Can Citizens Guess How Other Citizens Voted Based on Demographic Characteristics? *

Noam Titelman *London School of Economics and Political Science*
Benjamin E Lauderdale *University College London*

How well do citizens understand the associations between social groups and political divisions in their societies? Previous research has indicated systematic biases in how the demographic composition of party supporters are perceived, but this need not imply that citizens misperceive the likely voting behaviour of specific individuals. We report results from two experiments where subjects were provided with randomly selected demographic profiles of respondents to the 2017 British Election Study (BES) and then asked to assess either (1) which party that individual was likely to have voted for in the 2017 UK election or (2) whether that individual was likely to have voted Leave or Remain in the 2016 UK referendum on EU membership. We find that, despite substantial overconfidence in individual responses, on average citizens' guesses broadly reflect the actual distribution of groups supporting the parties and referendum positions.

Introduction

Public discussions of voter behaviour sometimes suggest that social groupings align much more strongly and simply with voter behaviour than is actually the case. As Ford and Cowley (2019) lament:

> It's not that there are not under-pinning factors driving the way people vote, merely that voters are much more complicated than most discussion of this sort of analysis ever allows. Even individual voters are complex and contradictory, so this will certainly be true of any group of voters — whether we define them by place, or profession, or past vote or anything else.

It is not only pundits who tend to misperceive associations between voter behaviour and demographic characteristics. Recent studies in political science have found that citizens (Levendusky and Malhotra, 2016; Mildenberger and Tingley, 2019) as well as representatives (Broockman and Skovron, 2018) can be biased on average when assessing the aggregate political attitudes of the public. These findings are consistent with an older literature on such biases in social psychology (Pronin, Puccio and Ross, 2002; Sherman, Nelson and Ross, 2003; Chambers, Baron and Inman, 2006; Todorov and Mandisodza, 2004; Shamir and Shamir, 1997). In contrast to these findings of bias, other researchers have found

*The authors would like to thank Chris Hanretty, Tom O'Grady, and workshop participants at Durham University, as well as the editor and reviewers at this journal, for their feedback on earlier versions of this paper; **Corresponding author**: n.titelman@lse.ac.uk

that citizens' average ex-ante forecasts of aggregate electoral outcomes are often (but not always) close to accurate (Lewis-Beck and Stegmaier, 2011; Rothschild and Wolfers, 2011; Murr, 2011; Graefe, 2014; Murr, 2016; Boon, 2012), illustrating that citizens' can collectively form unbiased assessments of one another's votes in some instances. Of course there is no reason to expect a single, consistent answer to all questions of the form: "do these [citizens/representatives] have unbiased perceptions of [measure of public opinion or voting behaviour]?" The direction, magnitude, and consequences of biases may vary substantially across different contexts.

Our focus in this paper is specifically on public perceptions about the relationship between socio-demographic characteristics and vote choice. Two recent studies in the US find that people tend to "overestimate the extent to which party supporters belong the party-stereotypical groups" (Ahler and Sood, 2018) and that "evangelicals tend to overestimate the percent of Republicans who are evangelicals and overestimate the percent of Democrats who are secular (seculars exhibit more muted, but opposite patterns)." (Claassen et al., 2019).

These studies have asked respondents to make assessments at the population-level, with prompts that ask respondents for $p(X|vote)$: the proportion of people with a given characteristic (X) among those voting for a particular party ($vote$). These "compositional" questions are interesting because they tell us about the "images" of party supporters that respondents bring to mind. Ahler and Sood (2018) provide experimental evidence that misperceptions about the composition of party supporters are consequential because they increase perceived distance of individuals from the parties they do not support.

Our study complements this work by asking respondents to report their beliefs about $p(vote|X)$ instead of $p(X|vote)$. That is, instead of asking what proportion of the people who voted a given way have a particular demographic attribute, we ask what proportion of the people with given demographic attributes voted in a particular way. Where the "compositional" question asked by previous studies is useful to assessing "party images", our "behavioural" question tells us about the assumptions that individuals make about the political behaviour of a *specific* person, based on that person's demographic characteristics. Both compositional and behavioural assessments are important quantities to understand if our goal is to assess the political assumptions that citizens are making about one another.

Both of these quantities, $p(vote|X)$ and $p(X|vote)$, are likely to be difficult for respondents to report on a survey. They ask respondents to report quantities that could only be measured accurately using cross-tabulations of nationally representative surveys. In general, survey respondents struggle with questions that ask for shares of groups in the population (e.g. Joslyn and Haider-Markel, 2018; Kunovich, 2017). Mistakes in reporting probabilities can take the form of overly extreme probabilities

(e.g. Kahneman, 2011) or probabilities overly close to 50%, depending on circumstances (Baron et al., 2014; Atanasov et al., 2017). In terms of the specific information required to answer accurately, the compositional question $p(X|vote)$ is more difficult than the behavioural question $p(vote|X)$, as only the latter is typically reported in the media when presenting demographic breakdowns of election results. Indeed, Ahler and Sood (2020) propose that citizens' understandings of these proportions might be linked. They argue that citizens might be more familiar with $p(vote|X)$ and therefore recover $p(X|vote)$ by implicitly calculating (perhaps inaccurately) the relationship between the two: $p(X|vote) = p(vote|X)p(X)/p(vote)$. There are multiple ways that citizens might err in applying Bayes rule, but the most likely are by failing to implicitly multiply $p(vote|X)$ by $p(X)/p(vote)$ at all, or by holding inaccurate beliefs about the base population proportions of $p(X)$. Implicit in Ahler and Sood's argument is the idea that citizens might hold accurate beliefs about $p(vote|X)$. We test if, in fact, citizens can report accurate beliefs about this probability.

We examine citizens' perceptions about $p(vote|X)$, assessing perceptions about many social groupings (X) jointly rather than one at a time. Our two experiments consist of presenting profiles of voter characteristics (such as income, education, social class, ethnicity, religion, place of residence, age, etc). In the first experiment we ask a group of respondents to assess which party that individual was likely to have voted for in the 2017 UK election. In the second experiment we ask another group of respondents whether that individual was likely to have voted Leave or Remain in the 2016 UK referendum on EU membership. The profiles of characteristics presented were randomly selected from the profiles of respondents to the face-to-face survey of the 2017 British Election Study (BES), so we know the true reported vote choice in both the 2016 referendum and 2017 election for each treatment profile, and the treatment profiles are representative in distribution of the voters in the referendum and election. This allows us to benchmark public perceptions against the actual demographic associations in a variety of ways.

We find that on *average*, citizens' perceptions broadly reflect the actual demographic associations of voting. Across a very large number of demographic attributes and the two different vote choices, we find only a single attribute where respondents are, in the aggregate, directionally mistaken (on average respondents think that holding a university degree was associated with voting Conservative in 2017, when in fact it was associated with voting Labour). Otherwise, for both the "old" political divide of party and the "new" political divide of Brexit, respondents' assessments are responsive to variation in profiles in qualitatively correct ways, and often capture the relative strength of associations well. At the same time, while *average* beliefs track reality reasonably well, at the individual-level guesses are noisy and overconfident, and so respondents do not perform well in probabilistic assessments like Brier score. We show that this reflects the difficulties of making probabilistic assessments of what

proportion of people with a given profile will have voted in a specific way. The accuracy of respondents' perceptions increases with their level of political attention, but is not consistently predicted by any other measured characteristic of the respondent.

Whereas previous work by Ahler and Sood (2018) found that respondents caricature party supporters, and do so more when they are more interested in politics, we do not find any such tendency. While we examine a different setting (the UK rather than the US), we believe it is more likely that these different findings are the result of the different way in which we elicit respondents' understandings of how political divides intersect with social and demographic groups in the population. Compositional questions make it easier to overstate demographic associations with vote, because demographic characteristics are presented one-at-a-time. In contrast, the behavioural question that we ask requires respondents to evaluate each demographic attribute in the context of many at once, to think about a particular person with a full profile of attributes. In this context, overstating one demographic association requires ignoring others. We find that respondents do not do this, at least not on average with respect to any particular attribute. This is true even though respondents give far too many extreme responses, frequently (and implausibly) stating that certain profiles are 100% or 0% likely to have voted Leave, Remain, Conservative or Labour.

Our findings are mostly consistent with another recent study, which assesses US respondents' ability to infer the Trump/Clinton vote choices of profiles that as they reveal a mix of social/demographic characteristics as well as political attitudes (Carlson and Hill, 2021). Like their study, we find that individual-level assessments are noisy but that there are not major biases in those assessments. The inclusion of political attitudes (e.g. on abortion and partisanship) in the Carlson and Hill experiment means that their study answers a different question than ours. They find partisanship is the attribute that most increased the accuracy of guesses, followed by the profile's reported most important problem. While closely related methodologically, their experiment is designed to assess respondents' beliefs about the links between other individuals' political attitudes and vote choice, while ours is focused on the perceived links between social groups and political positions.

As Ahler and Sood (2020) observe, there are a number of mechanisms that could explain errors in citizens' reported beliefs, some of which involve consistently mistaken beliefs and some of which involve different internal logical inconsistencies in citizens' beliefs. In the conclusion, we suggest future research strategies for resolving some of the outstanding puzzles in this area, using a combination of the research design that we employ here along with those previously employed by Ahler and Sood.

The Role of Citizens' Perceptions of Group Political Behaviour

Why does it matter what citizens believe about the demographic patterns of voting? The substantial cognitive and informational demands placed on citizens by democratic institutions have led to a number of theories about the mechanisms through which they process these demands. Political sophistication is often defined as the ability to deploy political knowledge to make connections with other forms of knowledge (Luskin, 1987, 1990). One early articulation envisions citizens holding different 'levels of sophistication', varying according to their ability to recognize and judge social groups and the ideology associated with different political parties (Campbell et al., 1960; Converse, 1964). In this definition, citizens with higher levels of sophistication are those capable of making ideological judgements, while people with more moderate sophistication are those who perceive parties in a group-centric fashion, as representing a coalition of groups' interests. There is a body of literature that finds most citizens perceive politics in a more group-centred fashion than an ideological one (Converse, 1964; Kinder and Kalmoe, 2017; Kalmoe, 2019), with a general conclusion that "people are naturally more group-oriented than ideological and that, in any case, most 'ideologues' are probably familiar with the groups comprising each party's coalition" (Kalmoe, 2019).

Within the group-centric perspective, Campbell et al. (1960) differentiated between those who, when evaluating parties, only mention a single group and those who can reference multiple groups in conflict. In other words, it is possible that a more complex group-centric perspective is also related to higher sophistication. Group-centric perspectives can vary widely in their 'sophistication' according to their accuracy and the extent to which they encompass multiple, potentially overlapping, social groupings. Indeed, there are several academic (presumably sophisticated) perspectives on parties which envision them primarily as group-based coalitions, in which different interest groups come together to coordinate policy demands (Cohen et al., 2009; Bawn et al., 2012). From this perspective: "...while parties include ideological elements, collections of intense policy demanding groups define parties" (Kalmoe, 2019).

Partisanship is often conceptualized in the literature as way to ease decisions by giving cues or heuristic guidance for people, with relatively little need for information on the candidates and the electoral context (e.g. Fiorina, 2002). These cues are usually thought of as policy stances of the party and its candidates, but they may as well be cues on the social groupings of party members. –>

Opinion-based Identity and Brexit

While voting and support for political parties are often the focal political behaviour, we can expect similar patterns for other salient opinion-based divisions (Bliuc et al., 2007; McGarty et al., 2009).

Hobolt, Leeper and Tilley (2020) find that, after the 2016 EU referendum, identification as "Leavers" and "Remainers" became at least as strong as party identities. The socio-demographic determinants of Brexit voting are different from those for the party divide. While age and education are the main predictor of this opinion-based division, "measures of social class (such as income, occupation and housing tenure) continue to matter more for partisan identities than for Brexit identities despite sharp falls in class voting in Britain in recent decades" (p.14). This is consistent with previous research on the determinants of Brexit vote that has found that remain voters tended to hold social liberal values, and also were more likely to be younger and hold more educational qualifications, while leave voters tended to hold social conservative values, and tended to be older and hold fewer educational qualifications (e.g. Alabrese et al., 2019; Goodwin and Heath, 2016; Dassonneville, 2016). There are reasons to believe these social cleavages became increasingly relevant partly because of generational changes in the British electorate, which has become more educated and racially diverse (e.g. Sobolewska and Ford, 2019). The Brexit divide seems to rival party in terms of their potential to shape citizens' views about the political alignment of social groups. Hobolt, Leeper and Tilley (2020) find that in terms of trait stereotype—positive in-group perception and negative out-group perception—the Brexit divide might be stronger than the partisan divide.

Thus, past research gives us reason to suspect that citizens' own social and political identities and their perceptions of the social and political identities of others are interrelated. This makes it important to know when perceptions are shaped by real demographic patterns, as well as in which circumstances they overstate or caricature those patterns (Ahler and Sood, 2018; Claassen et al., 2019). At the same time, people hold multiple political identities, and these may mobilize distinct aspects of their social identities. The existence of a long-standing (but evolving) party system in the UK, alongside the more recent "pseudo-party" system of Brexit vote and identity, provides a unique environment to examine how citizens understand the complex demographic associations with political behaviour.

Data and Methods

Our experiment consists of presenting real profiles of voter characteristics and then asking respondents to assess (1) which party that individual was likely to have voted for in the 2017 UK election or (2) whether that individual was likely to have voted Leave or Remain in the 2016 UK referendum on EU membership. The profiles of characteristics presented to respondents were those of individuals randomly selected from the 2017 British Elections face-to-face Survey (BES).[1] Because each "treatment

[1]BES respondent profiles were randomly sampled with the probability of sampling proportional to the BES 2017 with result weights (*wt_vote*). This ensured that the profiles presented to respondents of the experiment were nationally representative of British voters, based on self-reported turnout,[^whynotvalidated] in the 2017 election. These weights are constructed using demographic weights targeted to the voting eligible population and weighting to Great Britain turnout and vote results. Not exactly the same people voted in the 2016 referendum and the 2017 election, so this means that the profiles were slightly

profile" corresponds to a real BES respondent, each sampled profile has a true vote choice in both the 2016 referendum and 2017 election, and it is possible to benchmark public perceptions against reality.[2]

This experimental design follows a trend towards the use of more complex survey designs, particularly involving multidimensional randomisations of complex treatments. The most widely applied such designs are conjoint experiments, which independently randomise a large numbers of attributes in order to enable estimation of *average marginal component effects* (Hainmueller, Hopkins and Yamamoto, 2014). Our design is not a conjoint experiment, because the attributes are not independently randomised, instead we randomly select full profiles of attributes from a population survey (the BES) using population weights. This means that the profile attributes we present to respondents are effectively sampled from the population joint distribution of those attributes.

There are two reasons that we do not use a conjoint design here, one of which is general and one of which is specific to our application. In general, one threat to the external validity of conjoint experiments comes from the potential for the independent randomization distribution to consequentially shape the results (De la Cuesta, Egami and Imai, 2019). Since the *average marginal component effects* (AMCEs) average over the treatment distribution, an independent distribution may not be innocuous for the external validity of any findings. One manifestation of this problem is the fact that with independent randomization, implausible or impossible combinations of attributes may occur. The more specific reason that we adopt this design is that, unlike the many conjoint experiments which interrogate voter preferences, in our application there is a right answer. We know the votes of the individual respondents to the BES; we would not know the votes of hypothetical profiles generated by randomising individual attributes.

The cost of randomising the attributes at the full profile level, rather than the individual attribute level, is that differences in mean response, comparing all responses to profiles with different attribute levels, lose their causal interpretation (they are no longer unbiased estimators of the AMCEs). We can, nonetheless, form *model-based* rather than *design-based* estimates of the causal effects of respondents seeing particular attribute levels, through the use of regression. For the purposes of this

unrepresentative with respect to 2016 referendum voters, however not to an extent that is consequential for our purposes.

[2]Gender and region did not present missing values (they are used for the sampling process). To deal with missing attributes of the voters' profiles, due to non-response, two strategies were followed. For all attributes, apart from ethnicity and religion, missing values were randomly imputed using STATA to fill in missing values using a multivariate imputation through chained equations (MICE). In other words, we imputed multiple variables iteratively via a sequence of univariate imputation models, one for each imputation variable, with fully conditional specifications of prediction equations (mi impute chained command in STATA). This imputation strategy relies on assumptions to model the relationship between variables. Specifically, multiple linear regression was used for age, logistic regression for home status, subjective class, and subjective family class, and ordinal logistic for education and income. Gender, region, and vote (EU referendum vote for Brexit experiment and General Elections vote for the party experiment), where used as predictors. For ethnicity and religion, "unknown" category was included in the experiment as a possible level of these attributes. Figure 1 in the online appendix details missingness patterns before imputation. There are only 3.1% missing values for the Brexit experiment and 3.2% for the parties' experiment and these are mainly concentrated in the income attribute, which is strongly predicted by other attributes, such as home status. We are therefore confident this imputation does not distort the profiles' distribution in any consequential way.

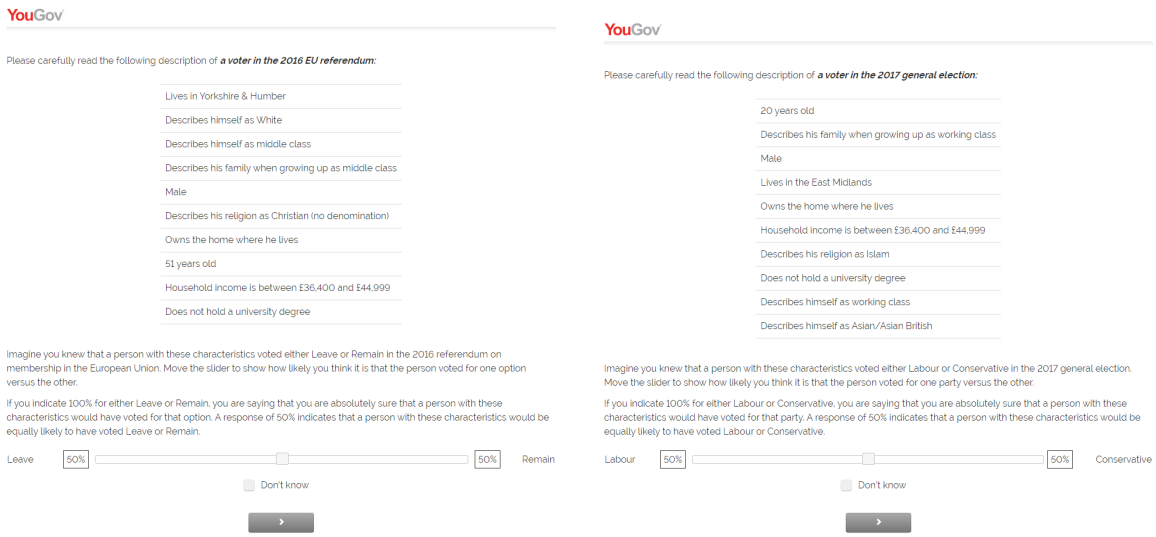

Figure 1: Survey prompts with example profile for Brexit experiment (left) and party experiment (right).

experiment, it makes sense to sacrifice having simple experimental comparisons for all attributes in exchange for having a meaningful external benchmark. Crucially, because the full profiles are themselves randomly assigned to respondents, the design still allows us to assess the causal effects of different attributes appearing in the treatment profiles, subject to modelling assumptions about how the effects of different attributes aggregate.

Our experiment was fielded by YouGov in June 2019. The prompt for the Brexit experiment first asked the respondents to carefully read a table with 10 demographic attributes of the voter. It then asked the respondent to assign how likely it is this voter voted for either Leave or Remain in a slider (that automatically made sure the sum of the two percentages resulted in 100%). The slider allowed integer percentage responses from 0 to 100. The party experiment prompt followed a similar format with the addition of making explicit that the profile voter had cast his or her vote for either Labour or Conservative. Immediately above the slider, the prompt included a statement that aimed to explain to respondents how the scale works. Specifically, it explained that choosing any value other than 0 or 100% implies uncertainty. For the Brexit experiment, this read "If you indicate 100% for either Leave or Remain, you are saying that you are absolutely sure that a person with these characteristics would have voted for that option. A response of 50% indicates that a person with these characteristics would be equally likely to have voted Leave or Remain."

The prompt was repeated three times per respondent with different profiles. The order in which the attributes were listed, and which ends of the slider corresponded to Leave, Remain, Conservative or Labour, were randomised per respondent. 1694 respondents were recruited for the Brexit experiment and 1688 respondents for the party experiment. We use sample weights provided by YouGov that make the data nationally representative for the British population on standard demographic and past vote

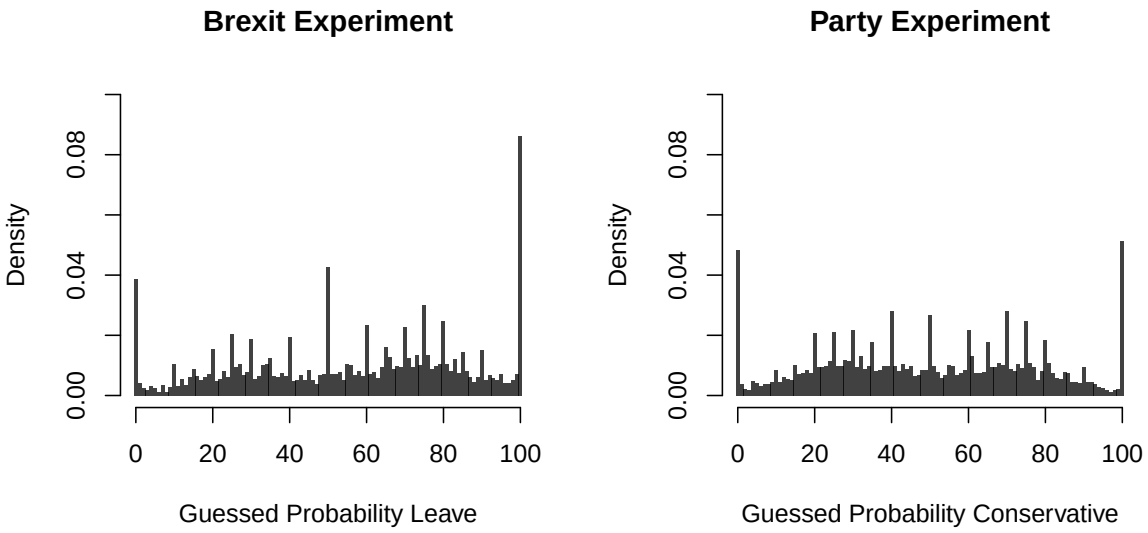

Figure 2: Distributions of guessed probabilities for voting Leave versus Remain (left), and Conservative versus Labour (right).

variables.

Determinants of Respondent Guesses

Figure 2 shows the distributions of guessed probabilities for voting Leave versus Remain, or Conservative versus Labour. Despite our efforts in the survey prompt to make clear that 0% and 100% responses are excessively strong statements, as they imply no uncertainty whatsoever, they remain common responses to the prompt.

Because the experimental profiles were randomly sampled from the BES, we can benchmark general perceptions on average across all profiles. Do respondents accurately perceive the general tendency of voters in the UK to support Labour versus the Conservatives and Leave versus Remain? The average guess for the party experiment is 49.8% Conservative vote (95% interval 48.8-50.8), slightly lower than the true value of 51.4% of the two-party vote and the proportion of the BES profiles which corresponded to Conservative voters, which was 51.5% (95% interval 48.5-54.4). In the Brexit experiment, the overall average guess is 56.5% Leave vote (95% interval 55.4-57.5), which is slightly greater than both the true value of 51.9% and the proportion of the BES profiles which corresponded to Leave voters, which was 50.3% (95% interval 47.6-53).[3] While these differences are statistically significant, they are not substantively large.

[3]The BES estimates for our Brexit experiment are slightly smaller than the referendum result because the sample is weighted to correspond to general election voters rather than those who voted in the referendum. Thus, on average, respondents perceived profiles as being more likely to correspond to Leave voters than they ought to have, and were very close to accurate for Remain voters.

Differences in Mean Guesses By Respondent Vote and Profile Vote

As an initial check on whether respondents are able to distinguish at all between Leave and Remain or Conservative and Labour profiles, we can calculate the average response given the true votes of the profiles that respondents observed. We find that the average guessed probability of a Leave vote was 52.7 (51.5%-54%) for BES profiles that actually voted for Remain, and 60.1 (58.8%-61.3%) for those that actually voted for Leave. We find that the average guessed probability of a Conservative vote was 46.6 (45.4%-47.9%) for BES profiles that actually voted Labour, and 53 (51.8%-54.2%) for those profiles that actually voted Conservative. Thus, we see clear evidence that responses were, on average, affected by information in the profiles in a way that made them more accurate than would have occurred if respondents were guessing without reference to the profile. They were more likely to guess higher probabilities of a Leave vote when the profile really was a Leave voter rather than a Remain voter; they were more likely to guess higher probabilities of a Conservative vote when the profile really was a Conservative voter rather than a Labour voter.

We can ask a similar question with respect to respondents' own vote history. Since the treatment profiles are randomly assigned to respondents, any difference that we see as a function of respondents' own vote history must be an indication of bias in how respondents perceive the votes of other citizens. We find that for both the party experiment and Brexit experiment there are small, but statistically significant differences predicted by respondents' previous vote. In the party experiment we find that respondents that voted for Labour in the 2017 general election underestimated the probabilities of Conservative vote, with an average guess of 47.2% (95% interval 45.7-48.7) while respondents who voted for Conservative were, on average, unbiased in their guesses, with an average guess of 51.4% (95% interval 50-52.9). In the referendum experiment, all respondents tended to overestimate Leave vote. However, this bias was stronger among leave voters, with an average of 59.3% (95% interval 57.9-60.7) versus an average of 54.5% (95% interval 53.1-55.9) for those who voted remain. While both experiments provide evidence of a tendency for respondents to make guesses about the profiles that tend slightly towards their own positions, the differences in average guess by respondents' own votes are still smaller than the differences by the profile's true vote.[4]

Differences in Mean Guesses By Profile Attribute

Because the profiles in our experiment are drawn from the real joint distribution of voters, we can analyse accuracy, subsetting by profile attribute values and comparing to the BES. The cross-tabulated BES distributions of vote by these attributes provide an appropriate benchmark for actual voting behaviour

[4]This may seem like a low standard, but respondents know their own vote and not the profile vote.

among individuals with these attributes, averaging over the actual distributions of other attributes that tend to come along with the attribute we are focusing on. Thus, for example, we can compare the guessed proportion of Leave voters for profiles with a university degree in the experiment ("*Guess*") to the proportion of Leave voters among (weighted) BES respondents ("*BES*") with a university degree. We are additionally able to compare to the true result of the election/referendum ("*Real*") when we subset by region.

Note that while it facilitates benchmarking, the non-independent randomization of profile attributes means that we cannot conclude from this analysis that it was a specific grouping variable that *caused* respondents to guess differently with respect to vote. It could be that it was other attributes, themselves associated with that attribute in the UK population, which led respondents to make different guesses.

In general, Figures 3 and 4 show that respondents' guesses are responsive to differences between groups. While on average guessed Leave vote is slightly too high, the differences between class groups, regions, income groups, home ownership status, gender, ethnicity, education and age are all in the right direction and are close to the correct magnitude for many attributes. Respondents appear to be substantially under-responsive to differences by age, income and ethnicity. In the party experiment, nearly all of the differences between groups are once again in the correct direction, with the sole exception of education. Respondents thought that profiles with university degrees were more likely to be Conservatives than those without, when in the BES the relationship goes the other way. Here, there is a substantial underestimation of age and regional differences, while the association with income is very close to correct.

Regression Analysis of Guesses by Attributes

These one-attribute-at-a-time analyses tell us about the general tendency of respondents to hold accurate perceptions of profiles with different attributes. But because profile attributes are correlated in the UK population, and therefore also in our experimental treatment distribution, the one-at-time analysis does not tell us the extent to which respondents are changing their responses due to particular profile attributes. It could be that respondents only perceive the importance of some of these attributes, change their responses in response only to those attributes, but nonetheless appear responsive to other attributes which are correlated with the ones that they know about. While our design's non-independent randomisation sacrifices experimental balance of profile attribute effects, the experimental design still rules out omitted variables and we can identify the causal effects of attributes subject to modelling assumptions (De la Cuesta, Egami and Imai, 2019), which are in our analysis the assumption of additivity of the attribute effects on a logit scale. The possibility of at-

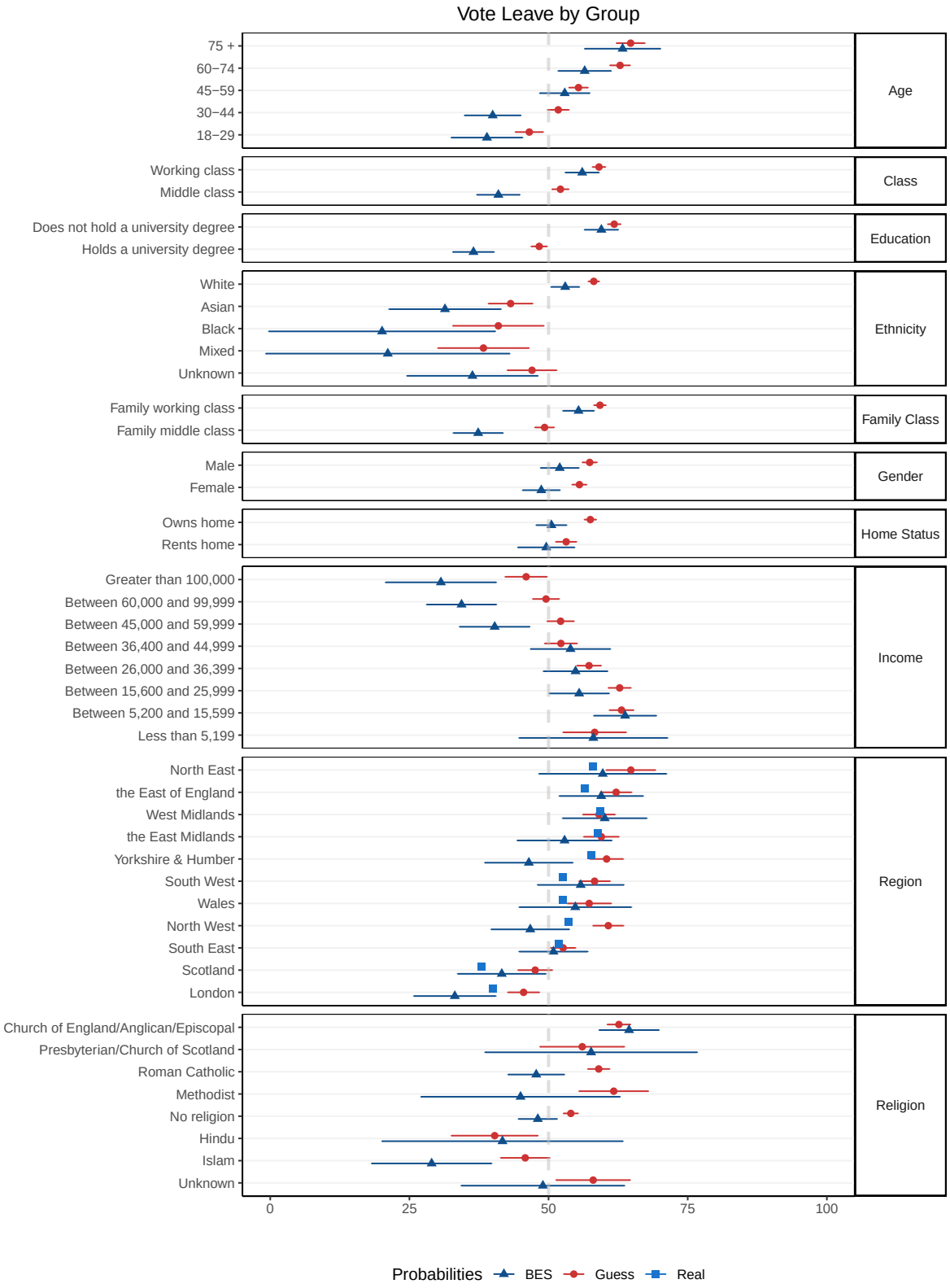

Figure 3: Average guess of vote versus BES estimates and known results by profile attribute for Brexit experiment.

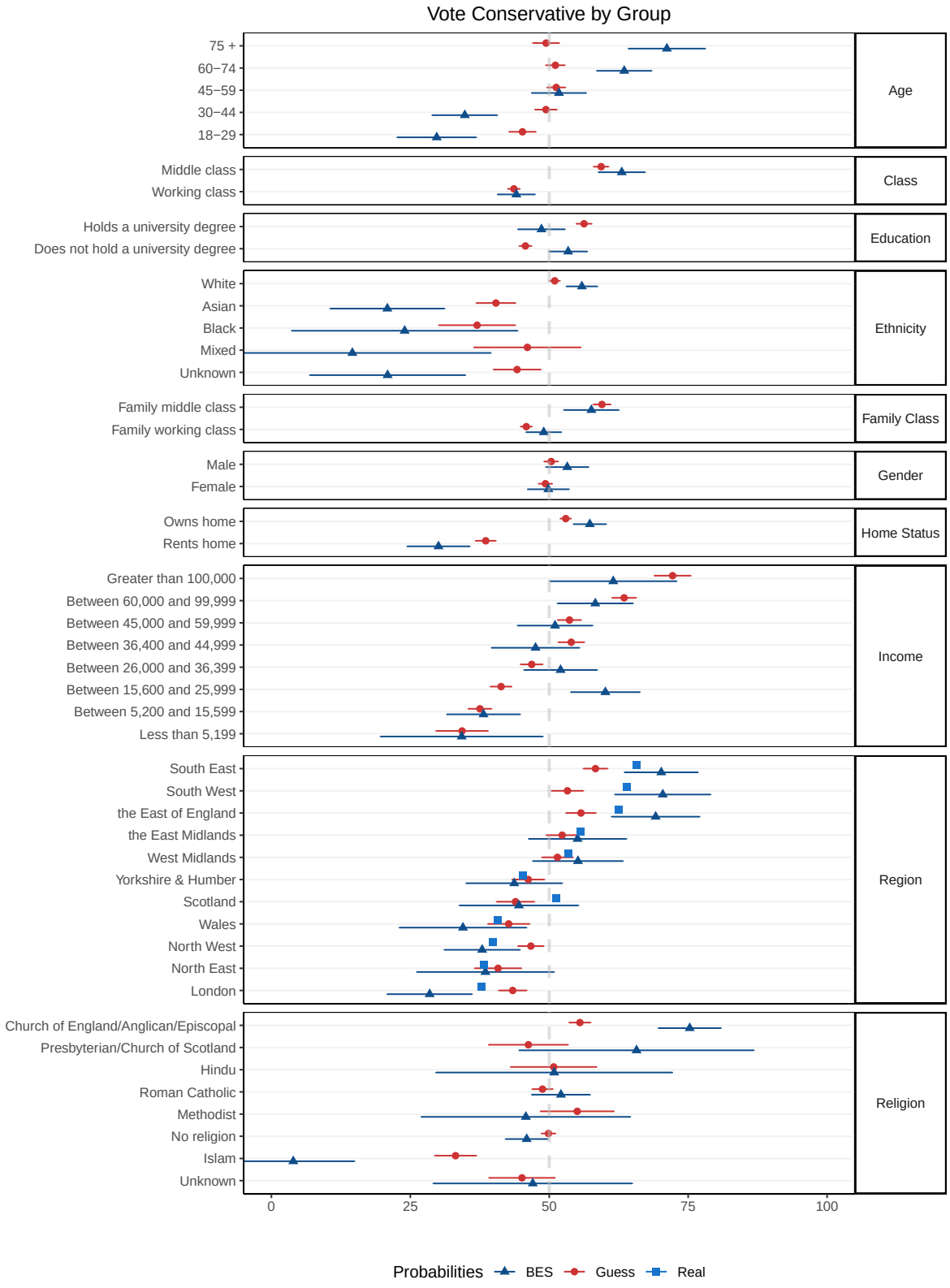

Figure 4: Average guess of vote versus BES estimates and known results by profile attribute for party experiment.

tribute confounding motivates moving to a multiple regression analysis of responses, to attempt to distinguish which of the profile attributes are influencing respondents.

The relevant benchmark for a regression model predicting respondent guesses as a function of profile attributes is the equivalent regression model predicting vote choice among BES profiles. In the analysis below, we use as modelling assumptions a (fractional) logistic regression for the guess (rescaled to the $[0, 1]$ interval) and a logistic regression for the binary vote choice, so that the coefficients are directly comparable.[5]

The individual coefficients shown in Figures 5 and 6 can be interpreted in a causal way. In other words, they represent the expected change in the odds of guessing a probability, by an average respondent, brought upon by a change in the presented profile from the base category to the measured category, averaged over the distribution of the other attributes. For example, the coefficient for "male" represents the expected change in odds of a guessed probabilities, for the average respondent, of being presented a random male profile rather than a random female profile, holding all other attributes constant. Our findings follow largely similar patterns to the single attribute analysis from before. There are some exceptions: we see responses tracking regional differences in the single attribute analyses in Figures 3 and 4, but Figures 5 and 6 suggest that this is mostly because of demographic variation by region as opposed to direct effects of the region label. Overall, the magnitudes of the partial associations are either close to correct or underestimated, but only in the case of education in the party experiment is the association significantly in the wrong direction. Respondents are, on average, responsive to most of the attributes provided in the experiment, holding constant all of the others.

Comparison of Predicted Probabilities

If we use both of these models to construct predicted probabilities for the BES profiles, we see that the predicted probabilities are correlated to a substantial degree. For the Brexit experiment, the predicted probabilities constructed using the BES vote data and using the experimental guesses are correlated at 0.82. For the party experiment, the equivalent correlation is 0.54. The fact that the coefficients from the model fit to the guesses tend to be attenuated relative to the model fit on the BES vote choice data means that the predicted probabilities from the former are also attenuated with respect to the predicted probabilities from the latter (see Figure 7).

[5] We obtain very similar results using a linear probability model for both the guesses and the BES vote data, however this model does lead to invalid predictions for the binary vote choice for some profiles.

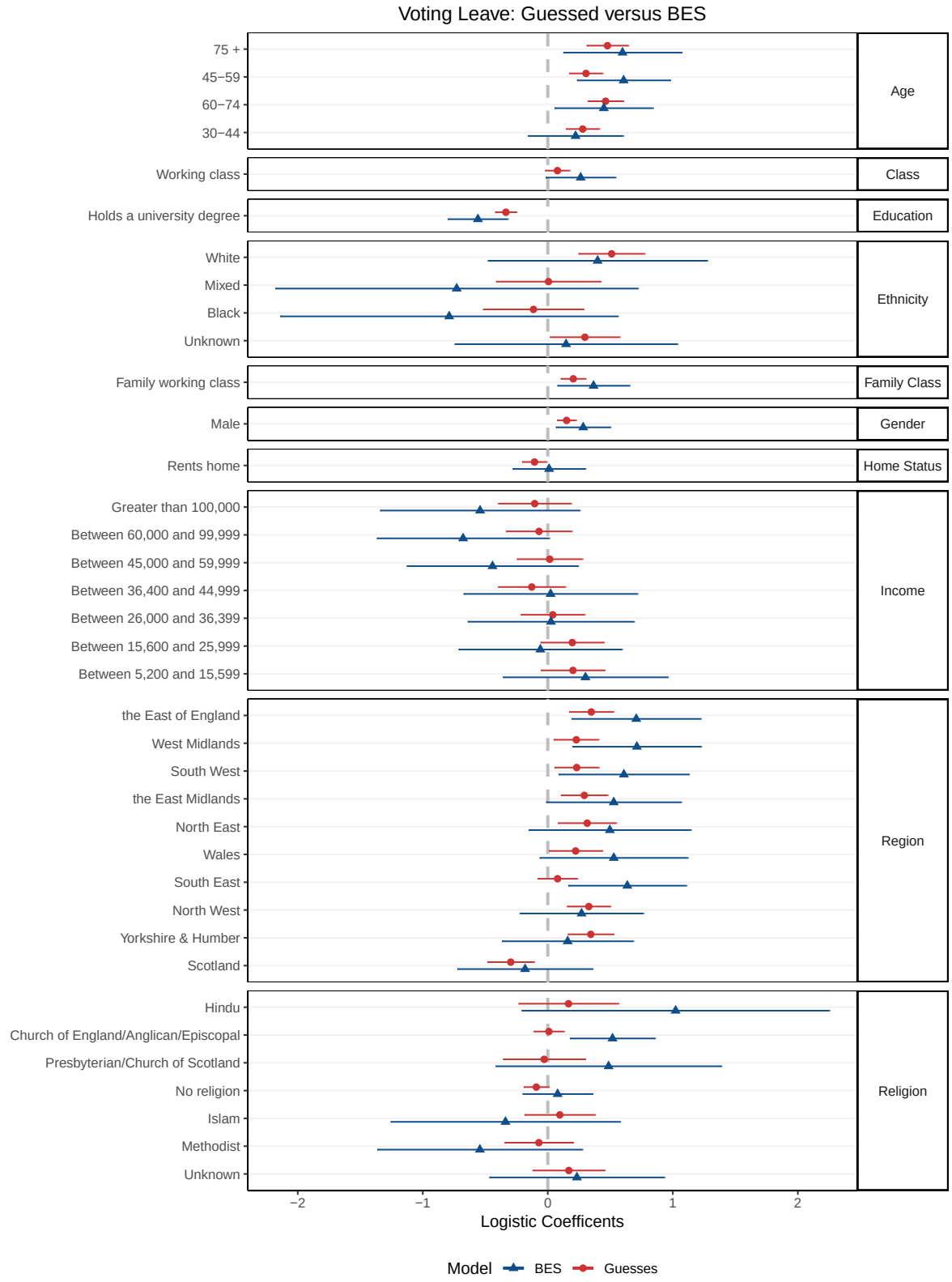

Figure 5: Regression coefficients for guess of vote versus BES estimates by profile attribute for Brexit experiment.

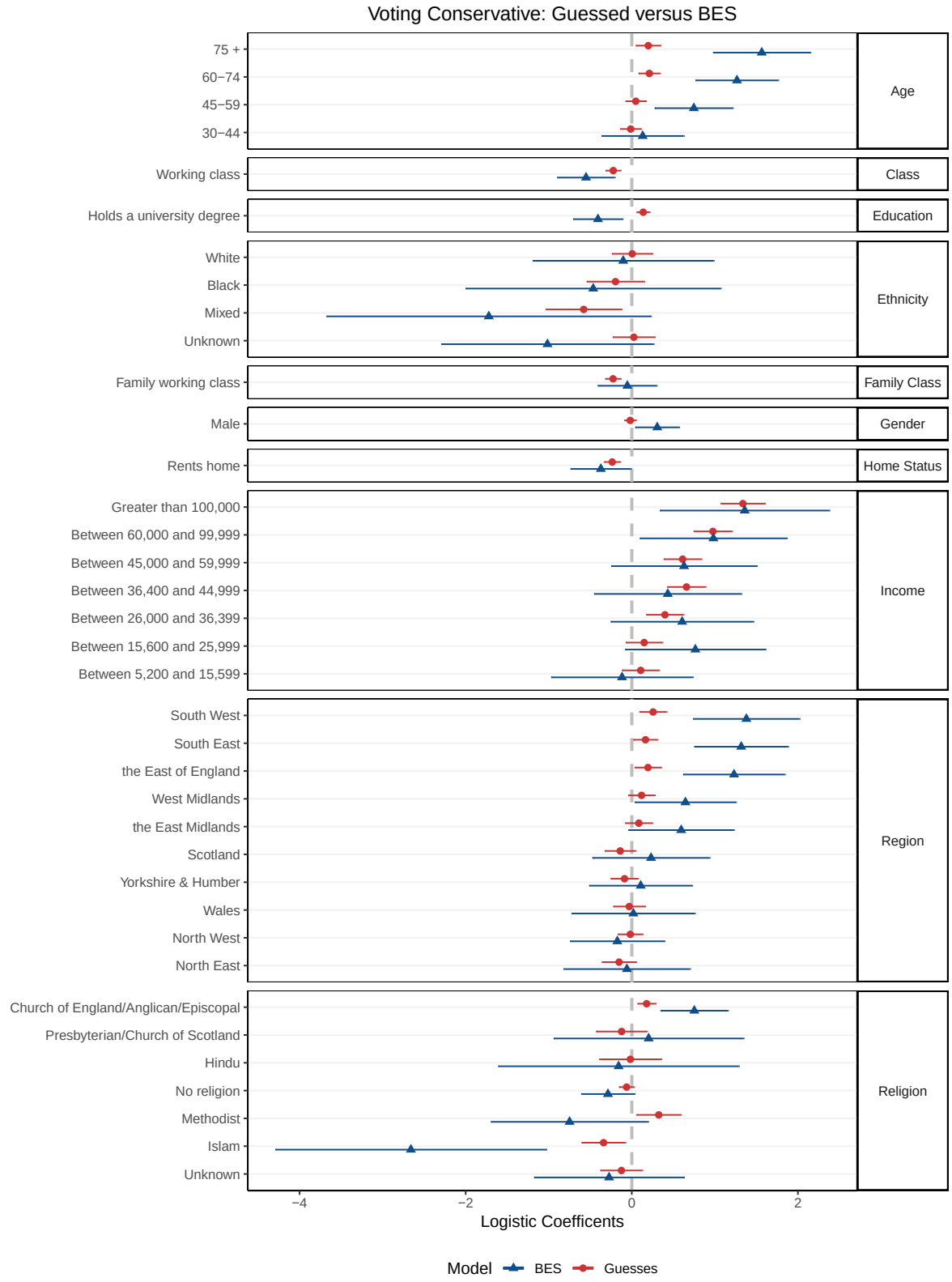

Figure 6: Regression coefficients for guess of vote versus BES estimates by profile attribute for party experiment.

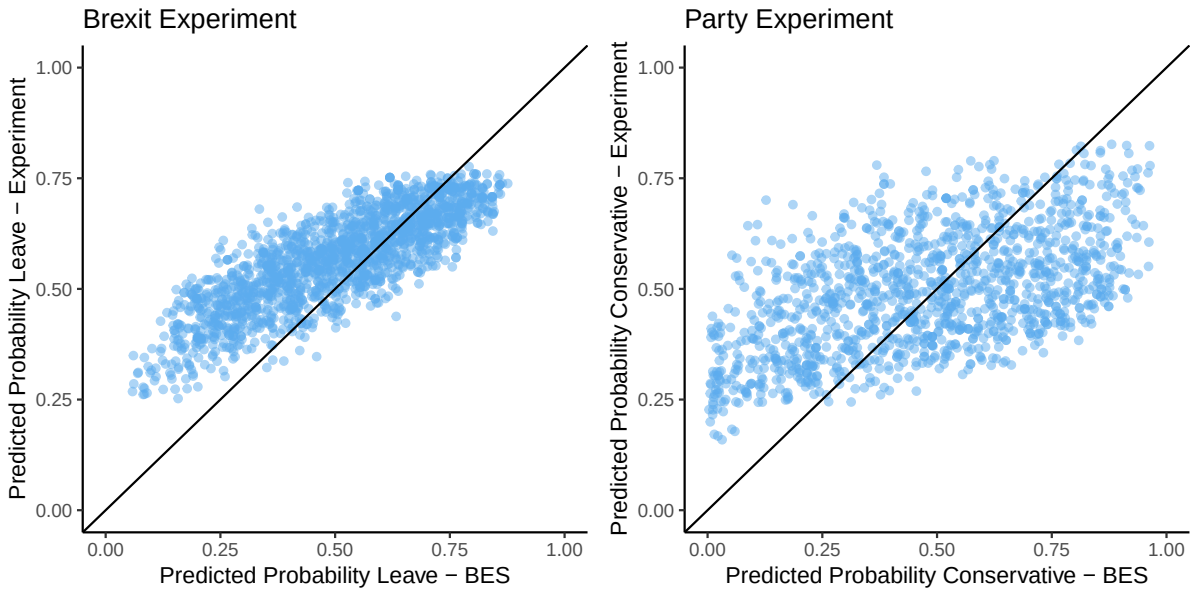

Figure 7: Predicted probabilities based on experimental responses as a function of predicted probabilities based on BES vote choice.

Determinants of Respondent Accuracy

Thus far, we have focused on whether respondents' guesses vary in the right ways given variation in the profiles, on average. But average variation in the profiles is not the only variation of interest. Is the good average performance the result of high quality individual-level guesses, or simply a lot of idiosyncratic error that cancels out? Figure 8, by comparison to Figure 7, shows that there is a great deal of idiosyncratic error. Which respondents to our experiment are more or less able to provide accurate responses? There are many ways to answer these questions, but here we use two measures of the accuracy of guesses, one which assesses the quality of the percentages reported by respondents as probabilistic forecasts, and one which assess only the direction of the guess.

First, we use the Brier Score, a tool from forecast evaluation, to assess respondents guesses as probabilistic predictions (Brier, 1950). If N is the total number of predictions, f_i is the probability reported by a respondent and o_i is the true vote of the profile shown to that respondent (which may take the values of 1 or 0):

$$\text{Brier Score} = \frac{1}{N}\sum_{i=1}^{n}(f_i - o_i)^2$$

Smaller Brier scores imply better predictions. Here, the measure enables us to assess the accuracy of respondents' guesses about the referendum and election vote by comparing their prediction to the actual votes associated with the voter profile that they observed. A convenient feature of the score is that it is simply an average of a quantity that we can calculate for each response. This means

that in addition to calculating the score overall, we can fit regression models for $Y_i = (f_i - o_i)^2$ to model how the Brier score, which is to say predictive accuracy, varies as a function of respondent characteristics. Note that this depends on only the guess and the true value for each response to our survey experiment, so we can model this quantity as a function of profile characteristics, respondent characteristics, or both.

Second, we use "correct dichotomised guesses" to assess respondents' guesses in a way that reduces sensitivity to their ability to use a probability scale effectively. Here, if the profile is actually a Leave voter, we count any guess from 51% Leave to 100% Leave as correct, a guess of 50% as half correct, and any guess from 0% to 49% Leave as incorrect. This approximates the assessment that we could have done if we had asked respondents simply for their best guess, rather than for a probability. Merely assessing whether the respondent's guess was in the correct direction makes sense if one is concerned that respondents understand that probabilities above 50% imply that an option is more likely than the alternative, but find it difficult to express the degree of confidence using a probability scale.

The overall Brier score for all responses (using survey weights) is 0.302 for the Brexit experiment and 0.291 for the party experiment. In both cases this is worse (higher) than the score of 0.25 that results from simply guessing 50% for every profile in both experiments. This is not surprising given that many respondents provide 0% and 100% responses, which are always overly confident probabilistic assessments given the limited predictive power of the profile attributes that respondents saw in the experiment. To generate a benchmark for what good guesses would look like in this task, we can compare the guessed results to the Brier score obtained by using the BES predicted probabilities as f_i. Any remaining difference can be attributed to either the respondents' lack of knowledge or their difficulty at communicating it as a probability. These benchmark Brier scores are 0.088 and 0.102 for the Brexit and party experiments respectively. These values are far better (lower) than the respondents achieved as well as being substantially better than what would result from guessing 50% on all profiles, because the profile variables are moderately predictive of vote choices in both experiments.

We can assess the extent to which poor reporting of probabilities is the problem by analysing the proportion of correct guesses when we dichotomise the guesses as described earlier. We find that, under this criterion, 56.3% (95% interval 54.6-58) of respondents in the Brexit experiment correctly guessed the vote of the respective profile. Similarly, 56.4% (95% interval 54.7-58.1) of respondents in the party experiment guessed correctly. If we similarly dichotomize the fitted probabilities from the benchmark model fit to the BES data, we find that 63.4% (95% interval 61.7-65.1) of profiles in the Brexit experiment and 59.7% (95% interval 58.1-61.4) in the party experiment could have been guessed correctly based on the dichotomised probabilities from the logistic regression fit on the BES data. By

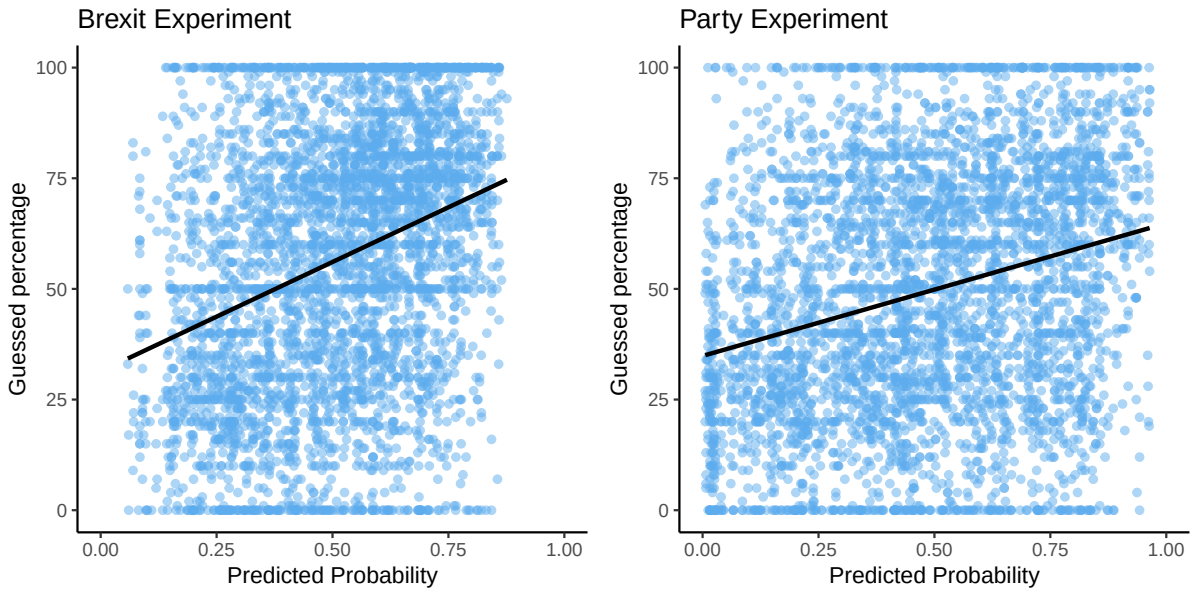

Figure 8: Guessed percentages for each response in the experiment as a function of the predicted probability for the experimentally provided profile using the BES vote regression model.

this standard, respondents perform reasonably, given the limits of what was possible using a basic demographic model with the data that they were presented with. The fact that the guesses look so much better when assessed dichotomously reinforces the point that the poor predictive performance by Brier score derives in large part from the fact that people struggle to think probabilistically or to report their beliefs in this way (e.g. Kahneman, 2011; Baron et al., 2014; Atanasov et al., 2017).

Respondent-level Predictors of Accuracy

In Table 1 we report the results of a regression predicting Brier scores and correct dichotomous guess proportions, for both experiments. The strongest source of respondent-level heterogeneity across the two experiments is that respondents who pay more attention to politics tend to do a much better job at guessing the probabilities of someone voting in a given way. Going from the lowest (0) to the highest (10) level of attention is associated with an increase of 7.5 and 12.4 percentage points in the proportion of profiles with the correct dichotomised guess in the Brexit and party experiments, respectively and all else equal. The fact that we see this association in both Brier scores and correct dichotomised guess tells us that it is primarily an association with knowledge, rather than with the ability to accurately report probabilities.

Political attention is the only respondent attribute that is consistently and strongly predictive of Brier scores as well as correct dichotomised guesses across both experiments. Higher educational attainment is associated with better (lower) Brier scores on the Brexit experiment, but not the party experiment. In both experiments, the region where respondents make the worst guesses by Brier

score, all else equal, is London. This difference is only marginally significant from other regions, and is not present in the party experiment when assessed by dichotomised guess, but it is plausible that people in London might have a poorer understanding of how people around the UK vote than do respondents elsewhere, simply because London is a bit of a political outlier among UK regions.

	Brier Score		Correct Dichotomized Guess	
	Brexit Exp.	Party Exp.	Brexit Exp.	Party Exp.
Intercept	0.414***	0.342***	0.453***	0.522***
	(0.045)	(0.048)	(0.073)	(0.080)
Political Attention	−0.008***	−0.009***	0.008*	0.012***
	(0.002)	(0.002)	(0.004)	(0.004)
Party Vote: Labour	−0.003	−0.001	0.024	0.002
	(0.013)	(0.013)	(0.021)	(0.021)
Party Vote: Liberal Democrat	−0.033	0.026	0.052	−0.014
	(0.022)	(0.021)	(0.036)	(0.035)
Party Vote: SNP	0.030	−0.025	−0.060	0.061
	(0.038)	(0.036)	(0.061)	(0.062)
Party Vote: Plaid Cymru	−0.109	0.026	0.105	−0.192*
	(0.107)	(0.068)	(0.172)	(0.115)
Party Vote: UKIP	0.086**	0.055	−0.041	−0.058
	(0.039)	(0.040)	(0.062)	(0.068)
Party Vote: Green	−0.016	−0.029	0.026	0.036
	(0.042)	(0.037)	(0.067)	(0.062)
Party Vote: Other	0.074	0.006	−0.158	−0.013
	(0.071)	(0.062)	(0.115)	(0.105)
Party Vote: Don't Know	−0.009	−0.017	0.024	0.031
	(0.037)	(0.035)	(0.060)	(0.059)
EU Ref Vote: Leave	0.013	0.023*	0.001	−0.021
	(0.013)	(0.012)	(0.020)	(0.020)
EU Ref Vote: Did not vote	0.029	0.027	0.008	−0.012
	(0.028)	(0.024)	(0.045)	(0.040)
Age	−0.000	−0.000	0.000	−0.000
	(0.000)	(0.000)	(0.001)	(0.001)
Education Level: 1	−0.118***	−0.051	0.187***	0.017
	(0.034)	(0.038)	(0.056)	(0.064)
Education Level: 2	−0.040	0.005	0.050	−0.029
	(0.025)	(0.027)	(0.041)	(0.045)
Education Level: 3	−0.100***	−0.019	0.120***	0.006
	(0.026)	(0.028)	(0.042)	(0.047)
Education Level: 4	−0.111***	0.017	0.149***	−0.064
	(0.030)	(0.030)	(0.048)	(0.050)
Education Level: 5 and above	−0.089***	−0.003	0.114***	−0.030
	(0.025)	(0.027)	(0.041)	(0.046)
Education Level: Other	−0.094***	−0.004	0.138***	−0.048
	(0.026)	(0.028)	(0.042)	(0.047)
Female	0.002	−0.013	−0.008	0.004
	(0.011)	(0.010)	(0.017)	(0.018)
Region: North West	0.046	0.028	−0.113**	0.026
	(0.031)	(0.031)	(0.049)	(0.053)
Region: Yorkshire and the Humber	0.020	0.023	−0.069	0.050
	(0.032)	(0.032)	(0.051)	(0.055)
Region: East Midlands	−0.008	0.025	−0.030	0.002
	(0.033)	(0.033)	(0.053)	(0.055)
Region: West Midlands	0.042	0.005	−0.106**	0.047
	(0.032)	(0.033)	(0.052)	(0.055)
Region: East of England	0.004	0.033	−0.057	−0.041
	(0.031)	(0.032)	(0.050)	(0.054)
Region: London	0.058*	0.056*	−0.111**	−0.018
	(0.031)	(0.031)	(0.049)	(0.053)
Region: South East	0.039	0.013	−0.094*	0.029
	(0.030)	(0.031)	(0.048)	(0.052)
Region: South West	0.031	0.015	−0.077	0.042
	(0.032)	(0.032)	(0.052)	(0.054)
Region: Wales	0.009	0.004	−0.118**	0.048

	Brier Score		Correct Dichotomized Guess	
	Brexit Exp.	Party Exp.	Brexit Exp.	Party Exp.
	(0.037)	(0.037)	(0.060)	(0.062)
Region: Scotland	0.021	0.042	−0.014	−0.026
	(0.035)	(0.034)	(0.057)	(0.058)
R^2	0.025	0.013	0.017	0.011
Adj. R^2	0.016	0.005	0.008	0.002
Num. obs.	3308	3394	3308	3394

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 1: Coefficient Estimates for a Regression Model for Brier Score and Correct Dichotomized Guess by Respondent Characteristics

Finally, we also assessed whether accuracy was related to aggregate similarity between the respondent and the evaluated profile, summarizing the difference between the respondent and the treatment profile using the Mahalonobis distance (Mahalanobis, 1936). Table 1 in the online appendix shows the result of this analysis. We find no evidence that respondents are more or less accurate in guessing the votes of profiles that are more or less similar to their own profile.[6]

The association between political attention and accuracy in guesses is not linear across the eleven categories of the 0-10 self-report, but is largely explained by the poor (high) scores of the lowest two groups in the political attention scale. As Figure 9 shows, despite the different sets of respondents in the two experiments, there is a distinctive non-monotonic pattern to the predictive performance of respondents across the difference levels of the attention measure, with those giving the "1" response on the 0-10 scale performing worst and those giving the "9" response performing best. The non-monotonicity likely reflects a non-monotonicity in how people respond to the self-assessment of political attention as a function of their real awareness of politics rather than non-monotonicity in the relationship between political attention and performance in this experiment. While it is clear that the 0s and 1s perform substantially worse than individuals expressing greater attention to politics, there is no clear trend above the two lowest levels: there is little difference between those who report a political attention of 2 and those who report a 10.

We note here the echo of Converse's conclusion that both the middle and higher strata of political sophistication can recognize the group alignment of political divides. In contrast, the lowest strata of political sophistication pays "too little attention to either the parties or the current candidates to be able to say anything about them" (Converse, 1964, p.16). Specifically, Converse claimed, the lack of linking information between the parties or policies and social groups' interests explain this lack of connection, which is consistent with our findings here.

[6]The Mahalonobis distance was measured using six attributes with available information on respondents. These attributes were: gender, region of residence, ethnicity, income, age, and education level.

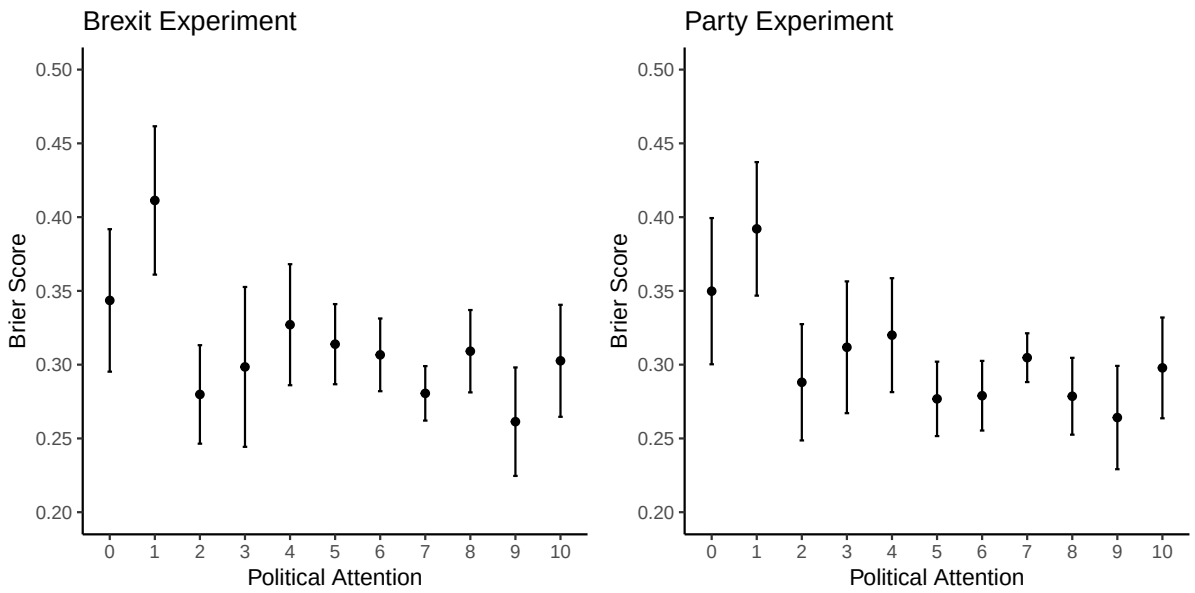

Figure 9: Brier score by respondent self-reported attention to politics.

Discussion and Conclusion

Our analysis examines both individual-level and aggregate-level accuracy, because both are important features of public understanding of how different social groups vote. It is important to know if there are systematic biases that show up in the aggregate, but also whether individuals tend to have much usable information about these questions. If individual citizens have wildly divergent beliefs about the likely voter behaviour of their fellow citizens, that is important to know even if these divergent beliefs average out to something close to reality. There is a long "wisdom of crowds" tradition of observing that while individuals may be inaccurate, they may nonetheless be accurate on average (Wallsten and Diederich, 2001; Surowiecki, 2005). This is often explained as resulting from individuals each having only a few pieces of relevant information, for example their social networks (e.g. Leiter et al., 2018), with the process of averaging cancelling out the resulting idiosyncratic errors. This pattern of individual level imprecision combined with aggregate-level accuracy is clearly evident in our data, not only because different individuals may know about the political associations of different attributes, but also because of errors in probability reporting. Individual citizens are poor at guessing how other specific citizens vote but the average guesses broadly reflect how major political cleavages relate to a variety of demographic characteristics.

The novelty of the Brexit divide means that respondents must have paid recent attention to these political cleavages, a finding further confirmed by the role of political attention in predicting accuracy, both for the older cleavage of party and the newer cleavage of Brexit. However, at the same time that we see evidence of very recent information intake in the Brexit experiment, there are some attributes

which suggest that party stereotypes are "sticky" (Green, Palmquist and Schickler, 2004; Lupu, 2013). In the party experiment education and age are strongly predictive of the actual distribution of voters, while class and economic attributes are less so. Respondents underestimate the age relationship, which makes sense in that it is newly strong; the education association with voting *used* to be that holding a degree predicted voting Tory (Heath, 2016; Ball, 2013), but that is no longer true. With respect to the "old" cleavage of party, some of respondents' errors may be because they have not updated in response to political realignments.

We find some egotistic bias, where respondents overestimate the probabilities that others have voted as they did. However, we do not find that $p(vote|X)$ accuracy is worse when respondents are asked about profiles that are more dissimilar to them, the egotistic bias applies across similar and dissimilar profiles. Thus it seems that performance in this task is less dependent on respondent's immediate social environment and more on general political knowledge. It remains to be studied if guesses on p(X) might be more dependent on immediate social environment. This contrasts with Carlson and Hill (2021) findings that respondents guesses become more accurate (less biased) for profiles that are more similar to the respondents' own profile. They explain this association as a manifestation of different-trait bias, as individuals are likely to assume that out-group members are more homogeneous than in-group members. This could be a relationship that is present for the political attitudes included in Carlson and Hill's experiment but not for demographic characteristics.

The different political contexts of the US and UK make comparisons to many of the studies we cite difficult. While our results are broadly consistent with the US study which asks the most similar questions (Carlson and Hill, 2021), we cannot rule out the possibility that US and UK citizens simply respond very differently to these kinds of survey prompts. While both countries have relatively strong two party systems, there is no shortage of political differences that could be relevant to how citizens perceive one another. We do not know whether UK studies asking questions similar to those of Ahler and Sood (2018) would find similar results to those that they find.

Regardless, our findings present an interesting puzzle in light of recent work by Ahler and Sood (2018) and Claassen et al. (2019). Those papers indicate that when asked *compositional* questions, about the demographic distributions of party supporters, respondents tend to stereotype or caricature, overstating the demographic distinctiveness of parties. The accuracy of perceptions is lower for citizens with greater interest in politics (Ahler and Sood, 2018, p969). Our paper asks a *behavioural* question about the voting of individuals with a given set of characteristics, $p(vote|X)$ rather than $p(X|vote)$, and finds no tendency of respondents to overstate the relevance of any particular attributes to guessing the vote choice of an individual. The accuracy of guesses is higher for those paying more attention to politics. Aside from the differing political context, one possible reconciliation of

these results is that respondents' inability to report percentages/proportions accurately simply manifests itself in different ways in the different experimental designs. Another possible reconciliation is that people are just inconsistent, giving answers to one kind of question that are mathematically inconsistent with the answers they would give to the other kind of question, for example, because of the representativeness heuristic that Ahler and Sood (2020) propose.

Another way of phrasing these key outstanding puzzles, which goes to the heart of the concerns raised by Ahler and Sood (2018), is to ask whether citizens *really* believe their overconfident guesses. Is the problem with reporting or with their beliefs? Ahler and Sood (2018) are unable to substantially improve the accuracy of party compositions by providing incentives to reduce expressive misreporting or by providing population base rates, which they take to suggest that citizens' beliefs are meaningfully erroneous (p969-971). Ahler and Sood (2018) further demonstrate through a series of experiments (p976-978) that the effect of correcting misperceptions about party composition is small, but non-zero, for perceptions about the extremity of opposing partisans.

For our experiment, the corresponding question is whether, for example, when someone reports 100% probability of a particular profile voting Leave, that level of certainty really guides how they would interact with and think about someone with those characteristics. Are citizens going through the world making *extremely strong* snap judgements about the political alignments of those around them, at least when given occasion to think about the politics of those people at all? Our finding that there is no one dominant pattern of such snap judgements in the aggregate does not mean that individuals are not doing this. Indeed, the implication of their numerical responses taken literally is that they are. The extent to which this is a reporting problem, as opposed to a belief problem, is less amenable to the kinds of tests used by Ahler and Sood, since the objects of evaluation in our experiments are unknown individuals rather than parties about which respondents already have other views that might be influenced by a corrective treatment.

The most compelling way forward would be to ask a much richer set of questions to individual respondents, including questions about $p(vote|X)$ and $p(X|vote)$ as well as the base rates $p(vote)$ and $p(X)$, in order to better establish which responses are consistent with one another and with reality, and which are not. While past studies have now analysed all of these quantities, they have done so in different contexts and individually rather than all in the same survey. A study of this type would be a useful next step in clarifying the complicated pattern of findings across this study and those that have been published previously.

References

Ahler, Douglas and Gaurav Sood. 2020. Typecast: A Routine Mental Shortcut Causes Party Stereotyping. (March 10, 2020). Available at SSRN: https://ssrn.com/abstract=3550117 or http://dx.doi.org/10.2139/ssrn.3550117 Rochester, NY: .

Ahler, Douglas J and Gaurav Sood. 2018. "The Parties in Our Heads: Misperceptions about Party Composition and Their Consequences." *The Journal of Politics* 80(3):964–981.

Alabrese, Eleonora, Sascha O. Becker, Thiemo Fetzer and Dennis Novy. 2019. "Who Voted for Brexit? Individual and Regional Data Combined." *European Journal of Political Economy* 56:132–150.

Atanasov, Pavel, Phillip Rescober, Eric Stone, Samuel A Swift, Emile Servan-Schreiber, Philip Tetlock, Lyle Ungar and Barbara Mellers. 2017. "Distilling the Wisdom of Crowds: Prediction Markets vs. Prediction Polls." *Management science* 63(3):691–706.

Ball, Stuart. 2013. *Portrait of a Party: The Conservative Party in Britain 1918-1945*. Oxford University Press.

Baron, Jonathan, Barbara A Mellers, Philip E Tetlock, Eric Stone and Lyle H Ungar. 2014. "Two Reasons to Make Aggregated Probability Forecasts More Extreme." *Decision Analysis* 11(2):133–145.

Bawn, Kathleen, Martin Cohen, David Karol, Seth Masket, Hans Noel and John Zaller. 2012. "A Theory of Political Parties: Groups, Policy Demands and Nominations in American Politics." *Perspectives on Politics* 10(3):571–597.

Bliuc, Ana-Maria, Craig McGarty, Katherine Reynolds and Daniela Muntele. 2007. "Opinion-Based Group Membership as a Predictor of Commitment to Political Action." *European Journal of Social Psychology* 37(1):19–32.

Boon, Martin. 2012. "Predicting Elections: A 'Wisdom of Crowds' Approach." *International Journal of Market Research* 54(4):465–483.

Brier, Glenn W. 1950. "Verification of Forecasts Expressed in Terms of Probability." *Monthly weather review* 78(1):1–3.

Broockman, David E. and Christopher Skovron. 2018. "Bias in Perceptions of Public Opinion among Political Elites." *American Political Science Review* 112(3):542–563.

Campbell, Angus, Philip E Converse, Warren E Miller and Donald E Stokes. 1960. *The American Voter*. University of Chicago Press.

Carlson, Taylor N and Seth J Hill. 2021. "Experimental Measurement of Misperception in Political Beliefs." *Journal of Experimental Political Science* In Press.

Chambers, John R, Robert S Baron and Mary L Inman. 2006. "Misperceptions in Intergroup Conflict: Disagreeing about What We Disagree About." *Psychological Science* 17(1):38–45.

Claassen, Ryan L., Paul A. Djupe, Andrew R. Lewis and Jacob R. Neiheisel. 2019. "Which Party Represents My Group? The Group Foundations of Partisan Choice and Polarization." *Political Behavior* .

Cohen, Marty, David Karol, Hans Noel and John Zaller. 2009. *The Party Decides: Presidential Nominations before and after Reform*. University of Chicago Press.

Converse, Philip E. 1964. The Nature of Belief Systems in Mass Publics. In *Ideology and Its Discontents*, ed. David Apter. New York: Glencoe Free Press.

Dassonneville, Ruth. 2016. "Volatile Voters, Short-Term Choices? An Analysis of the Vote Choice Determinants of Stable and Volatile Voters in Great Britain." *Journal of Elections, Public Opinion and Parties* 26(3):273–292.

De la Cuesta, Brandon, Naoki Egami and Kosuke Imai. 2019. Improving the External Validity of Conjoint Analysis: The Essential Role of Profile Distribution. Working paper. Available at: https://imai.fas.harvard.edu/research/files/conjoint.pdf.

Fiorina, Morris P. 2002. "Parties and Partisanship: A 40-Year Retrospective." *Political Behavior* 24(2):93–115.

Ford, Robert and Philip Cowley. 2019. "Has Brexit Really Divided Britain?" https://unherd.com/2019/11/how-much-has-brexit-divided-britain/.

Goodwin, Matthew J. and Oliver Heath. 2016. "The 2016 Referendum, Brexit and the Left Behind: An Aggregate-Level Analysis of the Result." *The Political Quarterly* 87(3):323–332.

Graefe, Andreas. 2014. "Accuracy of Vote Expectation Surveys in Forecasting Elections." *Public Opinion Quarterly* 78(S1):204–232.

Green, Donald P, Bradley Palmquist and Eric Schickler. 2004. *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*. Yale University Press.

Hainmueller, Jens, Daniel J Hopkins and Teppei Yamamoto. 2014. "Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments." *Political analysis* 22(1):1–30.

Heath, Anthony. 2016. *Understanding Political Change: The British Voter 1964-1987*. Elsevier.

Hobolt, Sara B, Thomas J Leeper and James Tilley. 2020. "Divided by the Vote: Affective Polarization in the Wake of the Brexit Referendum." *British Journal of Political Science* pp. 1–18.

Joslyn, Mark R. and Donald P. Haider-Markel. 2018. "Motivated Innumeracy: Estimating the Size of the Gun Owner Population and Its Consequences for Opposition to Gun Restrictions." *Politics & Policy* 46(6):827–850.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow (Kindle Edition)*. Penguin Books.

Kalmoe, Nathan P. 2019. "Speaking of Parties\ldots Dueling Views in a Canonical Measure of Sophistication." *Public Opinion Quarterly* 83(1):68–90.

Kinder, Donald R and Nathan P Kalmoe. 2017. *Neither Liberal nor Conservative: Ideological Innocence in the American Public*. University of Chicago Press.

Kunovich, Robert M. 2017. "Perceptions of Racial Group Size in a Minority-Majority Area." *Sociological Perspectives* 60(3):479–496.

Leiter, Debra, Andreas Murr, Ericka Rascón Ramírez and Mary Stegmaier. 2018. "Social Networks and Citizen Election Forecasting: The More Friends the Better." *International Journal of Forecasting* 34(2):235–248.

Levendusky, Matthew S and Neil Malhotra. 2016. "(Mis) Perceptions of Partisan Polarization in the American Public." *Public Opinion Quarterly* 80(S1):378–391.

Lewis-Beck, Michael S and Mary Stegmaier. 2011. "Citizen Forecasting: Can UK Voters See the Future?" *Electoral Studies* 30(2):264–268.

Lupu, Noam. 2013. "Party Brands and Partisanship: Theory with Evidence from a Survey Experiment in Argentina." *American Journal of Political Science* 57(1):49–64.

Luskin, Robert C. 1987. "Measuring Political Sophistication." *American journal of political science* pp. 856–899.

Luskin, Robert C. 1990. "Explaining Political Sophistication." *Political Behavior* 12(4):331–361.

Mahalanobis, Prasanta Chandra. 1936. On the Generalized Distance in Statistics. In *Proceedings of the National Institute of Science of India*. Vol. 12 National Institute of Science of India pp. 49–55.

McGarty, Craig, Ana-Maria Bliuc, Emma F Thomas and Renata Bongiorno. 2009. "Collective Action as the Material Expression of Opinion-Based Group Membership." *Journal of Social Issues* 65(4):839–857.

Mildenberger, Matto and Dustin Tingley. 2019. "Beliefs about Climate Beliefs: The Importance of Second-Order Opinions for Climate Politics." *British Journal of Political Science* 49(4):1279–1307.

Murr, Andreas E. 2016. "The Wisdom of Crowds: What Do Citizens Forecast for the 2015 British General Election?" *Electoral Studies* 41:283–288.

Murr, Andreas Erwin. 2011. ""Wisdom of Crowds"? A Decentralised Election Forecasting Model That Uses Citizens' Local Expectations." *Electoral Studies* 30(4):771–783.

Pronin, Emily, Carolyn Puccio and Lee Ross. 2002. "36. Understanding Misunderstanding: Social Psychological Perspectives.".

Rothschild, David and Justin Wolfers. 2011. "Forecasting Elections: Voter Intentions versus Expectations." *Available at SSRN 1884644* .

Shamir, Jacob and Michal Shamir. 1997. "Pluralistic Ignorance across Issues and over Time: Information Cues and Biases." *Public Opinion Quarterly* pp. 227–260.

Sherman, David K, Leif D Nelson and Lee D Ross. 2003. "Naïve Realism and Affirmative Action: Adversaries Are More Similar than They Think." *Basic and Applied Social Psychology* 25(4):275–289.

Sobolewska, Maria and Robert Ford. 2019. "British Culture Wars? Brexit and the Future Politics of Immigration and Ethnic Diversity." *The Political Quarterly* 90:142–154.

Surowiecki, James. 2005. "The Wisdom of Crowds: Why the Many Are Smarter than the Few." *Abacus: New Edition* 39.

Todorov, Alexander and Anesu N Mandisodza. 2004. "Public Opinion on Foreign Policy: The Multilateral Public That Perceives Itself as Unilateral." *Public Opinion Quarterly* 68(3):323–348.

Wallsten, Thomas S and Adele Diederich. 2001. "Understanding Pooled Subjective Probability Estimates." *Mathematical Social Sciences* 41(1):1–18.