.8	45.8	45.3	29.4
Centrist	45.6	20.4	20.8	44.9
Moderate right	10.8	2.5	2.4	9.4
Extreme right	2.4	0.4	0.4	2.2

Note: There are no statistically significant differences between the UK ESS (with design weights) and the CCC under the sample matching approach.

Table 4. Descriptive statistics on the raw unadjusted combined CCC and ESS dataset, proportionally weighted dataset and sample matching dataset (n=11,404)

Sample	Variable	Non-protester	Novice	Middle	Stalwart	Adjusted F (p)
Unadjusted Combined Dataset	**Age (mean)**	51	37.1	43.4	44.6	***
	Has degree (%)	14.9	74.7	83.5	80.0	***
	Male (%)	44.0	47.8	49.2	55.3	***
	Political interest (mean)	2.5	2.9	3.4	3.5	***
	Left-right					
	Don't know	13.5	22.0	6.7	5.7	
	Extreme left	3.7	10.4	20.0	49.6	
	Moderate left	11.4	32.4	49.0	34.4	***
	Centrist	54.1	29.1	21.8	7.7	
	Moderate right	14.7	5.5	2.2	2.6	
	Extreme right	3.1	0.6	0.4	0.0	
Proportional Weighting Approach	**Age (mean)**	46.9	35.5	41.5	42.7	***
	Has degree (%)	15.3	74.8	83.5	81.1	***
	Male (%)	48.4	51.7	53.4	59.2	***
	Political interest (mean)	2.5	3.0	3.4	3.5	***
	Left-right					
	Don't know	13.7	22.1	6.8	5.9	
	Extreme left	3.5	10.8	19.6	50.6	
	Moderate left	11.4	32.2	48.7	33.4	***
	Centrist	54.3	29.1	22.3	7.7	
	Moderate right	14.3	5.3	2.2	2.3	
	Extreme right	2.7	0.6	0.4	0.0	
Sample Matching Approach	**Age (mean)**	46.7	40.9	45.6	48.9	n.s
	Has degree (%)	15.4	32.5	38.7	45.5	***
	Male (%)	48.4	47.4	54.7	58.0	n.s
	Political interest (mean)	2.5	2.7	3.2	3.1	***
	Left-right					
	Don't know	13.7	11.1	3.3	3.2	
	Extreme left	3.6	9.4	8.4	32.1	
	Moderate left	11.4	18.8	31.3	25.0	***
	Centrist	54.3	47.0	45.6	33.3	
	Moderate right	14.3	12.6	9.2	6.5	
	Extreme right	2.7	1.1	2.0	0.0	

Note: The adjusted F is an adjusted Chi2, except for the scale variable (age and political interest) for which the adjusted ANOVA scores are derived from a regression (age and political interest as dependent variables and sub-group as the independent variable).

Table 5: Multinomial logistic regression on the raw unadjusted combined CCC and ESS dataset, proportionally weighted dataset and sample matching dataset (n=11,404)

	Raw	Proportionally weighted	Sample matching
Non-protesters (base category)			
Novices			
Age	-0.04***	-0.04***	-0.02
Political interest	0.40***	0.47***	0.28
Extreme left	0.06	0.15	0.99*
Moderate left	-0.31	-0.30	0.40
Centrist	-1.51***	-1.54***	-0.06
Moderate right	-1.96***	-1.99***	-0.05
Extreme right	-2.21*	-2.36*	-0.72
Has degree	2.57***	2.58***	0.76**
Is male	-0.27	-0.08	0.14
Constant	-3.00***	-5.80***	-5.72***
Middle			
Age	-0.02***	-0.01***	0.00
Political interest	0.87***	0.93***	0.80***
LR			
Extreme left	1.33***	1.34***	1.46***
Moderate left	0.80***	0.72***	1.53***
Centrist	-0.98***	-1.08***	0.69***
Moderate right	-2.22***	-2.28***	0.20
Extreme right	-2.10***	-2.29***	0.50
Has degree	2.81***	2.82***	0.78***
Is male	-0.20	-0.03	-0.03
Constant	-4.05***	-6.65***	-6.45***
Stalwarts			
Age	-0.01***	-0.01***	0.01
Political interest	0.84***	0.99***	0.58
LR			
Extreme left	2.38***	2.35***	2.92***
Moderate left	0.60*	0.42	1.36**
Centrist	-1.90***	-2.08***	0.39
Moderate right	-1.93***	-2.16***	-0.14
Extreme right	-15.71	-19.81***	-12.08***
Has degree	2.71***	2.74***	1.20***
Is male	-0.49***	-0.25*	-0.21
Constant	-5.53***	-8.63***	-8.72***
Adj F		359.73***	468.67***
Pseudo R2	0.41		

Table 6: Multinomial regression on the dependent variable for protesters only (novices, the middle group or stalwarts) with demonstration effects included

	Raw	Proportionally weighted	Sample matching
Novices (base category)			
Middle group			
Age	0.02***	0.02**	0.03*
Political interest	0.50***	0.51***	0.52**
Left-right self-placement (base = don't know)			
Extreme left	1.12**	1.04**	0.23*
Moderate left	1.15***	1.13***	1.19
Centrist	0.60*	0.59*	0.69
Moderate right	0.09	0.13	0.68
Extreme right	0.24	0.29	1.37
Has degree	0.33	0.35	0.27
Is male	-0.00	-0.01	-0.37
Demo effects (base = The Big If 2013)			
National climate march 2009	-.018	-0.22	0.60
May Day 2010	0.07	0.03	0.44
Take Back Parliament 2010	0.80	0.90	2.35**
No to Hate Crime Vigil 2010	1.15*	1.11*	0.87
Unite Against Fascism 2010	-0.13	-0.24	0.28
Fund Our Future 2010	2.75***	2.67**	1.00
Climate March 2010	0.32	0.36	-0.09
2nd National Student demo 2010	0.25	0.28	0.88
Million Women Rise 2010	1.09	1.08	1.91
TUC March for Alternative 2010	0.44	0.00	0.60
Occupy 2011	-0.25	-0.31	-0.31
London Pride Parade 2012	1.09	0.99	0.33
Constant	-1.49*	-1.47*	-1.33
Stalwarts			
Age	0.02**	0.02**	0.03
Political interest	0.34**	0.39**	0.14
Left-right self-placement (base = don't know)			
Extreme left	1.82 ***	1.78***	1.11
Moderate left	0.92**	0.87**	0.84
Centrist	-0.25	-0.29*	0.20
Moderate right	-0.51	0.43*	1.27
Extreme right	-12.23	-12.40***	-12.32***
Has degree	0.27	0.35	0.42
Is male	-2.20	-0.01	-0.92*
Demo effects (base = The Big If 2013)			
National climate march 2009	-1.28	-1.28	-0.68

May Day 2010	0.02	-0.01	-1.19
Take Back Parliament 2010	2.00**	2.11**	4.46***
No to Hate Crime Vigil 2010	1.54*	1.52*	2.35
Unite Against Fascism 2010	-0.20	-0.20	1.32
Fund Our Future 2010	3.59***	3.59***	2.14*
Climate March 2010	0.56	0.67	-0.79
2nd National Student demo 2010	1.24*	1.34*	1.59
Million Women Rise 2010	1.87*	1.91*	2.92*
TUC March for Alternative 2010	-0.28	-0.35	0.78
Occupy 2011	-0.56	-0.57	-0.59
London Pride Parade 2012	2.78***	2.65***	2.13*
Constant	-2.88***		
AdjF		27.13***	18.88***
R2	0.14		

Notes: This table is illustrative of demonstration effects in the distribution of novices, the middle group and stalwarts. The demonstration effects are most notable in the raw data, and the sample matching approach renders many of those differences insignificant. Note: only demonstrations with significant differences in the distribution of the dependent variable are shown in this table.

Appendix: Variable codings

Variable name	Variable description	Coding
demonstrator	Has participated in demo in past 12 months (ESS data only)	1=yes 0=no 99= no answer
Age	Calculated age of respondent	N/A (dropped those <15 and >105 years of age)
DV	Protest participation (non-protesters from ESS, protesters from CCC)	0=non-protester 1=novice 2=middle 3=stalwart 4=ESS protester 99=missing
polint	Political interest	1=not at all (includes don't know) 2=not very 3=quite 4=very much 99=no answer
demoever	Participation in demonstrations ever (CCC only)	1= Not before 2= 1-5 times 3=6 or more times 99=missing

demo12	Participation in demonstrations in the past 12 months (CCC only)	1= Not before 2= 1-5 times 3=6 or more times 99=missing
parlia	Trust: National parliament	0=don't know 1=not at all 2=not very 3=somewhat 4=quite 5=very much 99=missing
voted	Voted in last elections	0=no 1=yes
politician	Contacted a politician (past 12 months)	0=no 1=yes
badge	Worn a badge / sticker (past 12 months)	0=no 1=yes
petition	Signed a petition (past 12 months)	0=no 1=yes
boycott	Boycotted products (past 12 months)	0=no 1=yes
satisdem	Satisfaction with democracy	0-10 scale recoded: 0 = don't know 1= 0 THRU 1 (not at all) 2= 2 THRU 3 (not very) 3= 4 THRU 6 (somewhat) 4= 7 THRU 8 (quite) 5= 9 THRU (very much)
gender	Gender	1=male 2=female
lr	Left-right self-placement	0=10 scale recoded: 0= don't know 1=0 THRU 1 (extreme left) 2= 2 THRU 3 (moderate left) 3= 4 THRU 6 (centrist) 4= 7 THRU 8 (moderate right) 5= 9 THRU 10 (extreme right)
degree	Has a university degree	1=yes 0=no

Notes

[1] We do not deny that the ESS has its own issues. Of particular concern is that it relies on self-reporting of protest participation rather than actual protest participation. In comparison, we know that the CCC protesters did actually protest. However, sample matching is

necessary not only to attempt to correct for selection bias in the CCC, but also to make statistically sound comparisons between protesters and non-protesters. Our analysis actually compares non-protesters (self-reported in the ESS) to protesters who protest with different degrees of intensity and frequency (self-reported in CCC), making the two samples broadly comparable in terms of their reliability.

[2] Note that Saunders et al (2012) and Saunders (2014) have a different classification of novices, returners, repeaters and stalwarts. In addition to splitting the middle category in two, they additionally – and crucially – failed to include the category 'non-protesters'. It is also important to note that age has barely any effect on the distribution of the 'ever' variable for extent of protest. In their earlier study, Saunders et al (2012) found that 15 per cent of young people (aged 18-25) could be classified as stalwarts, which is a higher percentage than other age groups. Since 18.2 per cent of the stalwarts in the sample are young (despite the relatively smaller numbers of young people in the sample), our operationalization does not unduly discriminate against young people.

[3] We are aware that the 'past 12 months' clause means that people who say 'no' to this question may have protested at some point in their lives. In this case, it is unfortunate that the question does not ask 'have you 'ever' participated in a legal street demonstration?'.

[4] Unlike many other studies that include left-right self-placement as a predictor, we retain (rather than drop) the 'don't know' category. This decision was made since dropping the 11.9% of the sample who said 'don't know' would result in the loss of 1352 and would potentially bias our sample. Moreover, people who answer 'don't know' are an interesting class of people unable or unwilling to classify themselves.

[5] The proportion of those participating in a legal demonstration differs by 0.1% between the ESS file and the new combined file. However, in the cleaned file (with some cases of non-protester from the ESS dropped due to missing values) the proportions are identical.

This version last updated 14/09/19.
9,736 words, all inclusive.