Understanding People's Choice When They Have Two Votes

Ludovic Rheault, University of Toronto

André Blais, University of Montreal

John Aldrich, Duke University

Thomas Gschwend, University of Mannheim

Abstract

This paper introduces a model of vote choice in mixed electoral systems where electors cast two votes. We propose a new approach to account for contamination effects, a phenomenon that we define as a causal influence making choices more similar across the vote decisions. Since causality entails a time-ordering, we argue that contamination arises only when voters choose sequentially. By making use of new survey questions asking respondents about the timing of vote decisions, we can estimate the magnitude of these contamination effects directly. The model is tested using Bayesian multinomial probit models and survey data collected at the time of the 2013 federal election in Germany. Our findings suggest that contamination effects are present only among voters with lower levels of education, and operate mostly from the list vote to the candidate vote. We also test a number of predictions about the determinants of the two vote choices in mixed systems.

Understanding People's Choice When They Have Two Votes

Understanding why people vote the way they do is the central question in the field of electoral studies. There is a vast literature on the factors that lead citizens to vote for a particular party or candidate in a given election. The question that we address here is how to make sense of the choices that people make when they have two votes. There is surprisingly little research on this question in spite of the fact that the number of elections where people make two choices is growing. According to Carter and Farrell, 35 democracies use a mixed system for their legislative elections, and the great majority of them have two votes, almost always a party list vote and a vote for a candidate in a local constituency (Carter and Farrell 2010, 27; Massicotte and Blais 1999). Because they offer a trade-off between proportionality and local representation, mixed-member proportional systems such as the one used in Germany have gained in popularity in the past decades (see Bormann and Golder 2013; Linhart, Raabe and Statsch 2018). Instances of parties competing for multiple offices concurrently also extend beyond mixed electoral systems. Many countries feature elections in which parties are running at the same time for senate and lower house seats, or federal and subnational parliaments. Although the model that we develop in this study is applied to mixed electoral systems, its main features are exportable to a wider range of elections with multiple votes.

The presence of two votes in mixed systems has spurred interest about the magnitude and determinants of ticket-splitting, that is, people voting for a candidate associated with a party different from the one that they support on the list vote (Burden 2009; Gschwend 2007; Gschwend, Johnston, and Pattie 2003; Helmke 2009; Karp et al. 2002; Moser and Scheiner 2005; 2009; Pappi and Thurner 2002). A number of studies have also investigated the presence or absence of contamination effects (or spillover effects), that is, whether vote choice in one election affects the choice in the other election (Cox and Schoppa 2002; Ferrara, Herron, and Nishikawa 2005; Hainmueller and Kern 2008; Herron and Nishikawa 2001; Karp 2009). Yet, as far as we can tell, there has been no systematic, comparative investigation of the factors that affect each vote when people have two decisions to make, one with respect to the choice of a candidate in the local constituency and one with respect to the choice of a party list. We intend to fill that gap here by formulating a number of predictions and testing them using data collected during the 2013 German federal election. Germany uses a mixture of proportional representation (PR) and first-past-the-post (FPTP) rules since 1953. This makes the country a natural choice to test our model, since we can reasonably expect the German public to have well established strategies for casting the two votes, unlike voters in newly adopted mixed systems.

A key contribution of this study is to introduce a new framework to estimate contamination effects at the micro-level. So far, most of the debate surrounding the existence of contamination effects between electoral systems has relied upon macro evidence (see e.g. Cox and Schoppa 2002; Herron and Nishikawa 2001; Ferrara 2004; Moser and Scheiner 2004; Maeda 2008; Hainmueller and Kern 2008; Crisp, Potter, and Lee 2012). Building upon Duvergerian principles, many of those studies are looking to assess contamination by testing whether mixed electoral systems constrain the number of parties the same way plurality systems do, as opposed to pure PR systems. However, since mechanisms leading to contamination ultimately depend on the decisions of voters, the macroscopic approach leaves a large part of the phenomenon unobserved. Existing attempts to account for contamination at the micro level are reviewed in the next section, but overall, most of them use indirect estimates. Spillovers across vote decisions have also been studied in contexts where presidential elections are held concurrently with other national or subnational elections (Ferejohn and Calvert 1984; Campbell and Sumners 1990; Ames 1994; Samuels 2000), then again, most of the time using aggregate data or indirect measures of coattail effects. Our approach provides a direct measure of causal effects across vote decisions, making it possible to overcome identification issues while remaining general enough to be implemented in most situations involving simultaneous elections.

A Model of Multiple Vote Choice

We begin by introducing a model in which voters must choose multiple times between a set of parties competing in concurrent elections. Our model accounts for the possibility of "contamination" between the choices of individual voters. We define a *contamination effect* as the causal effect of a voter's decision to support party in one election on the preference of that voter for the same party in another election. We adopt a strict interpretation of causality that entails the existence of a time ordering between vote decisions: a condition for the existence of a contamination effect from an election s to an election t is that the voter's decision in s has been made before the decision in t. Consequently, our definition rules out the possibility of simultaneous contamination for individual voters. This avoids postulating the existence of simultaneous causation, a problematic concept at the philosophical level (see Granger 1969; Mellor 2002, Ch. 17). Our model involves discrete choices that are made only once in each election, which precludes the possibility of a feedback once the decision in one election has been made. For each pair of elections, either a voter makes one decision before the other, or else both decisions are made at the same time, in which case we rule out the possibility of contamination. As we illustrate later on in this study, it is possible to establish the sequence of the vote decisions for individual voters by collecting information about the timing of these choices in surveys.

Let us denote the event of a voter choosing in election s before election t as $s \to t$. Furthermore, we denote the contamination effect of a vote y_{sj} for party j in election s on the preference for party j in election t as θ_{st}. For the German mixed compensatory system with a local candidate vote C and a national party list vote L, this allows us to write the utilities of a voter for each election as:

$$u_{Lj} = \gamma_{Lj} + \boldsymbol{x}_j' \boldsymbol{\alpha}_j + \theta_{CL} \mathbf{1}(C \to L) y_{Cj} + \varepsilon_{Lj}, \tag{1}$$

$$u_{Cj} = \gamma_{Cj} + \boldsymbol{x}_j' \boldsymbol{\beta}_j + \theta_{LC} \mathbf{1}(L \to C) y_{Lj} + \varepsilon_{Cj}, \tag{2}$$

where $\boldsymbol{x}_j$ is a vector of covariates (some of which may be voter-specific, some party-specific), the γ parameters are intercepts, $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_j$ are vectors of parameters for each election, and $\mathbf{1}(.)$ is the indicator function equaling one if the expression between brackets is true and zero otherwise. For each election t, the ε_{tj} represent random disturbance terms. For the purpose of this paper, we assume that disturbances are multivariate normally distributed within each election, with mean vector zero and covariance matrix $\boldsymbol{\Sigma}_t$. The notation $\theta_{st}\mathbf{1}(s \to t)y_{sj}$ reflects the definition introduced above: a contamination effect θ_{st} is the impact of a vote decision y_{sj} on u_{tj}, given that the decision y_{sj} has been made *before* the decision in election t.[1] Since the events $(s \to t)$ and $(t \to s)$ cannot be true at the same time, our model is "recursive": the choice y_{sj} cannot depend in turn on u_{tj} if $(s \to t)$ is true. By construction, the terms $\mathbf{1}(s \to t)y_{sj}$ represent the part of the variation in y_{sj} that is strictly exogenous to u_{tj}.

The natural expectation is that the contamination effects, if they exist, are positive. In other words, choosing j in an election s increases the likelihood of choosing j in election t. We believe that this captures the intuition behind the contamination argument as commonly invoked in the previously cited literature on electoral systems. If contamination occurs, the outcome of election t becomes more similar to that of election s than it would otherwise have been. Thus, it must be the case that for some voters, the decision to vote for j in context s increases the chance of voting for j in context t.

We note that alternative theoretical predictions are possible. It could be that a decision to vote for j in election s reduces the likelihood of voting for the same party in election t, for instance if we assume that voters cast split-ticket votes to balance out the composition of government (see e.g. Lewis-Beck and Nadeau 2004). Such a hypothesis would have important implications for the study of electoral systems, since it could lead to an aggregate outcome in election t that becomes more dissimilar to that of election s, compared the outcome of t that would have been observed had that election taken place alone. An additional benefit of our approach is that we can test the relevance of such an alternative expectation by examining

[1] Notice that the contamination effect parameters θ are not specific to party. In multi-party elections, such models will likely be estimated with multinomial models in which party-specific variables have constrained coefficients across the alternatives j. As a result, there would be only one parameter θ_{st} for each pair of elections $\{s, t\}$.

the sign of the estimated θ_{st} parameters. In particular, we expect sophisticated voters to split their vote more frequently, which has consequences for the sign of these estimates and, as a result, the nature of contamination effects. We account for this possibility in our empirical section.

With regard to mixed electoral systems more specifically, our approach expands on existing attempts to measure contamination effects, which have been mostly indirect so far. For instance, Ferrara, Herron, and Nishikawa examined whether local candidate ratings have an independent impact on the list vote choice in the 1999 New Zealand elections, controlling for a host of other factors (Ferrara, Herron, and Nishikawa 2005, Ch. 5). They found that they do and inferred the existence of a contamination effect of the candidate vote on the list vote (Ferrara, Herron, and Nishikawa 2005, 73, Table 5.1). This effect is estimated to be rather weak. For his part, Karp (2009) used data on the 2002 New Zealand election and determined whether the presence of an incumbent increases the propensity to cast a list vote for the incumbent's party in the case of the two main parties (Labour and National) and whether the presence of a local candidate enhances the likelihood of supporting the party's candidate in the PR vote. The author reported a small incumbency effect for Labour vote but none for National and no candidate presence effect for the minor parties. His conclusion is that "the overall impact of candidate effects appears to be quite small" (Karp 2009, 49).

As far as we can tell, no previous empirical study has attempted to measure contamination effects *from* the list vote *toward* the candidate vote per se. To the extent that the list vote is more important than the candidate vote—since it is the former that determines how many seats a party obtains—voters should focus their attention on party lists first and foremost. We thus expect a stronger contamination from the list vote toward the candidate vote than the other way around. We also expect contamination effects to occur mostly among the less sophisticated voters. Sophisticated voters should not only be more likely to cast split-ticket votes, a pattern that would offset contamination effects, they should also rely more easily on distinct evaluative criteria for each of the two votes.

Determinants of Vote Choice in Mixed Electoral Systems

In addition to measuring contamination effects, we make predictions about the factors affecting vote choice in a mixed electoral system with a list vote and a local candidate vote. Our model comprises five basic proximate determinants of vote choice: party ratings, leader ratings, local candidate ratings, local chances of winning, and coalition ratings. Put simply, we propose that a voter is more likely to vote for a party when she thinks of herself as close to a party, when she likes the party, its leader and its local candidate, when she believes that the party's candidate has some chance of winning locally, and when she likes the coalition that the party is associated with.

We understand that the vote decision is affected by other considerations, perhaps the most obvious being issue positions and ideology. But we assume that these are more distant factors whose effect is basically indirect. We assume that voters' ideology and attitudes affect how much they like or dislike the various parties and leaders, and that these likes and dislikes in turn determine the final vote choice. This research strategy was followed by Page and Jones (1979) and Rahn et al. (1990), among others. We focus here on the more proximate factors.

The question that we address is whether these factors have a similar influence on the two votes. In terms of equations 1 and 2, we want to know how $\boldsymbol{\alpha}$ differs from $\boldsymbol{\beta}$. The first prediction concerns the party's perceived chances in the local constituency and local candidate ratings. These two considerations should affect the candidate vote choice but they should have no impact on the list vote. Logically, whether one likes the local candidate should have little influence on the decision to support a party list, all else equal. To be sure, Klingemann and Wessels (2001) noted previously that the German mixed system "[...] was specifically designed to strengthen personal ties between representatives and constituents (p. 279)." However, citing empirical evidence, the authors also argue that such a personalization of politics through local candidates probably has little effect on the list vote, and is limited to the candidate vote (see Klingemann and Wessels 2001, 279-280). In the same manner, since the number of seats a party gets from the list vote is practically unaffected by the number of votes it gets in a constituency, there is no reason to defect from a party list simply because that party is unlikely to win in the local constituency. We thus predict these two factors to affect only the candidate vote.

The second prediction has to do with party and leader ratings. We expect these two factors to have a stronger effect on the list than on the candidate vote. The list vote entails expressing support for a given party, and we should thus observe that how much one likes a party has a strong effect on the propensity to vote for that party list. The marginal effect should be weaker in the case of the candidate vote since people are explicitly asked to express support for a person. We do expect a positive association, however, between party ratings and the candidate vote since people may well prefer to be represented in their local constituency by a person associated with a party that they like and trust even if they do not particularly like that person.

The same should apply to leader ratings. Poguntke and Webb (2005) argued that leaders have become increasingly important in contemporary political parties. They are the public face of the party during election campaigns, they exercise considerable control over the extra parliamentary party and its resources, and they lead the elected members in the legislature. Voting for a party is thus also implicitly voting for its leader. Again we expect a weaker effect of leader ratings on the candidate vote because the latter entails supporting a particular person in the constituency. We nevertheless anticipate a positive association because, everything else being equal, one should prefer the local candidate to be under the direction of a “good” party leader.

The third prediction is about coalition ratings. The hypothesis is that coalition preferences affect only the list vote. There is empirical evidence that in countries where coalition governments are the norm, people’s vote choice depends not only on how they feel about the parties but also on how they feel about the coalitions that could be formed after the election (Abramson et al. 2008; Bargsted and Kedar 2009; Blais et al. 2006; Gschwend and Hooghe 2008; Meffert and Gschwend 2010). Such considerations, however, should enter the calculus decision only for the list vote since the candidate vote has no consequence on the number of seats won by the various parties (and thus on the likelihood of different coalitions).

To sum up, we expect that vote choice in mixed systems is characterized by

$$\text{Local Chances of Winning: } \beta^1 > \alpha^1 = 0; \tag{3}$$

$$\text{Local Candidate Ratings: } \beta^2 > \alpha^2 = 0; \tag{4}$$

$$\text{Party Ratings: } \alpha^3 > \beta^3 > 0; \tag{5}$$

$$\text{Leader Ratings: } \alpha^4 > \beta^4 > 0. \tag{6}$$

Again, the parameters β are marginal effects in the latent utilities for the candidate vote, whereas the parameters α are the corresponding effects in the list vote. Since all these variables are party-specific, we denote them with a single coefficient. For coalition ratings, we expect that

$$\text{Coalition Ratings: } |\alpha_j^5| > |\beta_j^5| = 0, \; \forall\, j. \tag{7}$$

Notice that coalition ratings are voter-specific variables, which implies one parameter per party. We use absolute values since the signs can vary depending on which coalition is considered, and which party is used as a reference for comparisons. Finally, we expect that contamination effects are more important from the list vote toward the candidate vote, rather than the other way around:

$$\text{Contamination: } \theta_{LC} > \theta_{CL} > 0. \tag{8}$$

As explained before, we also expect contamination effects to be largest among the less sophisticated voters.

The 2013 German Federal Election

Our empirical tests rely upon survey data on the 2013 German federal election. As mentioned in the introduction to this paper, Germany uses a mixed compensatory system where voters cast two votes on the same ballot, one for a local candidate in their constituency and one for a party list.[2] The party list election takes place under a PR system whereas the local candidate election uses a FPTP system. Our survey includes samples from two major Landers, Bavaria and Lower Saxony, encompassing both variants of the German party system: in Bavaria, the Christian Social Union (CSU) runs in lieu of its sister party, the Christian Democratic Union (CDU), although both formations are united at the federal level under a single leader.

The 2013 election gave the center-right Christian Democratic Union/Christian Social Union (hereafter CDU for short) a plurality of seats, for the third consecutive time. In terms of popular vote, the CDU received above 40% of the ballots in both election types, distancing the center-left Social Democratic Party (SPD) by some 10 percentage points.[3] Angela Merkel's party also secured a comfortable advance from the second-runner in terms of the seat distribution, obtaining slightly less than 50% of the seats compared to 30.5% for the SPD. On the other hand, the Christian Democrats' traditional coalition partner, the Free Democratic Party (FDP), did not garner enough list votes to meet the legal threshold of 5%, and as a result, was not allocated any seat. The two other minor parties having received seats (the Greens and the Left, with about 10% of the deputation each) being ideologically at odds with Merkel's party, this complicated the formation of a coalition controlling a majority of seats in the Bundestag. The two largest parties eventually formed an unlikely coalition government.

Our data come from Internet campaign surveys conducted by Harris/Decima. The Bavarian survey was conducted between September 16 and 21, 2014, with 4,762 respondents, and a post-election wave conducted between September 23 and 28, with 4,041 from the first-wave responding to the second wave questionnaire. The Lower Saxon survey was conducted between September 12 and 19 with 1001 respondents. A post-election wave was held between September 23 and 30, with 789 respondents. Overall, the contact rate was 12% and the response rate 11.5%. In both cases, a stratified, quota based sampling approach was used, the quotas being established for age, gender, and education. The vote distributions in the sample closely match the observed vote distribution in the total population of voters, especially in the case of Bavaria. As a result, our multivariate analysis does not include sample weights. Our empirical models will contain a dummy variable called Bavaria, equaling one if a respondent is from Bavaria and zero if from Lower Saxony.[4]

The dependent variables are the two vote choices as reported during the post-election waves. For the purpose of our analysis, we focus on the five main parties in each election:

[2] Notice that electoral agencies in Germany refer to the candidate and list votes as the "first" and "second" votes, respectively. We do not use these labels here.

[3] Official results are taken from the web site of the German Federal Returning Officer (www.bundeswahlleiter.de).

[4] We tested whether our main results are affected by the combination of two Landers. The results presented below were replicated after including interaction terms with the Bavaria dummy variable. These interaction effects were for the most part insignificant, suggesting that our results are not driven by one of the two regions.

CDU, SPD, Greens, FDP, and Left. The other parties garnered marginal vote shares, and we do not have measures of party and leader ratings for these smaller parties. The cross-tabulation in Table 1 shows the relationship between the two vote choices. As can be seen, most voters opt for a straight-line ticket, the main diagonal containing the largest proportions. Some proportion of Green and SPD voters were also keen on splitting their vote between these parties, which are usually expected to become coalition partners should the opportunity to form the government come about. Likewise, voters supporting the FDP list were somewhat likely to pick a CDU candidate in their local constituency, and vice-versa.

Table 1: Cross-Tabulation of Observed List and Candidate Votes

		List Vote				
		CDU	SPD	Greens	FDP	Left
Candidate Vote	CDU	87%	3%	1%	8%	1%
	SPD	4	78	14	1	4
	Greens	5	16	69	3	7
	FDP	18	3	0	79	0
	Left	3	10	0	0	87
Total (N = 2,694)		1,319	757	270	190	158

The table reports the percentage distribution of the list vote across the candidate vote choice in the survey sample. Percentages may not add up to 100 due to rounding. The bottom row reports the frequencies by list vote choice.

As mentioned above, our model includes five proximate determinants of the vote: party ratings, leader ratings, local candidate ratings, local chances, and coalition preferences. We also consider age, education, gender, and party identification as control variables. Table A1 in the Appendix reports descriptive statistics for these variables, which have all been rescaled between 0 and 1.

Party and leader ratings are based on questions asking respondents how much they like or dislike the various parties and leaders on a 0 to 10 scale (later rescaled from 0 to 1). Table A1 shows that the CDU was the most liked party while the Left was the most disliked. As for party leaders, Angela Merkel enjoyed an advantage with an average rating of 0.687 (about 7 out of 10), followed by Peer Steinbrück from the SPD. Overall, the ordering of average ratings in the sample is consistent with the support each party received at the time of the 2013 federal election.

The measurement of local candidate ratings differs from the previous two variables. We make use of responses to a survey question asking whether there is a candidate that the respondent particularly likes in the constituency, and if yes, from which party. Only around 30% of those who voted mentioned a candidate, almost all of them referring to a CDU or SPD candidate. The Local Chances variable is the score given by respondents to the perceived chances of each party winning in their constituency, on a 0 to 10 scale (again rescaled to run from 0 to 1). Typically, CDU and SPD candidates were perceived to have the best chances of winning in the constituency.

Coalition preferences were tapped through questions asking people how much they like or dislike (on a 0 to 10 scale) different government coalitions that could be formed after the election. For the purpose of this study, we focus on the two most plausible coalitions, those involving CDU with the FDP on the one hand and SPD with the Greens on the other hand. The actual turn of events leading to a Grand Coalition between the CDU and SPD was an unlikely outcome, which is why it would make little sense to explain vote choice using preferences over such an unusual coalition. To create our coalition ratings variable, we simply subtract the score given to the SPD–Greens coalition from the score given to the CDU–FDP coalition. The resulting variable is rescaled into the $[0, 1]$ range. The mean score is 0.523, suggesting that the CDU and FDP formed a slightly more popular coalition than did the SPD and Greens (0.5 indicates indifference between the coalitions).

Next, to account for contamination effects, we create time-ordered vote choice variables using information on the reported timing of the vote decisions. All post-election wave respondents were asked squarely whether they made one of the two vote decisions before the other. The wording of that survey question is "Which vote did you decide upon first?" and the response categories included "The candidate vote", "The list vote", "I decided both at the same time" and "Don't know" (translated from the German original version). Table 2 shows the distribution of responses to this question. Overall, about 36% of the respondents made their decision regarding the candidate vote before they chose which party list to support, whereas approximately 18% did the opposite. We measure the expressions $\mathbf{1}(L \to C)y_{Lj}$ from equation (2) using binary variables measuring the list vote choice conditional on having made a decision regarding party lists before the decision regarding local candidates. The expressions $\mathbf{1}(C \to L)y_{Cj}$ are constructed in a similar fashion, the other way around.

Table 2: Timing of the Vote Decisions

Timing	Frequency	Percentage
Candidate Vote First	962	36%
Both at the Same Time	1,243	46%
List Vote First	489	18%
Total	2,694	

Distribution of responses to the survey question tapping the timing of the two vote decisions.

In the last part of our empirical analysis, we reassess our hypotheses after accounting for our respondents' level of sophistication. We measure sophistication using educational attainment. Since survey quotas were established in part on this variable, its distribution closely matches that of the actual German population, making it the most reliable indicator of sophistication at our disposal. This variable contains three categories that account for the specificity of the German education system, in which different types of high schools coexist. The first category contains respondents with lower secondary or incomplete secondary schooling (about 20% of our final sample), the second category contains those with standard secondary schooling or technical degrees (37% of the sample), while the third category contains respondents with high secondary degrees or college education (42% of our sample). We create interaction terms by multiplying this educational attainment measure with our party-specific variables. For reasons exposed previously, we are especially interested in testing for the existence of contamination effects conditional on the level of education.

The choice of empirical estimators for our vote models requires some consideration. Previous research on binary probit models with simultaneous equations suggests that when exogeneity can be achieved, univariate estimators are to be preferred (Monfardini and Radice 2007). We adopt a similar strategy by sampling parameters from separate multinomial probit estimators. Even though our models contain recursive elements (the contamination effects), the time-ordering restrictions that we impose in Eqs. 1–2 ensure that the components $\mathbf{1}(s \to t)y_{sj}$ on the right-hand side of the latent utility functions do not introduce a correlation with the error terms ε_{tj}. This follows from the fact that $(s \to t)$ and $(t \to s)$ cannot be true at the same time, eliminating the possibility of feedback across equations. We also tested multinomial logit models estimated jointly by stacking the data for both elections. We report these results in Tables A3–A5 in an appendix, and show that they are consistent with the main findings of this paper. The multinomial probit models do not require making the independent of irrelevant alternatives (IIA) assumption, which is why we focus on these results in the main paper.

We rely upon Bayesian implementations of the multinomial probit model using the marginal data-augmentation algorithm proposed by Imai and van Dyk (2005*a*;*b*). This approach makes use of a Gibbs sampler to successively draw the latent utilities and the parameters, allowing to find credible posterior distributions of the form $p(\alpha|D)$, where we denote our data as D for short. For each model, we use non-informative priors for the parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and the θ's, namely a multivariate normal distribution with mean zero and an identity covariance matrix $\boldsymbol{I}$. The priors for the covariance matrices $\boldsymbol{\Sigma}_L$, $\boldsymbol{\Sigma}_C$ are drawn from the inverse-Wishart distribution using the default values proposed by Imai and van Dyk (2005*b*), the degrees of freedom being set to five—the number of alternatives—and the scale parameter to one. Notice that covariances matrices are of size $(J-1) \times (J-1)$ (where J is the number of parties) since one alternative serves as the base category. Moreover, the first diagonal elements of the covariance matrices are constrained to one to provide a normalizing scale.

Empirical Findings

We start by computing three samples of 1,000,000 Markov chain Monte Carlo (MCMC) draws, using different sets of starting values each time, for each of the two vote models. In both cases, a first MCMC sample is computed after setting the starting values of our parameters to zero. The next two samples use overdispersed starting values following the sequences $(-1, 1, -1, ...)$ and $(1, -1, 1, ...)$, respectively. This leaves us with a total of six million MCMC draws. To proceed with our empirical analysis, we make use of the last 500,000 draws from each of the three MCMC samples, once again for both the list vote and candidate vote equations. The appendix provides detailed information on diagnostic statistics, which support the conclusion that the chains have successfully converged to a stationary posterior distribution.

Table 3 reports the mean and the 95% credible intervals of the posterior distributions for our parameters of interest. The left section of the table reports statistics for the list vote equation, whereas the right section reports candidate vote estimates. Overall, the models perform well, correctly predicting 84.2% and 84.6% of actual vote choices, respectively for the list and candidate votes. This represents of a 69% proportional reduction in errors for the list vote model, and of 67% for the candidate vote model. Notice that the CDU serves as the reference category for voter-specific variables. Moreover, since all explanatory factors are on the same $[0, 1]$ scale, the means and quantiles of the posterior distributions can be compared in size.

Our first prediction is that local chances of winning and local candidate ratings affect the candidate vote but not the list vote, as stated in equations (3) and (4). The findings are consistent with this prediction. As shown in the left section of Table 3, the mean of the posterior distribution for the local chances of winning parameter α^1 falls close to zero in the list vote equation. In contrast, the posterior distribution of the corresponding parameter β^1 in the candidate vote equation appears unambiguously positive. A similar conclusion holds for local candidate ratings. Our results suggest that an important reason why the major parties do better than their junior partners in the candidate election is that some supporters of small parties are willing to defect at the constituency level, because they do not want to waste their vote on a candidate who is unlikely to win. To facilitate visualization of this result, Figure 1a reports the posterior distributions for both parameters, stacked in the same graph.

However, Bayesian analysis provides more direct ways to test hypotheses about differences between these two models. We rely on two different approaches in what follows. First, we evaluate a prediction that $\beta^k > \alpha^k$ by testing whether the posterior distribution $p(\beta^k|D)$ is larger than the most credible value of α^k, namely the median of its posterior distribution, which we denote by $\widetilde{\alpha}^k$. Since both models include the same variables and are estimated using identical priors and specifications, the coefficients are on the same scale, allowing comparisons

Table 3: Bayesian Multinomial Probit Models of List and Candidate Votes

		List Vote		Candidate Vote	
Party	Variable	Mean	Credible Interval	Mean	Credible Interval
	Local Chances (α^1, β^1)	−0.021	[−0.270, 0.230]	0.549	[0.304, 0.801]
	Local Ratings (α^2, β^2)	0.139	[−0.009, 0.287]	0.461	[0.308, 0.614]
	Party Ratings (α^3, β^3)	3.053	[2.593, 3.530]	2.477	[2.010, 2.933]
	Leader Ratings (α^4, β^4)	1.274	[0.976, 1.588]	0.842	[0.561, 1.131]
	Candidate $\rightarrow$ List (θ_{CL})	0.209	[0.077, 0.348]		
	List $\rightarrow$ Candidate (θ_{LC})			0.322	[0.161, 0.493]
	Party ID	0.813	[0.663, 0.977]	0.550	[0.390, 0.713]
SPD	Coalition Ratings (α^5_2, β^5_2)	−1.297	[−1.940, −0.644]	−1.791	[−2.488, −1.157]
	Age	0.080	[−0.411, 0.569]	0.279	[−0.167, 0.728]
	Education	−0.194	[−0.478, 0.088]	0.207	[−0.053, 0.467]
	Gender	0.093	[−0.103, 0.293]	0.016	[−0.163, 0.196]
	Bavaria	−0.018	[−0.283, 0.245]	0.315	[0.069, 0.557]
	Intercept	0.697	[0.215, 1.176]	0.318	[−0.155, 0.806]
Greens	Coalition Ratings (α^5_3, β^5_3)	−1.061	[−1.828, −0.331]	−0.483	[−1.231, 0.221]
	Age	−0.365	[−0.965, 0.226]	0.090	[−0.486, 0.651]
	Education	0.213	[−0.146, 0.584]	0.258	[−0.074, 0.608]
	Gender	0.079	[−0.162, 0.324]	0.115	[−0.107, 0.345]
	Bavaria	0.173	[−0.146, 0.498]	0.172	[−0.117, 0.469]
	Intercept	0.110	[−0.485, 0.687]	−0.503	[−1.118, 0.083]
FDP	Coalition Ratings (α^5_4, β^5_4)	1.393	[0.499, 2.300]	0.158	[−0.854, 1.142]
	Age	−0.393	[−1.042, 0.250]	−1.296	[−2.062, −0.562]
	Education	0.120	[−0.282, 0.537]	−0.048	[−0.481, 0.405]
	Gender	−0.143	[−0.432, 0.133]	−0.106	[−0.431, 0.202]
	Bavaria	0.536	[0.105, 0.984]	0.639	[0.154, 1.174]
	Intercept	−1.267	[−2.093, −0.490]	−0.633	[−1.569, 0.222]
Left	Coalition Ratings (α^5_5, β^5_5)	−0.980	[−1.867, −0.104]	−0.791	[−1.696, 0.105]
	Age	−0.225	[−0.997, 0.539]	0.146	[−0.648, 0.935]
	Education	−0.479	[−0.946, −0.016]	−0.208	[−0.673, 0.255]
	Gender	0.058	[−0.270, 0.387]	0.139	[−0.192, 0.468]
	Bavaria	−0.071	[−0.455, 0.310]	0.163	[−0.209, 0.541]
	Intercept	0.567	[−0.136, 1.253]	−0.338	[−1.071, 0.358]
	% Correctly Predicted	84.2%		84.6%	
	Observations	2,694		2,694	
	Monte Carlo Draws	1,500,000		1,500,000	

Summary statistics of the posterior predictive distributions of parameters from the list and candidate vote equations, estimated with Bayesian multinomial probit models. The 95% credible intervals are reported between brackets.

based on point estimates. Generally speaking, we posit hypotheses of the form $H_1 : \beta^k > \alpha^k$ against the null $H_0 : \beta^k \leq \alpha^k$, and rely upon non-informative priors $P(H_0) = 0.5$ and $P(H_1) = 0.5$. We then estimate the posterior probability $P(H_1|D)$ numerically using the MCMC draws, by counting the proportion of draws larger than a reference value. In other words, we estimate the probability $P(\beta^k > \widetilde{\alpha}^k|D) = \int_{\widetilde{\alpha}^k}^{\infty} p(\beta^k|D)d\beta^k$ numerically. We also compute Bayes factors (B_{10}) as

$$B_{10} = \frac{P(H_1|D)}{P(H_0|D)} \bigg/ \frac{P(H_1)}{P(H_0)}, \tag{9}$$

which represent the odds of observing the data given that our hypothesis is true, relative to the null hypothesis. Following usual conventions, this value can be assessed using the scale proposed by Kass and Raftery (1995) for twice the log Bayes factors.[5]

Second, since our posterior distributions are approximately normal, a difference $\beta^k - \alpha^k$ is also normally distributed, and we can invoke the exchangeability assumption to compute the posterior distribution of this difference numerically, again using the MCMC draws of both parameters. This method accounts for the variance of both parameters. However, we need to make the strong assumption that the correlation between the coefficients is zero, which was not necessary using the first approach. Despite the caveat with the second method, replicating the tests with two different strategies helps to assess the robustness of our findings. Using the posterior distribution of differences between parameters, the probability $P(H_1|D)$ becomes:

$$P(\beta^k - \alpha^k > 0|D) = \int_0^{\infty} p(\beta^k - \alpha^k|D)d(\beta^k - \alpha^k), \tag{10}$$

which we estimate using the proportion of the difference between draws greater than zero.

Using these methods for testing our hypotheses, we find clear evidence supporting our first prediction. As can be seen from Table 4, the probability that local chances of winning and local candidate ratings have a larger influence on the candidate vote is close to 1, using either of the two approaches described above. The log Bayes factors are also above 10 in both cases, which suggest a very strong support for the two hypotheses.[6]

The second set of predictions, according to which party and leader ratings should have a stronger impact on the list vote than on the constituency vote, is also supported by the data. In both cases, the means of the posterior distributions are clearly lower in the candidate vote equation than in the list vote's. Starting with party ratings, the mean of $p(\alpha^3|D)$ is greater than

[5]The evidence in favor of H_1 is considered very strong if $2\ln(B_{10})$ is greater than 10, strong if between 6 and 10, positive if between 2 and 6, and barely worth mentioning if between 0 and 2 (Kass and Raftery 1995, 777). The null is supported if the value is negative.

[6]Note that Bayes factor cannot be computed in one case for which 100% of the draws are greater in size than the median in the other vote model.

Table 4: Hypothesis Testing

Hypothesis	H_1	H_0	Method 1		Method 2	
			$P(H_1\|D)$	$2\ln(B_{10})$	$P(H_1\|D)$	$2\ln(B_{10})$
Local Chances	$\beta^1 > \alpha^1$	$\beta^1 \leq \alpha^1$	1.000	N/A	0.999	13.769
Local Ratings	$\beta^2 > \alpha^2$	$\beta^2 \leq \alpha^2$	0.999	13.280	0.997	11.864
Party Ratings	$\alpha^3 > \beta^3$	$\alpha^3 \leq \beta^3$	0.992	9.652	0.959	6.312
Leader Ratings	$\alpha^4 > \beta^4$	$\alpha^4 \leq \beta^4$	0.998	12.537	0.979	7.719
Contamination	$\theta_{LC} > \theta_{CL}$	$\theta_{LC} \leq \theta_{CL}$	0.915	4.763	0.850	3.472
Coalition: SPD	$\lvert\alpha_2^5\rvert > \lvert\beta_2^5\rvert$	$\lvert\alpha_2^5\rvert \leq \lvert\beta_2^5\rvert$	0.070	−5.181	0.143	−3.587
Coalition: Greens	$\lvert\alpha_3^5\rvert > \lvert\beta_3^5\rvert$	$\lvert\alpha_3^5\rvert \leq \lvert\beta_3^5\rvert$	0.941	5.521	0.863	3.673
Coalition: FDP	$\lvert\alpha_4^5\rvert > \lvert\beta_4^5\rvert$	$\lvert\alpha_4^5\rvert \leq \lvert\beta_4^5\rvert$	0.996	11.158	0.966	6.706
Coalition: Left	$\lvert\alpha_5^5\rvert > \lvert\beta_5^5\rvert$	$\lvert\alpha_5^5\rvert \leq \lvert\beta_5^5\rvert$	0.665	1.370	0.617	0.952

Bayesian hypothesis tests based on the models reported in Table 3. Probabilities are computed numerically using the proportion of MCMC draws larger than a reference value. Method 1 compares the posterior distribution of the parameter of a vote model against the most credible value of the parameter in the other vote model (the median). Method 2 uses the difference between MCMC draws across vote models. B_{10} denotes Bayes factors. In all cases, we use non-informative prior probabilities $P(H_0) = 0.5$ and $P(H_1) = 0.5$.

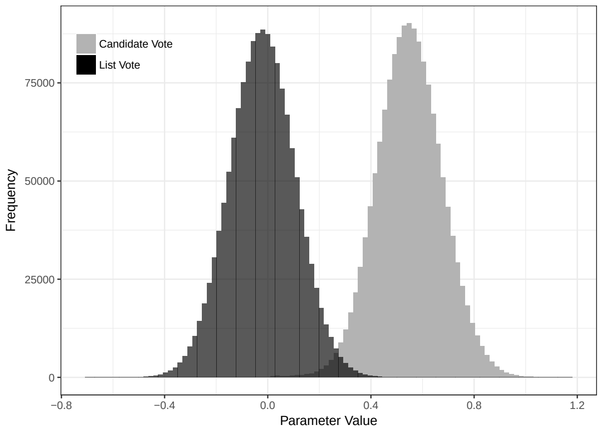
the 97.5 percentile of the corresponding parameter's density in the candidate vote equation, as shown in the third row of Table 3. The log Bayes factors fall between 5 and 10 using either of our two methods, which suggests strong evidence in favor of our hypothesis. Again, we report the densities in Figure 1c. We note that party ratings do influence the candidate vote, though to a lesser extent, which is consistent with our assumption that some people simply vote for their preferred party's candidate in the local constituency. The conclusions are very similar for leader ratings. The probability estimates for $P(\alpha^4 > \beta^4|D)$ are close to one, substantiating our prediction that leader evaluations have a stronger effect on the list vote.

On the other hand, the third prediction, that coalition preferences disproportionately affect the list vote, is only partially supported. Table 3 shows that quantiles of the posterior densities in the list vote equation are larger (in absolute terms) than those of the local vote equation, except for the SPD. Everything else being equal, the more one prefers the CDU–FDP coalition over the SPD–Green coalition, the less likely one is to vote for the SPD, Green and Left party lists, compared to the CDU. On the other hand, the FDP appears to benefit from those coalition preferences, respondents more supportive of the CDU–FDP alliance being even more likely to choose the FDP's list than the CDU's, everything else being equal. However, our results indicate that coalition preferences also affect the SPD candidate vote, which runs contrary to our initial expectation. Using Bayes factors (rows 6 to 9 of Table 4), we find supporting evidence of our hypothesis only when considering the Greens and the FDP, and strong evidence only for the FDP.

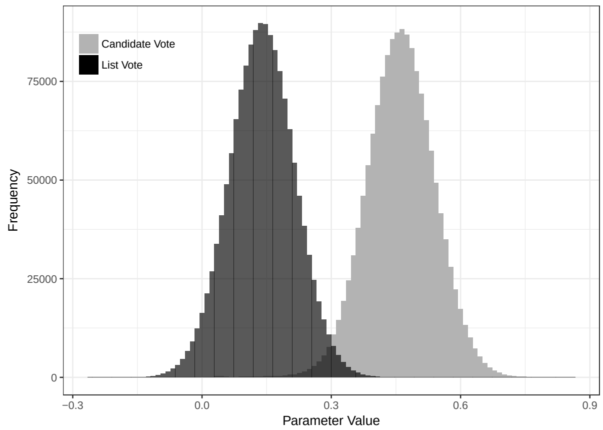
Finally, our results provide a concrete assessment of contamination effects across votes. As explained earlier, the predetermined vote choice variables enter as exogenous regressors in

Figure 1: Posterior Densities of List and Candidate Vote Parameters

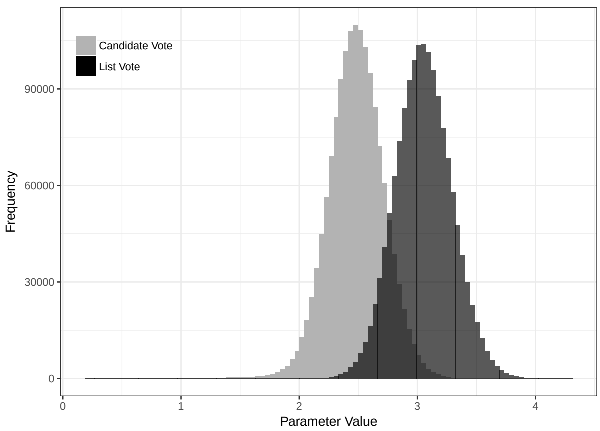
(a) Local Chances of Winning



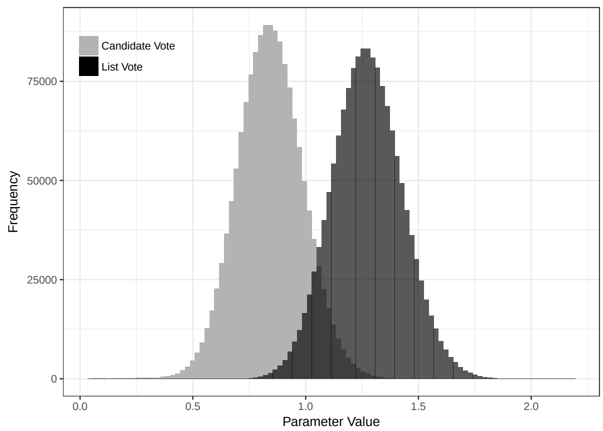
(b) Local Candidate Ratings



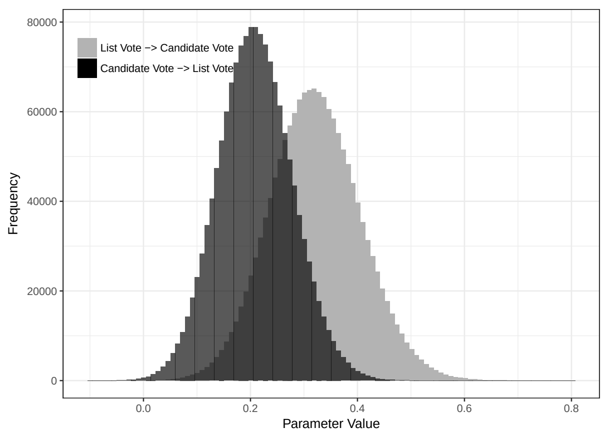
(c) Party Ratings



(d) Leader Ratings



(e) Contamination Effects



each equation, and they represent estimates of the causal effect of a previously reached vote decision on the other vote. We do find evidence of contamination effects in both directions and report the posterior densities in Figure 1e. As can be seen by comparing the fifth and sixth rows of Table 3, these effects appear larger from the list vote toward the candidate vote than the other way around. Interestingly, even though more respondents declared making up their mind about the local vote first (see Table 2), this pattern does not induce a larger contamination effect going in that direction. In fact, the opposite effect prevails: the smaller number of voters who chose in the list vote first were more likely to support the same party in their local constituency. Although consistent with our expectation, the evidence is not strong, as can be seen in the fifth row of Table 4. The posterior probability $P(\theta_{LC} > \theta_{CL}|D)$ approximates to 0.92 using our first approach, and 0.85 using the second approach, meaning that we are only 85% percent confident that θ_{LC} is larger than θ_{CL} using the more conservative test.

To illustrate the size of contamination effects, we compute changes in the predicted probability of choosing a local candidate (or party list), without and with an anterior decision to support the list (or local candidate) of that party. Table 5 reports these values. In each case, the posterior predicted probabilities are computed after setting all other variables at their sample means. As can be seen in the upper portion of Table 5, a voter having made a prior decision to support the CDU list is more likely to vote for a CDU local candidate. The marginal effect is reported in the upper-left cell of Table 5, and corresponds to approximately +11 percentage points. Obviously, this implies that the same voter is less likely to support candidates from the other parties. The size of the contamination effect is similar for a voter having decided to support the SPD list, about +12 percentage points. On the other hand, the spillovers are weaker for smaller parties. The bottom part of the table reports the change in predicted probabilities associated with contamination effects in the opposite direction. As anticipated from the previous discussion, those effects are smaller in magnitude.

The results differ when considering the level of sophistication of voters. To show this, we reassess our last prediction after including interaction terms between the level of education and each of the party-specific indicators. The rest of the specification is the same as before. Table 6 reports the results. The lower panel of the table is produced by adding the MCMC draws for the coefficients of each variable and those of their respective interactions with education. Notice that, for simplicity, we only report the summary of posterior distributions for variables with interactions. As can be seen by comparing the two panels of Table 6, contamination effects are essentially driven by voters with lower levels of education. Figure 2 compares the posterior densities in both vote models, for lowly and highly educated voters, which helps to summarize our findings. When considering respondents with a lower level of education, our estimates now clearly support our initial expectation that contamination

Table 5: Estimated Size of Contamination Effects

	List → Candidate				
	Change in Candidate Vote Probability				
List Vote	CDU	SPD	Greens	FDP	Left
CDU	0.112				
SPD		0.117			
Greens			0.043		
FDP				0.011	
Left					0.009

	Candidate → List				
	Change in List Vote Probability				
Candidate Vote	CDU	SPD	Greens	FDP	Left
CDU	0.077				
SPD		0.079			
Greens			0.030		
FDP				0.021	
Left					0.009

Marginal effects representing the change in the out-of-sample posterior predictive probability of choosing a party given that a voter previously made a decision to vote for the party indicated in the row header in the other election type. Probabilities are computed after setting all other explanatory variables of the models at their mean values.

effects are more important from the list vote toward the candidate vote than the other way around. The log Bayes factor is 13.22, suggesting strong evidence in favor of our initial hypothesis, as can be observed in Table 7. The same is not true for voters with a high level of education (see Figure 2b, showing that posterior densities are very close to each other). We also note that, apart from contamination effects, most of our other predictions are substantiated even more clearly when focusing on voters with a high level of education (see Table 7). These results reinforce the idea that sophisticated voters are more likely to consider separate criteria for each of the two vote decisions. In contrast, since the contamination argument implies a very simple decision-making heuristic, the fact that educated voters do not exhibit this behavior makes sense, substantively speaking.

Overall, our results on contamination suggest that such effects are rather modest in scope. They affect mostly the decision process of less sophisticated voters, and contamination from the list vote toward the candidate vote appears more sizable. This last finding lends credence to the view that the list vote, which determines how many seats the party gets in the legislature, is the most important of the two. Our results are also consistent with the literature at the party level, which shows that contamination effects are clearly from the PR component to the FPTP component. That is, in compensatory systems where the PR dimension dominates (as in the German case), small parties tend to nominate more candidates in the local constituencies than they would “normally” do in a FPTP election, because they have already decided to present them in the PR election (Ferrara, Herron, and Nishikawa 2005, 63). Our own conclusion is

Table 6: Bayesian Multinomial Probit Models with Education Interactions

	List Vote		Candidate Vote	
Variable	Mean	[Credible Interval]	Mean	[Credible Interval]
		LOW EDUCATION (Education = 0)		
Local Chances	0.452	[−0.062, 0.975]	0.504	[0.002, 1.022]
Local Ratings	0.244	[−0.040, 0.530]	0.300	[0.025, 0.583]
Party Ratings	2.253	[1.612, 2.921]	1.981	[1.354, 2.612]
Leader Ratings	1.516	[0.967, 2.082]	1.202	[0.679, 1.743]
Candidate → List	0.364	[0.102, 0.632]		
List → Candidate			0.987	[0.584, 1.409]
Party ID	0.516	[0.252, 0.789]	0.515	[0.247, 0.791]
		HIGH EDUCATION		
Local Chances	−0.212	[−0.541, 0.116]	0.566	[0.269, 0.869]
Local Ratings	0.083	[−0.132, 0.296]	0.511	[0.303, 0.724]
Party Ratings	3.766	[3.106, 4.447]	2.840	[2.226, 3.435]
Leader Ratings	1.141	[0.745, 1.555]	0.636	[0.285, 0.997]
Candidate → List	0.139	[−0.049, 0.334]		
List → Candidate			0.099	[−0.094, 0.302]
Party ID	0.935	[0.737, 1.152]	0.560	[0.369, 0.756]
% Correctly Predicted	84.1%		84.6%	
Observations	2,694		2,694	
Monte Carlo Draws	1,500,000		1,500,000	

Summary statistics of the posterior predictive distributions of parameters from the list and candidate vote equations, estimated with Bayesian multinomial probit models after including interaction variables between Education and each of the party-specific covariates. Only estimates for party-specific covariates are reported for simplicity, but the models include the same controls as in Table 3. The 95% credible intervals are reported between brackets. The Gelman-Rubin multivariate PSRF statistic approximates to 1.01 for both models.

that contamination effects, when they arise, are more likely to run in the same direction for voters too, from the more "important" component (PR) to the less important (FPTP).

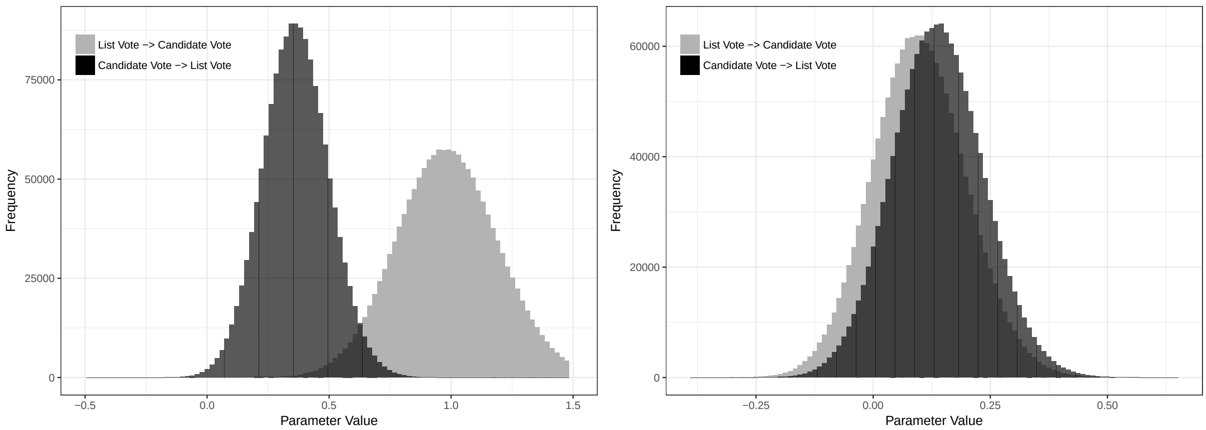
Conclusion

In many elections voters are faced with two decisions when they are making up their mind how to vote: which party list to support and which candidate to support in the constituency. As far as we can tell, no previous study has examined the determinants of the two votes using an integrated framework.

We formulated hypotheses regarding five proximate determinants of voting and their relative influence on the two vote decisions in mixed systems. As predicted, local chances of winning and local candidate ratings affect mostly the candidate vote, while party and leader ratings affect more strongly the list vote. We have also found coalition preferences to affect mostly (although not uniquely) the list vote. We believe that the methodology proposed here

Figure 2: Posterior Densities of Contamination Effects, by Education Level

(a) Low Education (b) High Education



is the most logical way to proceed. There are two votes, and scholars must provide an explanation for each. Because the factors driving voter behavior are expected to be similar for both votes, we argue in favor of a general model comprising the same variables in each vote equation. The goal is to determine whether some of the factors have a greater influence on the candidate vote than on the list vote.

We have found some differences, and these differences make sense. The most important is that perceptions of local chances affect the candidate vote but not the list vote. Moreover, our analysis helps to understand why ticket splitting is not more widespread. The fact is, for both votes, the most crucial proximate factor is how one feels about the parties. This is obvious with respect to the list vote but less so for the candidate vote. The two votes are first and foremost an expression of party preferences, and this is why most people support the same party with their two votes.

Moreover, our methodology allows us to make a contribution to the micro-foundations of contamination effects. We developed a general framework to estimate potential contamination effects between votes, and relied upon a direct measure of these effects by making use of information on the timing of vote decisions. Our findings suggest that contamination effects are limited to voters with lower levels of education. Previous studies have examined whether the list vote is influenced by the local candidate vote and they have found weak contamination effects. We observe a weak contamination effect in that direction using our survey data at the time of the 2013 German federal election. In particular, our data suggest that more voters make up their minds about local candidates first, before choosing a party list. Nonetheless, spillovers can go in the opposite direction, from the list vote to the candidate vote, and we have found stronger evidence of this type of effect. Overall, we believe that these findings make sense. We should expect people to pay greater attention to the most “important” vote,

Table 7: Hypothesis Testing by Level of Education

			Method 1		Method 2	
Hypothesis	H_1	H_0	$P(H_1\|D)$	$2\ln(B_{10})$	$P(H_1\|D)$	$2\ln(B_{10})$
			Low Education (Education = 0)			
Local Chances	$\beta^1 > \alpha^1$	$\beta^1 \leq \alpha^1$	0.576	0.613	0.554	0.437
Local Ratings	$\beta^2 > \alpha^2$	$\beta^2 \leq \alpha^2$	0.653	1.265	0.609	0.886
Party Ratings	$\alpha^3 > \beta^3$	$\alpha^3 \leq \beta^3$	0.793	2.685	0.720	1.888
Leader Ratings	$\alpha^4 > \beta^4$	$\alpha^4 \leq \beta^4$	0.868	3.772	0.788	2.627
Contamination	$\theta_{LC} > \theta_{CL}$	$\theta_{LC} \leq \theta_{CL}$	0.999	13.218	0.994	10.192
			High Education (Education = 1)			
Local Chances	$\beta^1 > \alpha^1$	$\beta^1 \leq \alpha^1$	1.000	N/A	1.000	15.941
Local Ratings	$\beta^2 > \alpha^2$	$\beta^2 \leq \alpha^2$	1.000	16.282	0.997	11.863
Party Ratings	$\alpha^3 > \beta^3$	$\alpha^3 \leq \beta^3$	0.996	10.841	0.979	7.648
Leader Ratings	$\alpha^4 > \beta^4$	$\alpha^4 \leq \beta^4$	0.994	10.328	0.967	6.771
Contamination	$\theta_{LC} > \theta_{CL}$	$\theta_{LC} \leq \theta_{CL}$	0.346	−1.274	0.388	−0.914

Bayesian hypothesis tests based on the models reported in Table 6. Integrals are computed numerically using the MCMC draws. Method 1 compares the posterior distribution of the parameter of a vote model against the most credible value of the parameter in the other vote model (the median). Method 2 uses the difference between MCMC draws across vote models, restricting the correlation of coefficients to zero. B_{10} denotes Bayes factors. In all cases, we use non-informative prior probabilities $P(H_0) = 0.5$ and $P(H_1) = 0.5$.

and the influence should flow mainly from the most to the least important decision.

References

Abramson, Paul R., John H. Aldrich, André Blais, Daniel Lee, and Renan Levine. 2008. "Coalition Considerations and the Vote." In *The Elections in Israel—2006*, ed. Asher Arian, and Shamir Michal. New Brunswick: Transaction Books.

Ames, Barry. 1994. "The Reverse Coattails Effect: Local Party Organization in the 1989 Brazilian Presidential Election." *American Political Science Review* 88(1): 95–111.

Bargsted, Matias A., and Orit Kedar. 2009. "Coalition-Targeted Duvergerian Voting: How Expectations Affect Voter Choice under Proportional Representation." *American Journal of Political Science* 53(2): 307–323.

Blais, André, John H. Aldrich, Indridi H. Indridason, and Renan Levine. 2006. "Do Voters Vote For Government Coalitions? Testing Downs' Pessimistic Conclusion." *Party Politics* 12(6): 691–705.

Bormann, Nils-Christian and Matt Golder. 2013. "Democratic Electoral Systems around the World, 1946–2011." *Electoral Studies* 32(2): 360–369.

Burden, Barry C. 2009. "Candidate-Driven Ticket Splitting in the 2000 Japanese Elections." *Electoral Studies* 28(1): 33–40.

Campbell, James E., and Joe A. Sumners. 1990. "Presidential Coattails in Senate Elections." *American Political Science Review* 84(2): 513–524.

Carter, Elisabeth, and David Farrell. 2010. "Electoral Systems and Election Management." In *Contemporary Democracies 3: Elections and Voting in the 21st Century*, ed. Lawrence LeDuc, Richard G. Niemi, and Norris Pippa. London: Sage.

Cox, Karen E., and Leonard J. Schoppa. 2002. "Interaction Effects in Mixed-Member Electoral Systems: Theory and Evidence From Germany, Japan, and Italy." *Comparative Political Studies* 35(9): 1027–1053.

Crisp, Brian F., Joshua D. Potter, and John J. W. Lee. 2012. "Entry and Coordination in Mixed-Member Systems: A Controlled Comparison Testing the Contamination Hypothesis." *The Journal of Politics* 74(4): 571–583.

Ferejohn, John A., and Randall L. Calvert. 1984. "Presidential Coattails in Historical Perspective." *American Journal of Political Science* 28(1): 127–146.

Ferrara, Federico. 2004. "Electoral Coordination and the Strategic Desertion of Strong Parties in Compensatory Mixed Systems with Negative Vote Transfers." *Electoral Studies* 23(3): 391–413.

Ferrara, Federico, Erik S. Herron, and Misa Nishikawa. 2005. *Mixed Electoral Systems: Contamination and Its Consequences.* New York: Palgrave MacMillan.

Gelman, Andrew, and Donald B. Rubin. 1992. "Inference from Iterative Simulation using Multiple Sequences." *Statistical Science* 7(4): 457–511.

Granger, Clive W. J. 1969. "Investigating Causal Relations by Econometric Models and Cross-spectral Methods." *Econometrica* 37(3): 424–438.

Gschwend, Thomas. 2007. "Ticket-Splitting and Strategic Voting under Mixed Electoral Rules: Evidence from Germany." *European Journal of Political Research* 46(1): 1–23.

Gschwend, Thomas, and Marc Hooghe. 2008. "Should I Stay or Should I Go? An Experimental Study on Voter Responses to Pre-Electoral Coalitions." *European Journal of Political Research* 47(5): 556–577.

Gschwend, Thomas, Ron Johnston, and Charles Pattie. 2003. "Split-Ticket Patterns in Mixed-Member Proportional Election Systems: Estimates and Analyses of Their Spatial Variation at the German Federal Election, 1998." *British Journal of Political Science* 33(1): 109–127.

Hainmueller, Jens, and Holger Lutz Kern. 2008. "Incumbency as a Source of Spillover Effects in Mixed Electoral Systems: Evidence from a Regression-Discontinuity Design." *Electoral Studies* 27(2): 213–227.

Heidelberger, Philip, and Peter D. Welch. 1983. "Simulation Run Length Control in the Presence of an Initial Transient." *Operations Research* 31(6): 1109–1144.

Helmke, Gretchen. 2009. "Ticket Splitting as Electoral Insurance: The Mexico 2000 Elections." *Electoral Studies* 28(1): 70–78.

Herron, Erik S., and Misa Nishikawa. 2001. "Contamination Effects and the Number of Parties in Mixed-Superposition Electoral Systems." *Electoral Studies* 20(1): 63–86.

Imai, Kosuke, and David A. van Dyk. 2005*a*. "A Bayesian Analysis of the Multinomial Probit Model Using Marginal Data Augmentation." *Journal of Econometrics* 124(2): 311–334.

Imai, Kosuke, and David A. van Dyk. 2005*b*. "MNP: R Package for Fitting the Multinomial Probit Model." *Journal of Statistical Software* 14(3): 1–32.

Kass, Robert E., and Adrian E. Raftery. 1995. "Bayes Factors." *Journal of the American Statistical Association* 90(130): 773–795.

Karp, Jeffrey A. 2009. "Candidate Effects and Spill-Over in Mixed Systems: Evidence from New Zealand." *Electoral Studies* 28(1): 41–50.

Karp, Jeffrey A., Jack Vowles, Susan A. Banducci, and Todd Donovan. 2002. "Strategic Voting, Party Activity, and Candidate Effects: Testing Explanations for Split Voting in New Zealand's New Mixed System." *Electoral Studies* 21(1): 1–22.

Klingemann, Hans-Dieter, and Bernhard Wessels. 2001. "The Political Consequences of Germany's Mixed-Member System: Personalization at the Grass Roots?" In *Mixed-Member Electoral Systems: The Best of Both Worlds?*, ed. Matthew Soberg Shugart, and Martin P. Wattenberg. Oxford: Oxford University Press pp. 279–296.

Lewis-Beck, Michael S., and Richard Nadeau. 2004. "Split-Ticket Voting: The Effects of Cognitive Madisonianism." *The Journal of Politics* 66(1): 97–112.

Linhart, Eric and Johannes Raabe and Patrick Statsch. 2018. "Mixed-Member Proportional Electoral Systems—The Best of Both Worlds?" *Journal of Elections, Public Opinion and Parties*. Published Online. https://doi.org/10.1080/17457289.2018.1443464

Maeda, Ko. 2008. "Re-Examining the Contamination Effect of Japan's Mixed Electoral System Using the Treatment-Effects Model." *Electoral Studies* 27(4): 723–731.

Massicotte, Louis, and André Blais. 1999. "Mixed Electoral Systems: A Conceptual and Empirical Survey." *Electoral Studies* 18(3): 341–366.

Meffert, Michael F., and Thomas Gschwend. 2010. "Strategic Coalition Voting: Evidence from Austria." *Electoral Studies* 29(3): 339–349.

Mellor, David Hugh. 2002. *The Facts of Causation*. New York: Routledge.

Monfardini, Chiara, and Rosalba Radice. 2007. "Testing Exogeneity in the Bivariate Probit Model: A Monte Carlo Study." *Oxford Bulletin of Economics and Statistics* 70(2): 271–282.

Moser, Robert G., and Ethan Scheiner. 2004. "Mixed Electoral Systems and Electoral System Effects: Controlled Comparison and Cross-National Analysis." *Electoral Studies* 23(4): 575–599.

Moser, Robert G., and Ethan Scheiner. 2005. "Strategic Ticket Splitting and the Personal Vote in Mixed-Member Electoral Systems." *Legislative Studies Quarterly* 30(2): 259–276.

Moser, Robert G., and Ethan Scheiner. 2009. "Strategic Voting in Established and New Democracies: Ticket Splitting in Mixed-Member Electoral Systems." *Electoral Studies* 28(1): 51–61.

Page, Benjamin I., and Calvin C. Jones. 1979. "Reciprocal Effects of Policy Preferences, Party Loyalties and the Vote." *The American Political Science Review* 73(4): 1071–1089.

Pappi, Franz Urban, and Paul W. Thurner. 2002. "Electoral Behaviour in a Two-Vote System: Incentives for Ticket Splitting in German Bundestag Elections." *European Journal of Political Research* 41(2): 207–232.

Poguntke, Thomas, and Paul Webb. 2005. *The Presidentialization of Politics*. Oxford: Oxford University Press.

Rahn, Wendy M., John H. Aldrich, Eugene Borgida, and John Sullivan. 1990. "A Social Cognitive Model of Candidate Appraisal." In *Information and Democratic Processes*, ed. John Ferejohn, and James Kuklinski. Champaign: University of Illinois Press pp. 187–206.

Samuels, David J. 2000. "The Gubernatorial Coattails Effect: Federalism and Congressional Elections in Brazil." *The Journal of Politics* 62(1): 240–253.

Appendix A: Additional Results

Table A1: Descriptive Statistics

Variable	Party	Mean	Std. Err.
Local Chances	CDU	0.816	0.191
	SPD	0.547	0.252
	Greens	0.338	0.253
	FDP	0.206	0.220
	Left	0.143	0.196
Local Ratings	CDU	0.212	0.409
	SPD	0.158	0.365
	Greens	0.015	0.122
	FDP	0.006	0.074
	Left	0.004	0.061
Party Ratings	CDU	0.633	0.296
	SPD	0.596	0.254
	Greens	0.481	0.287
	FDP	0.307	0.279
	Left	0.229	0.286
Leader Ratings	CDU	0.688	0.316
	SPD	0.520	0.305
	Greens	0.390	0.289
	FDP	0.296	0.258
	Left	0.331	0.306
Coalition Ratings		0.524	0.299
Party ID	CDU	0.320	0.467
	SPD	0.177	0.382
	Greens	0.068	0.252
	FDP	0.022	0.146
	Left	0.024	0.152
Age		0.428	0.199
Education		0.652	0.349
Gender (Female = 1)		0.438	0.496

The table reports descriptive statistics of our variables for the 2,694 respondents with non-missing observations. All variables are scaled between 0 and 1.

This appendix provides additional details on the models used in the main paper. Our multinomial probit models are sampled with three chains, which allows the computation of the Gelman-Rubin potential scale reduction factors (PSRF) to assess whether they have converged to a stationary distribution (Gelman and Rubin 1992). Table A2 reports these statistics for the parameters of both vote equations. A value close to one indicates that convergence has been achieved. As can be seen, for both equations, all values fall well below the conventional benchmark of 1.1, suggesting that we have successfully reached convergence. The conclusion holds when considering the upper limit of the 97.5% credible interval for this statistic. Table A2 also reports the p-values from the Heidelberger and Welch (1983) convergence tests. In all cases, we obtain values larger than 0.05, supporting the (null) hypothesis of stationary distributions and strengthening the conclusion that we have successfully reached convergence.

Tables A3 to A5 replicate the main models presented in the paper using a frequentist approach and multinomial logit models (also called conditional logit models when they include alternative-specific variables). We fit a joint model in a fashion similar to seemingly unrelated regressions, using the *suest* pre-built command in the Stata software package. The simultaneous model is fitted internally by duplicating and stacking the dataset, using interaction terms with all parameters of one model, and then fitting the full model user standard errors clustered by respondent. We cannot reproduce exactly the models we used in the main paper for testing the role of voter sophistication, but we rely on split samples to fit the joint model for respondents with lower and higher levels of education. We test differences across models using Wald tests and report the p-values in Table A5. Most of the key findings mentioned in the paper are substantively the same when computed with this alternative approach. For instance, we also find a stronger contamination effect from the list vote to the candidate vote, yet only among the less sophisticated voters.

Table A2: Convergence Diagnostic Statistics

		List Vote			Candidate Vote		
Party	Parameter	GR	GR (97.5%)	HW-p	GR	GR (97.5%)	HW-p
	Local Chances	1.0000	1.0000	0.98	1.0004	1.0012	0.54
	Local Ratings	1.0001	1.0003	0.20	1.0002	1.0004	0.58
	Party Ratings	1.0002	1.0002	0.84	1.0013	1.0019	0.40
	Leader Ratings	1.0001	1.0002	0.10	1.0001	1.0002	0.50
	Contamination Effects	1.0000	1.0000	0.72	1.0001	1.0005	0.30
	Party ID	1.0002	1.0007	0.54	1.0005	1.0013	0.17
SPD	Coalition Ratings	1.0004	1.0008	0.39	1.0014	1.0016	0.75
	Age	1.0001	1.0003	0.52	1.0001	1.0003	0.87
	Education	1.0000	1.0000	0.57	1.0000	1.0001	0.11
	Gender	1.0001	1.0003	0.67	1.0000	1.0001	0.73
	Bavaria	1.0001	1.0003	0.69	1.0000	1.0002	0.52
	Intercept	1.0002	1.0006	0.39	1.0002	1.0002	0.89
Greens	Coalition Ratings	1.0009	1.0028	0.20	1.0002	1.0008	0.57
	Age	1.0001	1.0005	0.63	1.0002	1.0007	0.87
	Education	1.0000	1.0000	0.92	1.0000	1.0000	0.83
	Gender	1.0002	1.0008	0.51	1.0000	1.0001	0.69
	Bavaria	1.0002	1.0007	0.17	1.0001	1.0004	0.67
	Intercept	1.0002	1.0007	0.09	1.0001	1.0001	0.74
FDP	Coalition Ratings	1.0003	1.0012	0.22	1.0006	1.0016	0.11
	Age	1.0001	1.0005	0.62	1.0001	1.0002	0.58
	Education	1.0000	1.0000	0.20	1.0008	1.0025	0.61
	Gender	1.0001	1.0002	0.31	1.0002	1.0006	0.87
	Bavaria	1.0001	1.0002	0.81	1.0000	1.0000	0.80
	Intercept	1.0002	1.0006	0.28	1.0004	1.0012	0.26
Left	Coalition Ratings	1.0005	1.0011	0.47	1.0004	1.0011	0.59
	Age	1.0004	1.0014	0.06	1.0002	1.0006	0.75
	Education	1.0000	1.0001	0.97	1.0009	1.0028	0.49
	Gender	1.0000	1.0001	0.89	1.0001	1.0003	0.67
	Bavaria	1.0001	1.0005	0.12	1.0002	1.0006	0.38
	Intercept	1.0003	1.0010	0.18	1.0004	1.0014	0.86
Cov. Matrix	SPD:Greens	1.0007	1.0014	0.53	1.0007	1.0025	0.45
	SPD:FDP	1.0014	1.0047	0.06	1.0005	1.0019	0.40
	SPD:Left	1.0010	1.0022	0.29	1.0012	1.0036	0.51
	Greens:Greens	1.0004	1.0014	0.66	1.0002	1.0006	0.52
	Greens:FDP	1.0019	1.0068	0.23	1.0011	1.0012	0.81
	Greens:Left	1.0010	1.0027	0.25	1.0014	1.0018	0.11
	FDP:FDP	1.0001	1.0004	0.84	1.0006	1.0010	0.44
	FDP:Left	1.0009	1.0031	0.09	1.0007	1.0009	0.23
	Left:Left	1.0004	1.0012	0.17	1.0005	1.0006	0.28
Multivariate PSRF			1.0062			1.0039	

Convergence statistics for the main models reported in Table 3. The first two columns for each equation report the Gelman-Rubin (GR) potential scale reduction factor (PSRF) statistics along with the 97.5% upper bound of their credible interval. The third column reports the p-value of the Heidelberger-Welch (HW) diagnostic test of stationary distributions. Values larger than 0.05 indicate that the test is passed, supporting the conclusion of a stationary distribution. The tests are performed on a combined sample of 1,500,000 MCMC draws for each vote model.

Table A3: Joint Multinomial Logistic Models of List and Candidate Votes

		List Vote		Candidate Vote	
Party	Variable	Estimate	Confidence Interval	Estimate	Confidence Interval
	Local Chances	−0.206	[−0.634, 0.223]	1.114	[0.621, 1.606]
	Local Ratings	0.253	[−0.002, 0.508]	0.840	[0.536, 1.144]
	Party Ratings	5.383	[4.588, 6.178]	4.947	[4.199, 5.694]
	Leader Ratings	2.055	[1.493, 2.617]	1.521	[0.979, 2.062]
	Candidate → List	0.206	[−0.041, 0.453]		
	List → Candidate			0.406	[0.051, 0.761]
	Party ID	1.233	[1.028, 1.437]	0.838	[0.627, 1.049]
SPD	Coalition Ratings	−4.225	[−5.478, −2.972]	−4.324	[−5.460, −3.188]
	Age	−0.089	[−0.981, 0.804]	0.369	[−0.464, 1.203]
	Education	−0.500	[−1.028, 0.027]	0.405	[−0.071, 0.880]
	Gender	0.039	[−0.323, 0.400]	0.055	[−0.273, 0.382]
	Bavaria	−0.159	[−0.643, 0.324]	0.613	[0.144, 1.081]
	Intercept	2.456	[1.448, 3.465]	1.226	[0.299, 2.153]
Greens	Coalition Ratings	−4.032	[−5.433, −2.630]	−2.464	[−3.828, −1.100]
	Age	−0.877	[−2.089, 0.334]	0.105	[−1.052, 1.261]
	Education	0.298	[−0.420, 1.017]	0.512	[−0.171, 1.196]
	Gender	0.059	[−0.382, 0.500]	0.277	[−0.172, 0.726]
	Bavaria	0.247	[−0.316, 0.810]	0.353	[−0.202, 0.908]
	Intercept	1.239	[0.037, 2.442]	−0.139	[−1.301, 1.023]
FDP	Coalition Ratings	2.805	[1.544, 4.066]	0.236	[−1.621, 2.094]
	Age	−0.582	[−1.604, 0.439]	−3.000	[−4.881, −1.118]
	Education	0.154	[−0.457, 0.765]	−0.119	[−1.007, 0.768]
	Gender	−0.194	[−0.630, 0.242]	−0.225	[−0.835, 0.385]
	Bavaria	1.018	[0.260, 1.775]	1.465	[0.188, 2.742]
	Intercept	−2.229	[−3.538, −0.920]	−0.916	[−3.073, 1.241]
Left	Coalition Ratings	−3.326	[−5.053, −1.598]	−3.035	[−4.963, −1.106]
	Age	−0.696	[−2.382, 0.989]	0.440	[−1.148, 2.029]
	Education	−1.110	[−1.917, −0.304]	−0.156	[−0.959, 0.647]
	Gender	0.128	[−0.432, 0.689]	0.507	[−0.107, 1.122]
	Bavaria	−0.291	[−0.941, 0.360]	0.287	[−0.436, 1.011]
	Intercept	2.300	[0.901, 3.699]	0.210	[−1.161, 1.582]
	Observations		2,694		2,694

Joint estimation of frequentist, multinomial logistic models using a stacked dataset and an interaction variable for each election. The joint model is computed using clustered standard errors by respondent. 95% confidence intervals are reported between brackets.

Table A4: Joint Multinomial Logistic Models of List and Candidate Votes

	List Vote		Candidate Vote	
Variable	Mean	[Credible Interval]	Mean	[Credible Interval]
		Low Education (Sub-Sample with Education < 1)		
Local Chances	−0.031	[−0.613, 0.552]	0.784	[0.036, 1.532]
Local Ratings	0.329	[0.004, 0.654]	0.665	[0.276, 1.055]
Party Ratings	4.852	[3.951, 5.754]	4.519	[3.588, 5.449]
Leader Ratings	2.415	[1.726, 3.104]	1.866	[1.148, 2.583]
Candidate → List	0.322	[0.020, 0.623]		
List → Candidate			1.067	[0.515, 1.618]
Party ID	0.989	[0.704, 1.274]	0.906	[0.607, 1.205]
		High Education (Sub-Sample with Education = 1)		
Local Chances	−0.592	[−1.220, 0.036]	1.116	[0.451, 1.782]
Local Ratings	0.169	[−0.255, 0.594]	1.110	[0.597, 1.624]
Party Ratings	6.242	[4.742, 7.741]	5.622	[4.340, 6.903]
Leader Ratings	1.590	[0.664, 2.515]	1.216	[0.371, 2.060]
Candidate → List	0.117	[−0.306, 0.540]		
List → Candidate			−0.038	[−0.532, 0.456]
Party ID	1.427	[1.120, 1.734]	0.875	[0.550, 1.201]
Observations		2,694		2,694

Joint estimation of frequentist, multinomial logistic models using a stacked dataset and an interaction variable for each election, on two different subsamples (low and high education). The joint model is computed using clustered standard errors by respondent. 95% confidence intervals are reported between brackets.

Table A5: Hypothesis Testing, Frequentist Models

Hypothesis	H_1	H_0	p-value
		FULL SAMPLE	
Local Chances	$\beta^1 \neq \alpha^1$	$\beta^1 = \alpha^1$	0.0000
Local Ratings	$\beta^2 \neq \alpha^2$	$\beta^2 = \alpha^2$	0.0008
Party Ratings	$\alpha^3 \neq \beta^3$	$\alpha^3 = \beta^3$	0.2929
Leader Ratings	$\alpha^4 \neq \beta^4$	$\alpha^4 = \beta^4$	0.0742
Contamination	$\theta_{LC} \neq \theta_{CL}$	$\theta_{LC} = \theta_{CL}$	0.2826
		LOW EDUCATION (Sub-Sample with Education < 1)	
Local Chances	$\beta^1 \neq \alpha^1$	$\beta^1 = \alpha^1$	0.0271
Local Ratings	$\beta^2 \neq \alpha^2$	$\beta^2 = \alpha^2$	0.1338
Party Ratings	$\alpha^3 \neq \beta^3$	$\alpha^3 = \beta^3$	0.4990
Leader Ratings	$\alpha^4 \neq \beta^4$	$\alpha^4 = \beta^4$	0.1396
Contamination	$\theta_{LC} \neq \theta_{CL}$	$\theta_{LC} = \theta_{CL}$	0.0079
		HIGH EDUCATION (Sub-Sample with Education = 1)	
Local Chances	$\beta^1 \neq \alpha^1$	$\beta^1 = \alpha^1$	0.0000
Local Ratings	$\beta^2 \neq \alpha^2$	$\beta^2 = \alpha^2$	0.0011
Party Ratings	$\alpha^3 \neq \beta^3$	$\alpha^3 = \beta^3$	0.4172
Leader Ratings	$\alpha^4 \neq \beta^4$	$\alpha^4 = \beta^4$	0.4692
Contamination	$\theta_{LC} \neq \theta_{CL}$	$\theta_{LC} = \theta_{CL}$	0.5727

Wald tests of the null of equal coefficients, using estimates from the joint multinomial logistic regression models reported in Tables A3 and A4.