When Do Voters Know Their Candidates?
A Data-Driven Approach to the German Federal Election

Oliver Rittmann (Corresponding Author)[a]
Marie-Lou Sohnius[b]
Thomas Gschwend[c]

Declaration of interest: none.

[a]orittman@mail.uni-mannheim.de
University of Mannheim
School of Social Sciences
A5, 6
68159 Mannheim
Germany

[b]marie-lou.sohnius@nuffield.ox.ac.uk
University of Oxford
Nuffield College New Road
Oxford
OX1 1NF
United Kingdom

[c]gschwend@uni-mannheim.de
University of Mannheim
School of Social Sciences
A5, 6
68159 Mannheim
Germany

When Do Voters Know Their Candidates? A Data-Driven Approach to the German Federal Election

Abstract

Voters' knowledge about their candidates plays a central role in theories of representation and accountability. Many electoral system incentives also work through some nominal component when translating votes into seats. We know that candidate knowledge is not only conceptually crucial for voters to have but also brings about electoral support. Surprisingly, we do not know much about the determinants under which citizens know their candidates. In lieu of existing causal explanations, we compile data on many factors related to candidate knowledge and use machine learning to study which factors best predict candidate knowledge out-of-sample across multiple German Federal elections. We find factors that describe candidate-, voter-, and district characteristics politically to be important predictors in contrast to factors that describe them socio-demographically. These findings can be a starting point for developing causal theories and have implications for the electoral reform debate in Germany.

1 Introduction

When do voters know their political candidates? Almost all electoral systems include a nominal component. Some electoral systems require voters to explicitly cast a nominal vote for one or more particular candidates. Even in some list PR systems, voters can cast a preference vote for some candidates or at least can see a printed list of candidates' names next to the respective party label on a party-list ballot. Without at least recognizing a candidate's name on the ballot, however, voters cannot systematically use any nominal information in their decision-making process. Thus, knowing political candidates allows voters to attach importance to evaluating candidates when casting their votes. This knowledge of candidates seems to be a minimal requirement for any personalized geographic incentive system of democratic representation to work. The folklorist understanding of democratic accountability also builds on that: Voters act as "rational god of vengeance and reward" (Key, 1964, p. 568) and hold politicians accountable by either re-electing a satisfactory incumbent or by punishing them and voting for an opponent instead (Fiorina, 1981; Kramer, 1971).

In addition to improving the quality of representation, voters' knowledge of candidates also plays a vital role in electoral politics. Kam and Zechmeister (2013) show experimentally that name recognition increases candidate support (and not necessarily the other way around). Voter's albeit minimal knowledge about their candidates is a foundational assumption in the literature about institutional incentives for strategic voting (e.g., Cox, 1997; Gschwend and Meffert, 2017; Rheault et al., 2020) — to figure out who is viable and who should be deserted — as well as for the literature on incentives and consequences of cultivating personal votes (e.g., Cain, Ferejohn and Fiorina, 1987; Carey and Shugart, 1995; Shugart, Valdini and Suominen, 2005).

Here, we study voters' candidate knowledge in the run-up of federal elections in Germany, where the nature of voters' relationship with their elected members of parliament (MPs) takes center stage in ongoing electoral reform discussions. While Germany has a paradigmatic mixed-member proportional (MMP) system, the size of its parliament (*Bundestag*) grew substantially because of additional seats needed to compensate for so-called "overhang" seats to ensure that the overall seat distributions stay proportional to the parties' list-vote share.[1] One proposal for decreasing the parliament's size is to reduce the number of electoral districts, leading to fewer candidates being elected in the nominal tier and, hence, *a priori* minimizing the chances that such "overhang" seats occur. Critics of this proposal argue that reducing the number of electoral districts — which implies that the average district will cover a larger geographic area and there will be more citi-

[1] If a party wins more electoral districts in a state than its list vote share in this state would warrant it can hold onto those excess seats, which are referred to as "overhang" seats.

zens to be represented per district MP — makes it increasingly difficult for citizens to get to know their local representatives. This has potentially serious implications for citizens' satisfaction with the electoral system and the state of democracy and how responsive they perceive the political system to be. While the existing evidence refutes these arguments for the case of Germany (Sohnius, Gschwend and Rittmann, 2022), common wisdom nevertheless still assumes that there is an especially strong relationship between voters and their nominally elected district representative that list MPs do not have, even though, as is the case in Germany, almost all list MPs also compete at the district level. But if district incumbents are not better known than list incumbents that unsuccessfully run in an electoral district, then the unique relationship between citizens and their local district MP has to be questioned.

Despite the importance of candidate knowledge, especially in the context of such lively arguments, we know surprisingly little about the causes that determine whether voters know the candidates on their ballot. In our study, we investigate under which circumstances voters in Germany know or do not know their local candidates. To learn about the factors contributing to candidate knowledge, we pursue a data-driven approach and identify variables that strongly predict candidate knowledge in out-of-sample predictions. We consider three sets of explanatory variables: Candidate-level characteristics, voter-level characteristics, and district-level characteristics. To assess the predictive power of different predictors, we compile a new dataset of voter-candidate dyads based on election surveys from three recent federal elections in Germany (2009, 2013, 2017). We match these voter surveys with detailed information about respondents' electoral districts, the district candidates on the respondent's ballot, and the candidates' prior political careers. We then divide this data into training and test data and use the training data to train a random forest model predicting whether survey respondents can recall the names and parties of the candidates who compete in their electoral district (Breiman, 2001*b*). Finally, we evaluate the performance of the trained model out-of-sample on the hold-out test data.

Our contribution is to identify which conceivable explanatory factors matter empirically so that scholars can start developing parsimonious causal theories based on them that future research can test. One theme that runs through our results is that variables that describe candidates, voters, and districts *politically* seem to hold valuable information for predicting candidate knowledge. In contrast, variables that describe candidates, voters, and districts *socio-demographically* seem to have little value for predicting candidate knowledge. Our findings have important implications not only for future research on the determinants of candidate knowledge and the development of causal explanations but also for our understanding of MMP systems and the two types of legislators they generate (Manow, 2015; Stratmann and Baur, 2002; Klingemann and Wessels, 2001). It

seems that, based on the German data we analyze in this study, district MPs and list MPs are perceived less differently than often assumed.

2 What does potentially explain candidate knowledge?

We are interested in factors that explain voters' knowledge of local district candidates. Factors contributing to candidate knowledge are as manifold as research on the topic. Much work has been devoted to individual factors explaining candidate knowledge, such as the effects of campaign spending (Coleman and Manna, 2000), online advertisement (Broockman and Green, 2014), or the type of election (Parker, 1981). Yet we know little about the relative importance of causes determining whether voters know the candidates on their ballot.

Giebler and Weßels (2017) were among the first to comparatively analyze determinants of candidate knowledge. In their work, they differentiate between three explanatory blocks: Candidate-related factors, voter-related factors, and context-related factors. Even though we consider a different set of explanatory variables, we consider this a useful theoretical frame for studying candidate knowledge. Specifically, the frame allows us to speak to a broader debate about the importance of district candidates for local representation. The focus on candidate-level explanations allows us to study whether candidates who already represent the district as an incumbent are more well-known among the electorate than non-incumbents and, more crucially, district candidates who are already members of the Bundestag but who gained their seat through the party list after they lost the district race. Much research argues that there are two classes of incumbent MPs, the more local district MPs, and list MPs. The former are usually seen as strong representatives of local district interests, while the latter are argued to represent the party interest. If the difference between these two types of incumbents has no explanatory power to predict whether they are known among voters, then the value of a local district mandate, as opposed to a non-local list mandate for local representation, may be overstated. The theoretical frame also opens the door to analyzing whether specific contextual factors that pertain to the electoral-district level, such as the geographic size of electoral districts, play a significant role in the personal link between voters and their local representatives. If, for example, the size of electoral districts helps explain candidate knowledge in a way where voters in larger districts are less likely to know their candidate, then this would indicate that increasing the size of electoral districts could be harmful to the quality of local personal representation.

2.1 Candidate-Level Explanations

The first set of explanatory factors for candidate knowledge focuses on the candidates themselves. This group of potential predictors is motivated by the idea that candidates can have specific attributes that make them more or less widely known among the electorate. We consider six factors that we group into aspects related to the political career of the candidate and aspects related to personal characteristics. A first expectation is that candidates' publicity should grow with the size of their electorate. In this context, it seems plausible that a candidate who receives 40% of the district votes is more well known than a candidate who receives 5% of the district votes. Closely connected to this idea is that there may be two front runners in a district race who get the majority of all votes and that the public attention is focused on those candidates.

Second, we consider party affiliation as a potential predictor of candidate knowledge. The majority of candidate votes (*Erststimme*) in Germany are won by the Christian Democratic Party group (CDU/CSU) and the Social Democrats (SPD). Furthermore, there is variation in the emphasis that different party groups put on local representation. For example, the CSU is known to emphasize the importance of district candidates for local representation. While this is plausibly a function of the number of district MPs a party has in their parliamentary group, it is reasonable to expect that voters should more widely know candidates of parties usually winning electoral district races and emphasizing the role of district MPs.

A third predictor that naturally comes into mind is the incumbency status of candidates. Previous studies have found that incumbents are more well-known among voters than candidates who do not already hold office (Kam and Zechmeister, 2013). However, the mixed-member system of the German Bundestag creates two different types of incumbents and therefore renders a binary classification of candidates into incumbents and non-incumbents insufficient. Following the two-vote principle, mixed-member systems open two ways of political representation. First, citizens can elect candidates to parliament via the candidate vote. They are nominally elected with a plurality of votes within their electoral district. Second, voters can help candidates into office with their second vote, with which they elect candidates from a party list. Importantly, candidacy in mixed-member systems is not always mutually exclusive. In practice, this creates a strategic incentive for legislators to pursue a dual candidacy and run concurrently in an electoral district and on a party list. This dual candidacy maximizes their chances of getting elected. The list candidacy offers district candidates a fallback option if they lose the district vote. In consecutive elections, legislators who lost their district vote but gained a seat through the party list regularly run in the district race again.

The German electoral system thus creates two types of incumbents in district races:

Candidates who won the district race at the previous election and candidates who lost the district race at the last election but gained a seat through the party list. Previous research considered only the first type as incumbents (Giebler and Weßels, 2017), but this could be an oversimplification. There may be different normative expectations for members of the Bundestag who won their seat through the district vote as opposed to those who won the seat through the party list. But despite the normative difference in their roles, the mandates do not differ. We thus have a particular interest in whether incumbency, especially incumbent type, predicts candidate knowledge.

2.2 Voter-Level Explanations

The second set of explanatory factors focuses on voter characteristics. This group of variables reflects the notion that voters differ from one another and that those differences may make them more or less knowledgeable of district candidates. We can roughly group this set of explanations into factors related to the political identity of voters, including their, for example, political interest and ideological leaning; and socioeconomic factors. One set of explanatory factors in the political domain describes the general relationship between a voter and the democratic system. These factors include voters' political interest, whether they are satisfied with the democratic system, their political knowledge about how the German electoral system works, whether they think that local representation is important in the German political system and whether they intend to turn out to vote at the upcoming election. This set of predictors is motivated by the assumption that voters with a more positive relationship with the political system should also be more informed about the actors in that system (Grönlund and Milner, 2006). Therefore, variables related to citizens' relationship with the political system may help to predict their knowledge of candidates.

Beyond the general relationship between voters and the political system, we also expect information about their ideological identity to plausibly help predicting candidate knowledge. Voters should be more likely to know the candidate they intend to vote for, even more so if it is the candidate of a party they identify with. The information conveyed in this set of explanations should not only help predict whether voters know local candidates in general but also predict *which* candidates they know.

As the third set of voter-level explanations, we consider socioeconomic factors. These include age (different cohorts are more or less invested in politics (Frazer and Macdonald, 2003)), gender (previous research found differences in political knowledge between men and women—even though this is now debated (Kraft and Dolan, 2022)), education (higher education may be correlated with higher political knowledge (Grönlund and Milner, 2006)), and voters' economic situation (voters who experience financial hardship may

have less capacity to engage with politics).

2.3 District-Level Explanations

Finally, we consider a set of district-level explanations. Here we have a particular interest in constituency characteristics. One group of predictors in this domain focuses on the general features of the electoral district. The set of features includes the population size, geographic size, and population density in the electoral district. A larger electoral district, in population or area, may make it more difficult for candidates to make themselves prominent in the district. Population density reflects the notion that it may not be the number of voters in a district that makes it more difficult for candidates to become well known but the concentration of voters within the district.

The second group of district-level predictors concerns the political competition in the electoral district. Here, we consider the electoral race's competitiveness and the district's number of competitors. We also consider the potential consequences of a reform of the electoral districts in 1998 that led to a restructuring of multiple electoral districts. This disruption of the local political landscapes may have affected the relationship between voters and district delegates.

Predictor Variable	Range/Categories	Mean	Description
Candidate-Level Predictors			
Voteshare at t	$[0.01, 0.60]$	0.19	Realized district vote share of the candidate at the upcoming election.
Party	{CDU, ..., AfD}	–	Party of the candidate.
Status at $t-1$	{New Candidate, District Incumbent, List Incumbent}	0.64 0.15 0.20	Incumbency Status in the previous legislative period. *New Candidate:* Did not run before; *District Incumbent:* District winner of previous election; *List Incumbent:* list MP, lost in district at previous election.
Frontrunner	$\{0, 1\}$	0.43	Is the candidate among the top-2 candidates in the district?
Age	$[18, 78]$	47.81	Age of the candidate in election year (Election year − Year of birth).
Female	$\{0, 1\}$	0.33	Does the candidate identify as a women?
Voter-Level Predictors			
Political interest	$[1, 5]$	2.67	Self-reported level of political interest from high (1) to low (5).
Party identification	$\{0, 1\}$	0.20	Does the respondent identify with the party of the candidate?
Voted for candidate	$\{0, 1\}$	0.20	Did the respondent intend to vote for the candidate?
Satisfaction with democracy	$[1, 5]$	2.52	Self-reported satisfaction with democracy in Germany from high (1) to low (5).
Turnout intention	$\{0, 1\}$	0.93	Self-reported intention to turn out to vote in the upcoming election.
Local representation important	$[1, 5]$	2.08	Agreement with "The MP should represent all citizens in the electoral district." from high (1) to low (5).
Political Knowledge	$\{0, 1\}$	0.54	Respondent correctly answered "Which vote decides how many seats each party will have in parliament?"
Age	$[18, 99]$	55.17	Age of the respondent in election year (Election year − Year of birth).
Female	$\{0, 1\}$	0.48	Does the respondent identify as a women?
Highschool	$\{0, 1\}$	0.34	Does the respondent hold a highschool degree?
Subjective economic situation	$[1, 5]$	2.38	Self-reported satisfaction with own economic situation from high (1) to low(5).
District-Level Explanations			
Population size	$[197.6, 377.4]$	276.29	Population size of the electoral district in 1000.
Geographic size	$[26.9, 6250.3]$	1322.86	Geographic size of the electoral district in km^2.
Population density	$[0.04, 12.63]$	0.87	Population density of the electoral district ($\frac{\text{Population Size}}{\text{Geogr. Size}}$).
Effective number of candidates	$[2.17, 5.82]$	3.67	Effective number of candidates in the electoral district ($\frac{1}{N}\sum_{i=1}^{N} p_i^2$).
Winning margin	$[0.00, 0.51]$	0.15	Difference in vote shares between district winner and second placed candidate.
Electorate change	$\{0, 1\}$	0.14	$\geq 50\%$ of the district's electorate changed through 1998 electoral district reform.

Table 1: Overview of all variables used to predict district candidate knowledge during the 2009, '13, and '17 federal elections in Germany.

3 Data & Methods

To gain insights about factors that matter for candidate knowledge, we compile a dataset based on pre-election surveys from three federal elections in Germany in 2017, 2013, and 2009 (GLES, 2019*a*,*b*,*c*). Within these surveys, respondents were asked whether they could recall the names and parties of district candidates running in their local constituency at the federal election.[2]

We match these voter surveys with detailed information from the German Federal Elections Officer about respondents' electoral districts, including population size and geographic size. We further match respondents with their district candidates, including information about the district candidates' demographic characteristics and their prior political careers.

The resulting data set includes 33,868 unique voter-candidate pairs. As every electoral district in every election has a different set of candidates running, each voter is paired with multiple candidates. The number of candidate pairs for one voter thus depends on the number of running candidates within their electoral district and varies within and between districts across time. Our dependent variables indicate whether the voter could name the specific candidate's name and party.[3] In the survey sample, 54.9% of the respondents know at least one local candidate. This knowledge differs vastly by the party. Historically, the vast majority of electoral districts were either won by the Christian Democratic party group (CDU/CSU) or by the Social Democrats (SPD). This is reflected in a much higher candidate knowledge of candidates of those parties than other parties' candidates. The most widely known candidates come from the CSU in Bavaria: Every second survey respondent in Bavaria (52.3%) was able to name the candidate of the CSU in their district. This rate is lower for the CDU and SPD candidates. About one out of three survey respondents named the local candidate of the CDU (36.2%) and the SPD (34.4%), respectively. These numbers drop for the district candidates of other parties. Only one out of four survey respondents (26.1%) could name at least one candidate running for another party.

To measure the independent variables, we compile data from various sources. All

[2]The exact wording of this question is documented in Appendix A. Note, there are a few pure list candidates each election, i.e., candidates that do not run in any electoral districts. Respondents are not asked about them in the survey measure we use.

[3]We do not count respondents who were able to call a candidate but confuse the party or are unable to name a party at all. However, this does not happen very often. In 2013 and 2017, conditional on being able to call a candidate, respondents in our data were able to name the correct party in 91.9% of the cases. In 3.6% of the cases, those survey respondents named an incorrect party; in 4.4%, they did not name any party. The survey data from the 2009 pre-election survey records candidate knowledge binary, indicating only whether respondents named the candidate together with the correct party or not.

predictor variables are listed and summarized in table 1. The first set of predictor variables describes the candidates themselves. To measure the size of a candidate's electorate, we use their vote share in the election year. It is important to note that this measure is only realized *after* the election and thus cannot be used to predict candidate knowledge in the future. We still prefer it to pre-election measures because survey-based measures of voting intentions are not sufficiently accurate on the district level. To measure incumbency status, we differentiate between three groups: Non-incumbents are candidates who do not hold a mandate in the Bundestag; district incumbents are candidates who have won the district in the previous election; list incumbents are members of the Bundestag who gained their seat via the party list. Importantly, the vast majority (96.8%) of the list incumbents ran unsuccessfully in the district before. All other candidate-level predictors are straightforward to measure: "Party" is a categorical predictor that denotes the party of a candidate, "Frontrunner" indicates whether a candidate was among the top 2 candidates in the district, "Age" measures the age of the candidate in the election year, and "Female" indicates whether the candidate identifies as a woman.

The second set of predictors describes voters. Here, we draw on a battery of survey items to measure political interest, whether a respondent identifies with the party of the district candidate, whether they reported voting for the candidate,[4] whether they are satisfied with how democracy works in Germany, whether they intend to turn out to vote at the upcoming election, whether they think that local representation is important, whether they are knowledgeable about the electoral system in Germany, and their age, gender, education, and subjective economic situation. The scales of all variables, as well as the question items for each variable, are summarized in table 1.

The final set of predictors describes the electoral districts. Here, we include the population size of the district, the geographic size of the district, and population density. Two additional variables describe the electoral competition in the district. As a measure of the number of competitors, we include the effective number of candidates (Laakso and Taagepera, 1979). As a measure of the competitiveness of the district race, we include the winning margin of the district winner (e.g., Gschwend, 2007). Finally, we include a variable that indicates whether at least 50% of the electorate changed due to an electoral district reform in 1998. This applies to 30 electoral districts that were most severely affected by the electoral district reform (Eisel and Graf, 2002).

We compile all these variables in one data set. After discarding all observations with at least one missing value on any of the variables, we are left with a data set of 11,245 voter-candidate pairs. This data set comprises 1,784 unique voters and 1,985 unique

[4]To express the intent to vote for a candidate, respondents did not need to know the name of the candidate. It was sufficient if they reported using their district vote to vote for the candidate of a specific party.

candidates and covers 86.6% of the district races from 2009 to 2017.

4 Predicting candidate knowledge

In the previous section, we collected a wide range of variables that possibly predict voters' candidate knowledge. Our next goal is to study which of those variables contain information that helps us to predict candidate knowledge. The predominant approach to such a task involves statistical models that assume a specific stochastic process, e.g., a logistic regression model. In this framework, the probability of a voter's candidate knowledge would be modeled as a transformation of some linear combination of our independent variables. This comes down to assuming to know the functional form of the relationship between the independent variables and the outcome, even though this is often not true (Breiman, 2001*b*). Beyond that, because this classical approach treats the functional form of a regression model and the set of independent variables as known, it puts little emphasis on model evaluation (Athey and Imbens, 2019). That is, it relies on standard errors as measures of uncertainty for a specific statistical model's parameters, but rarely asks whether parameters estimated within one model enable us to predict the outcome based on new data. But if the specified model is incorrect, then this assessment of uncertainty has limited value. In cases where there is little theoretical guidance about how a set of predictor variables is related to the outcome, pre-specifying a statistical model appears as a suboptimal idea as such a model requires us to make assumptions without theoretical backing that is likely to lead to incorrect conclusions. Under such circumstances, predictive modeling and machine learning offer a viable alternative. Instead of using theory to set up a statistical model, predictive modeling adopts a more inductive approach and treats the data-generating process as unknown. Instead, the data is used to determine the functional form of the relationship between the independent variables and the outcome (Molina and Garip, 2019; Grimmer, Roberts and Stewart, 2021). To assess modeling uncertainty and prevent overfitting, this approach uses a train-test set logic where a model's predictive abilities are evaluated based on data that was not used during the estimation of the model.

In light of these arguments, scholars have increasingly turned towards machine learning models to study various contexts (see for example Lupu and Warner (2022) and Kim and Zilinsky (2022)). Given the sparseness of theory surrounding the explanation of candidate knowledge, we adopt such a data-driven approach in this study. We start by splitting the data set into a train and test set, with 75% of the data going into the train set and 25% being reserved in the test set. Next, we use the training data to train and fine-tune a random forest model (Breiman, 2001*a*; Montgomery and Olivella, 2018) using the

Measure	Naive Model	Random Forest
Percentage of correctly predicted	0.681	0.784
Sensitivity (true-positive rate)	0.000	0.567
Specificity (true-negative rate)	1.000	0.891

Table 2: Random Forest Model Evaluation. Evaluation scores are based on out-of-sample predictions on a held-out test set. The naive model predicts the modal category (voter does not know the candidate) and serves as a benchmark for the random forest model.

`R`-package `mlr3` (Lang et al., 2019).[5]

Before turning to the importance of specific variables for predicting candidate knowledge, we start by checking how well our model predicts candidate knowledge in general. For that, we apply the trained random forest model to the hold-out test set. Table 2 shows the out-of-sample performance scores and benchmarks them against a naive model that predicts the modal category for each observation, i.e., any voter does not know any candidate. The trained random forest model performs adequately on the hold-out test set: For almost eight out of ten respondent-candidate pairs (78.4%), the model correctly predicts whether the respondent knows the candidate. The model is better able to correctly predict candidate knowledge for voter-candidate pairs where the voter does not know the candidate (89.1%) than when the voter *does* know the candidate (56.7%). Overall, these results indicate that our set of predictor variables indeed stores information that is predictive of voters' knowledge of district candidates in the run-up to the federal elections in Germany.

4.1 Which factors matter most for the prediction of candidate knowledge?

In the next step, we are interested in which of our predictive variables are most important for the model's ability to predict candidate knowledge and which variables are least important. That is, we want to learn about the relevant factors for predicting candidate knowledge. Figure 1 presents helpful quantities of interest to infer this — the variable importance measures for each predictive variable in the trained random forest model. The scores represent the rate by which the percentage of incorrectly predicted voter-candidate

[5]Three hyperparameters of the random forest model are tuned using random search (Bergstra and Bengio, 2012). Specifically, the search space contains the number of trees ($\texttt{n}_{\texttt{trees}} \in [20, 100]$), the number of randomly sampled variables used as candidates at each split ($\texttt{m}_{\texttt{try}} \in [2, 10]$), and the minimum number of observations in terminal nodes of a tree ($\texttt{nodesize} \in [10, 50]$). Random search is performed ten times. We select the model with the lowest classification error based on tenfold cross-validation in the test set as the best-fitting model (Neunhoeffer and Sternberg, 2019).

pairs increases when the information stored in one variable is taken from the model.[6] If this number equals one, this means that withholding the variable's information from the model does not decrease the model's predictive performance. We calculate variable importance scores based on in-sample and out-of sample predictions. We do this because a variable may have predictive value in the data that was used to train the model, but this may be a result of overfitting. To investigate the relevance of variables for the prediction of candidate knowledge, it is thus essential to check whether they help predict observations that were not used to train the model.

The first result is that while the trained model attributes at least some importance to the entire range of predictors based on the training data, once we apply the model to unseen data, the set of predictors contributing to the model's predictive ability shrinks considerably. This confirms our approach and underlines the importance of evaluating predictive models out of sample, as in-sample analysis may make some variables appear important even though they are not.[7]

One theme that runs through our results is that variables that describe candidates, voters, and districts *politically* appear to hold valuable information for the prediction of candidate knowledge, while variables that describe candidates, voters, and districts *socio-demographically* do not contribute to the prediction of candidate knowledge. Starting with candidate-level predictors, we find that candidates' age and gender carry little to no value in predicting candidate knowledge, even though the model emphasizes candidate age in the training data. Variables that *are* important to predict whether a candidate is widely known among voters are related to their electoral and political success. This is reflected by the fact that the realized vote shares of district candidates is the single most important predictor in our model, and their incumbency status is the third most important predictor.

The results on the voter level draw a similar picture: Variables that describe voters

[6]Precisely, we take the information from the model by shuffling the values of the predictor and recalculating the classification error of the trained model.

[7]It is important to point out that there are two theoretical reasons why a variable may seem important in the training data but unimportant in the test data. The first reason is that the model learned patterns in the training data that are unique to the training data and therefore do not generalize to the test data. This is what is called overfitting and we consider it to be the driving force behind disparities between in-sample and out-of-sample measures of variable importance. The second potential reason is that the model learns patterns that generalize beyond the training data but are not present in the test set due to data sparseness. For example, the model may learn from the training data that women between 60 and 65 with a high school degree have high knowledge of candidates. Suppose this is a general pattern, but by chance, there are simply no women between 60 and 65 with a high school degree in the test data. In this case, the model would have learned something meaningful from the predictors gender, age, and education. Still, it would not become visible in the out-of-sample variable importance assessment. Given our 75-25 train-test-split, we consider this option less likely, but we cannot rule it out. We are also fairly convinced that if the second option has an impact on our out-of-sample importance estimates, those impacts should not be so large that they change any of our substantive conclusions.

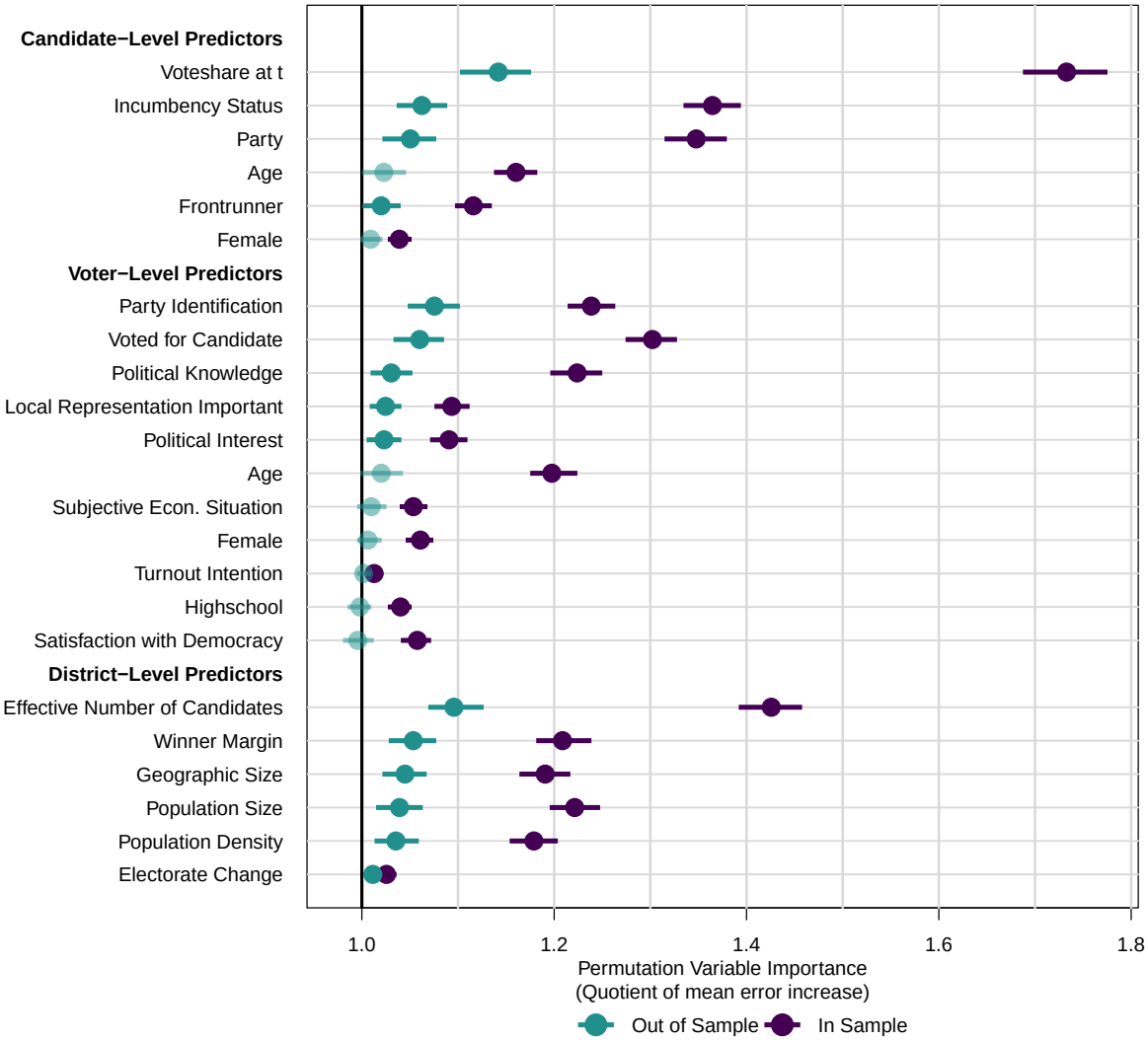

Figure 1: Permutation variable importance of all predictors in the random forest model. To calculate the importance score of one predictor variable, we shuffle the values of the predictor and recalculate the classification error of the model. This gives us an idea about how the model performs if we withhold the information of the specific predictor. The more the classification error decreases, the more important the variable for the predictive performance of the model. We divide the classification error of the permutation data set by the classification error of the full data set. If this quotient is equal to one, then withholding the information of the predictor has no effect on the predictive performance of the model and the variable has no importance for the predictive power of the model. Higher scores indicate higher variable importance. For each variable, we repeat the procedure 500 times. Points represent the mean variable importance, lines the center 95% of the distribution of estimates. Points are depicted transparently if this interval includes one. We show variable importance scores based on the training data set (in sample predictions) and on the test data set (out of sample predictions). The difference highlights the importance to keep potential overfitting when analyzing the model: A variable may seem important to predict outcomes in the training set, but this often does not generalize to the test set. To evaluate the informational value of specific variables, it is thus important to focus their contribution to out of sample predictions.

politically seem to be valuable to predict their knowledge of district candidates, while socio-demographic characteristics do not play an important role. Whether a respondent identifies with the candidates' party and whether they intend to vote for a candidate are the two most informative variables to predict whether a specific voter knows a particular candidate, followed by political knowledge about the political system, whether voters think that local representation is important, and political interest. While the relevance of political characteristics does not come as a surprise, it is remarkable how unimportant voters' socio-demographic characteristics are for predicting their candidate knowledge. Neither voters' formal education nor their age and gender seem to carry information that helps us to predict their candidate knowledge.

On the district level, we again observe a superiority of variables that describe electoral districts politically over variables that describe electoral districts demographically—even though here, variables like geographic size, population size, and population density seem to improve the model's prediction to some extent. Yet, the most important variables to predict candidate knowledge within an electoral district are closely connected to the political contest in the district. That is, the effective number of candidates within that district, followed by the winning margin in the district.

Another interesting result on the district level is that major disruptions of electoral districts in 1998 (Electorate Change) do not help to predict the level of candidate knowledge in the elections 2009–2017. We do not want to interpret this result in the sense that reforms of electoral districts do not affect local candidate knowledge. Still, the result suggests that more than ten years after the reform, there are no dramatic differences in the level of candidate knowledge between districts most highly disrupted by the reform and others.

Taken together, the results suggest that candidate knowledge is not the result of either candidates' or voters' socio-demographic characteristics. Instead, what matters for candidate knowledge is a candidate's ability to prevail in the political and electoral contest, the voter's political identity, and the nature of the electoral competition within a district. While these factors help us predict candidate knowledge, we should not make the mistake and interpret the results causally. For example, our model indicates that a candidate's electoral success is the top predictor of their prominence among voters. Still, it provides no answer to where a candidate's electoral success comes from. It is neither able to differentiate between causal directions (are candidates electorally successful because they are prominent among voters?; or are candidates prominent among voters because they are politically successful?), nor is it able to say anything about the roots of political and electoral success, which may, for example, be a function of their party's electoral success.

4.1.1 Direction of Effects

Our primary goal is to learn about the predictive value of a wide set of variables for voters' candidate knowledge. Yet, we are also interested in whether the most important variables of the trained model influence the predictions of the random forest model in a way that is in line with what we would expect theoretically. For this purpose, we select the three most important variables of each set of predictors and investigate how the model's average predicted probabilities change as a function of these variables. Figure 2 shows the results of this exercise.[8]

The results are mostly consistent with what theoretical expectations would suggest. Starting with candidate-level predictors in the top row of figure 2, we observe that higher votes shares of a candidate and already being an incumbent in the run-up of the election are associated with higher probabilities of being known among voters. Moreover, we observe higher predicted probabilities for candidates of the parties that traditionally win districts races (CDU/CSU and SPD), compared to other parties. These results are hardly surprising but confirm that the model learned sensible relationships.

One observation on the candidate level stands out to us: While incumbency status matters in general, there seems to be almost no difference in candidate knowledge with respect to the type of incumbency. Having won the district in the previous election seems to come with virtually no gain in prominence compared to candidates who only entered the parliament via the party list. Recall that the vast majority of list incumbents in our data (96.8%) ran in the district before but did not win the race. This implies that list incumbents were less electorally successful in the district prior to the election than district incumbents. Yet, by virtue of their list mandate, they seem to be almost as widely known in the district as district incumbents. Or in other words, district incumbents seem to enjoy no advantages over the list incumbents regarding their prominence in the electoral district.

Turning to the voter level, we find that respondents who identify with a candidate's party and who vote for a candidate are more likely to know the candidate. The same holds for respondents who know the German electoral system sufficiently well to know which of their two votes is decisive for the overall composition of the parliament—but the magnitude of this effect is much smaller than the magnitude of party identification and vote choice.

Regarding the effective number of candidates within an electoral district, we find that a higher number of candidates is associated with lower probabilities of candidate knowledge. To make sense of this result, it is helpful to remind ourselves about what

[8]It is again important to emphasize that none of the graphs allow a causal interpretation.

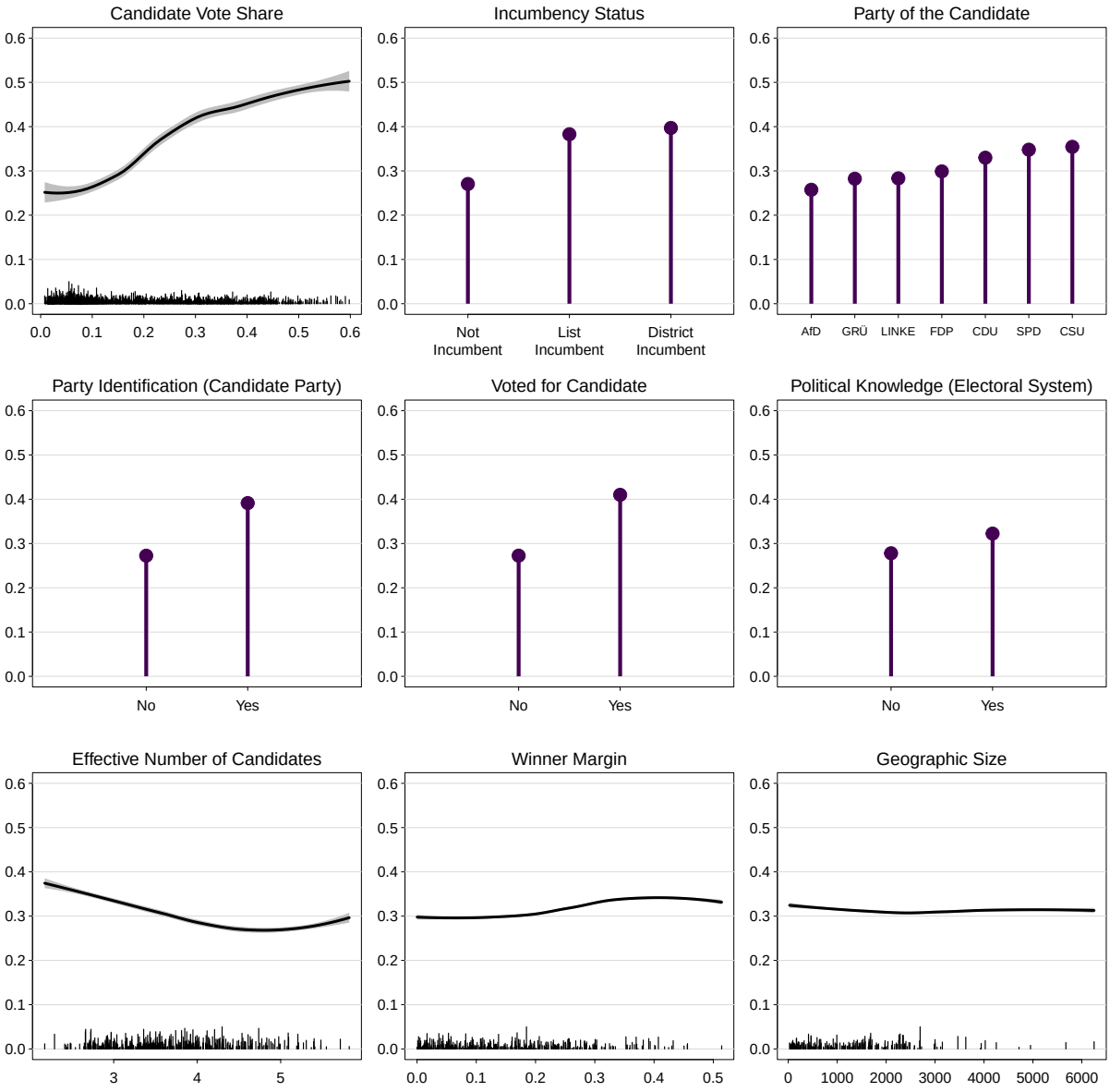

Figure 2: Partial Dependence Plots for the top three most important variables on the candidate level (top row), voter level (center row), and district level (bottom row). x-axes represent the predictor variables, y-axes show predicted probabilities to know a district candidate by the random forest model. Predicted probabilities conditional on a specific value c of the covariate x_k are calculated by creating a replicate of the predictor matrix X, replacing all observed covariate values x_k with c, using the trained model to predict probabilities for each unit in the replicate matrix, and averaging across all those predicted probabilities. Each panel shows how the model's predicted probability changes across the empirical range of the variables. The lines represent local polynomial regression lines fitted to the predicted probabilities with 95%-Confidence Intervals. It is important to note that these confidence intervals *do not* quantify sampling uncertainty around the predicted probabilities. Panels with continuous predictors also show the empirical distribution of the variable in the full data set (train and test set) at the bottom of the panels.

unique information the effective number of candidates adds to the model that is not captured by other variables.[9] Since the model has access to a candidate's vote share, what the effective number of candidates adds is information about the number of other auspicious candidates that the candidate competes in the district race. Thus, the negative association between the effective number of candidates and candidate knowledge suggests that a candidate's chances of being known decrease the more (serious) competitors they face. This may indicate that voters have a limited capacity to recall candidates' names and are overwhelmed when there are four or five equally promising competitors for a seat in their district.

Finally, we gather little information from the partial dependencies on the remaining two district-level variables. Candidate knowledge slightly increases when there is a very high winning margin (30 percentage points), but this does not happen very often and is thus based on relatively few observations. No clear trend is observable regarding the geographic size, but the variable's value may stem from interactions with other variables that are not visible in the aggregate.

4.2 Knowledge of Candidates who are already MPs

The previous analysis investigated the predictability of the knowledge of all candidates running for any of the parties that made it into parliament in the respective election year. Competition at the district level always includes candidates without a real chance of winning the district seat. To narrow the scope, we now focus on candidates who are already MPs. Notably, this does not only include district incumbents but also list incumbents.[10] One main goal of this analysis is to further investigate the remarkable finding from the previous section that district incumbents are no more prominent in their district compared to candidates who "only" hold a list mandate. Formally, only district incumbents are representatives of an electoral district. Thus one could argue that district incumbents should be more widely known in their district than members of the Bundestag who may be connected to the district because they ran in it before but whose mandate is not formally tied to the electoral district.

The subset of voter-incumbent candidate pairs comprises 4081 observations and is thus substantially smaller than the full data set. Incumbents were known in about every

[9] After all, the effective number of candidates is a function of the vote shares of all candidates in the district, and the candidate's vote share entered the model as a separate variable.

[10] Note that not all list incumbents of the *Bundestag* run in an electoral district, but many do. Our analysis does not consider candidates who exclusively run on a party list, as they have no unambiguous connection to a specific electoral district. Most district candidates who hold a list mandate are candidates who lost the district race at a previous election but entered the Bundestag via their party's list.

Measure	Naive Model	Random Forest
Percentage of correctly predicted	0.503	0.737
Sensitivity (true-positive rate)	0.000	0.702
Specificity (true-negative rate)	1.000	0.771

Table 3: Random forest model evaluation for the model trained on the subset of incumbent candidates. Evaluation scores are based on out-of-sample predictions on a held-out test set. The naive model predicts the modal category (voter does not know the candidate) and serves as a benchmark for the random forest model.

second voter-candidate pair (50.3%), providing a fairly balanced data set. We again use a 75:25 ratio to split the data into training and test data and train a random forest model following the same procedure as above. Table 3 shows the out-of-sample performance of the random forest model trained on the subset of candidates who are already MPs. The trained random forest model is able to correctly predict candidate knowledge in about three out of four voter-candidate pairs (73.7%), substantially improving predictive accuracy compared to the naive baseline model. Given the more balanced sample, it is no surprise that the Sensitivity-Specificity difference of the incumbent model is much less pronounced than in the full data model. Our incumbent model is only slightly better able to predict the lack of knowledge among those who do not know a candidate (77.1%) than the knowledge of candidates among those who know the candidate (70.2%).

Figure 3 shows variable importance measures of the random forest model trained on the subset of candidates who are already members of the Bundestag. The variables that have been most important to predict candidate knowledge of all candidates remain largely the same when focusing on only district and list incumbent candidates. Among the top most important predictors in the model are still candidate vote share (even though the importance shrank relative to other predictors), whether respondents intend to vote for a candidate or identify with the candidate's party, and the effective number of candidates. At the same time, the model confirms the finding that the socioeconomic characteristics of both candidates and voters seem to carry no information that helps us predict candidate knowledge.

Remarkably, the incumbency status that differentiates between list and district incumbents slipped down to the least important variable among all candidate-level predictors. This confirms the result that what matters for candidate knowledge is whether candidates hold a mandate in the parliament, but not how they got there — through winning their district or through their party list.

Figure 4 presents partial dependencies of the four most important variables of our incumbent model, confirming the findings from the full model. Voting for a candidate is associated with a higher probability of knowing the candidate, more competitors within a

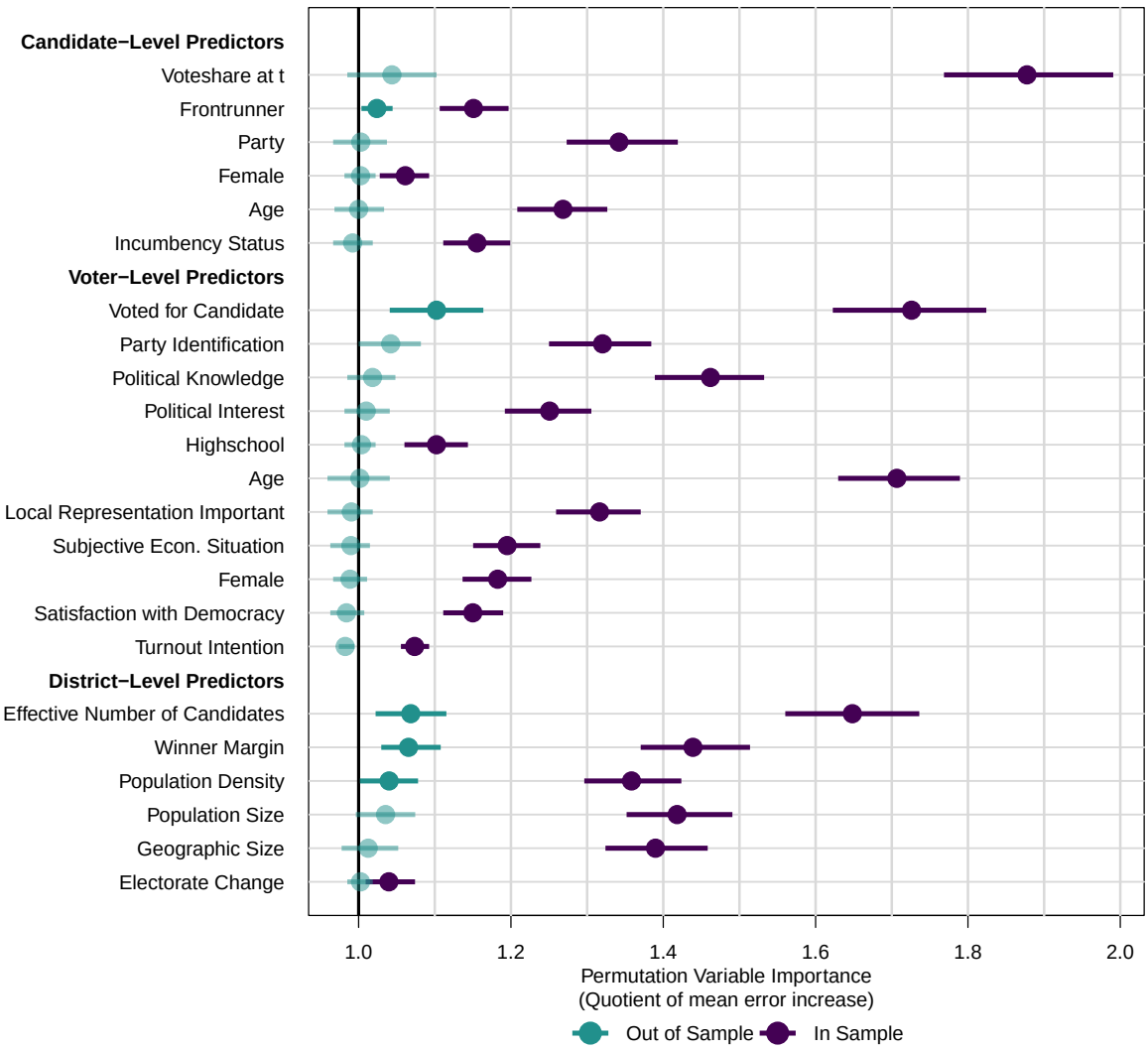

Figure 3: Permutation variable importance of all predictors in the random forest model fitted on the subset of candidates who are already members of the Bundestag prior to the election.

district are associated with lower probabilities of knowing candidates within that district, and a higher candidate vote share is associated with higher probabilities of knowing the candidate. Also, candidates in districts in which one candidate is far above all other candidates seem to be more well-known than candidates in more competitive districts. Still, this association only takes effect above an exceptional winning margin of about 30 percentage points.

The average predicted probability of knowing a candidate conditional on incumbency type increases by only 1.6 percentage points, from 49.5% to 51.1%, for district incumbents compared to list incumbents. Together with the previous results, this questions whether there is a unique value of being a district MP for being known within the electoral district, compared to a list MP.

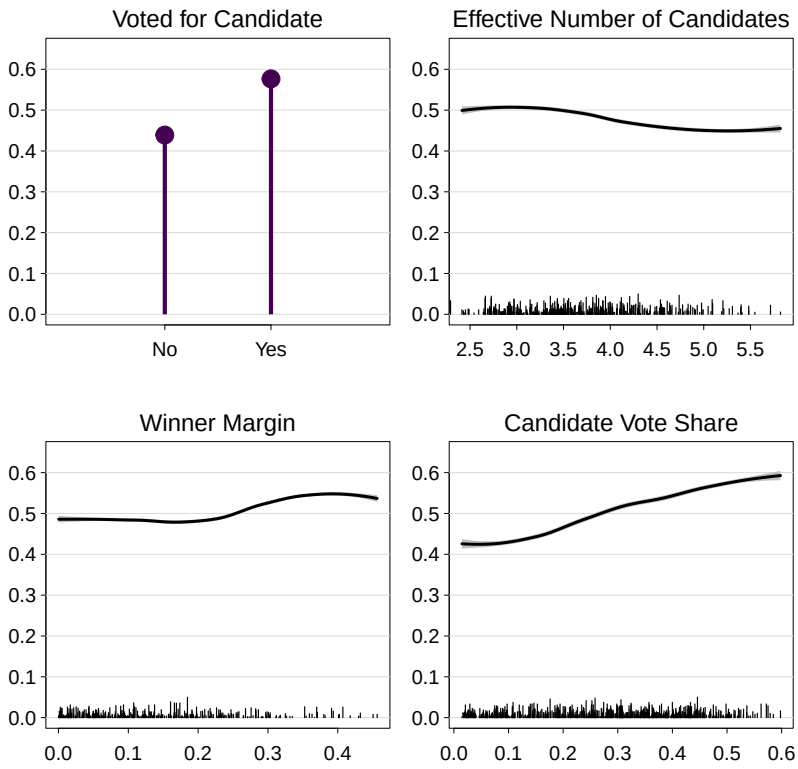

Figure 4: Partial dependence plots for the four most important variables. x-axes represent the predictor variables, y-axes show predicted probabilities to know a district candidate based on our random forest model.

5 Conclusion

Being able to name local candidates in the run-up of federal elections in Germany is far from general knowledge. Survey evidence suggests that about every second voter knows at least one candidate, one out of three voters can recall at least two candidates, and about 15% can name three or more candidates. While voters' candidate knowledge is an essential topic in democratic theory and a must-have for them, at least in some minimal form, for representation and accountability to work, current research is surprisingly innocent about its causes. We show how scholars can responsibly use a data-driven research design to identify potential explanatory factors from a kitchen-sink list of conceivable factors by treating the supposed data-generating process as unknown. In lieu of solid theoretical guidance, predictive modeling and machine learning can offer a viable alternative to assuming the functional form of the relationship between various explanatory factors and voters' candidate knowledge. Our criterion for identifying potentially explanatory factors is whether and how important they are in predicting voters' candidate knowledge as measured in surveys. Specifically, our quantity of interest is the *permutation variable importance* of each predictor in our random forest model. We rigorously evaluate the

model's predictive ability out-of-sample, i.e., based on data that was not used during the estimation stage.

Typically, MMP systems allow for dual candidacies, i.e., candidates' can compete in both tiers, the nominal tier as well as the party-list tier, at the same time. Consequently, incumbency status cannot be as clearly conceptualized as in first-past-the-post systems. There are two different types of incumbents depending on their election mode: district incumbents, who won the nominal district race in the previous election, and, list incumbents, who were elected through their party's list even if they might have lost their district race. We find that there is no difference between both incumbent types in terms of predicting voters' candidate knowledge. Respondents in our data do not know their district incumbents better than list incumbents. This might be surprising as candidates elected through a party list should have a priori no strong incentives to make themselves known to potential voters so that they can recall their name and party affiliation correctly. As dual candidates, however, even list incumbents have such incentives. They compete in their local district as well, running more candidate-centered election campaigns (Gschwend and Zittel, 2015; Zittel and Gschwend, 2008) and serve there as 'shadow' district representatives to increase their chances of winning this district the next time or serve as the local representative of their party (Lundberg, 2006; Manow, 2015). This finding has also implications for the current electoral reform debate in Germany. The supposed importance of district incumbents for local representation seems to be more of a myth used by some to discredit particular reform proposals. At least our finding that district incumbents are not better known than list incumbents suggests that voters do not share this myth. District MPs are, for that matter, no better MPs than list MPs.

Moreover, we find that political characteristics of either candidates, voters, or districts are more predictive of voters' candidate knowledge than social-demographic characteristics. This is normatively pleasing as existing inequalities in the propensity to know candidates that exist in society seem to get channeled only through politically charged characteristics such as vote intention, party identification, and political knowledge into actual candidate knowledge of voters. Based on these results, we suggest that those characteristics should be put center stage in future research to build causal theories under what conditions voters know their candidates. Scholars should then develop implications of such theories and test them with causal research designs.

A limitation of our study is given by its restriction to survey respondents who answered all questions that we used to predict candidate knowledge. It follows that our results describe the subset of survey participants with no missing values on the relevant survey items, and generalizability to the population of all survey participants and beyond is a matter of uncertainty. One concern is that missingness may be correlated with candidate knowledge and one or more predictor variables. For example, respondents without high

school education may be less likely to answer all required questions and at the same time less likely to know district candidates. If there is also a connection between the predictor and the propensity to answer all relevant survey questions, then restricting the analysis to respondents without missing values could bias the results regarding the importance of the respective predictor.

Finally, we would like to note that our findings about candidate knowledge depend on the assumption that knowing a candidate's name and party affiliation is diagnostic for thinking about them when making decisions. While we know of no research contradicting this assumption, it is yet conceivable that voters can remember a candidate's party affiliation but cannot recall their names (or vice versa). Fortunately, there are only a few respondents that do that. It is harder to imagine that respondents who neither recall their name nor party do actually consider a candidate's identity seriously.

References

Athey, Susan and Guido W. Imbens. 2019. "Machine Learning Methods That Economists Should Know about." *Annual Review of Economics* 11:685–725.

Bergstra, James and Yoshua Bengio. 2012. "Random Search for Hyper-Parameter Optimization." *Journal of Machine Learning Research* 13:281–305.

Breiman, Leo. 2001*a*. "Random Forests." *Machine Learning* 45:5–32.

Breiman, Leo. 2001*b*. "Statistical Modeling: The Two Cultures." *Statistical Science* 16(3):199–215.

Broockman, David E. and Donald P. Green. 2014. "Do Online Advertisements Increase Political Candidates' Name Recognition or Favorability? Evidence from Randomized Field Experiments." *Political Behavior* 36(2):263–289.

Cain, Bruce E., John A. Ferejohn and Morris P. Fiorina. 1987. *The Personal Vote: Constituency Service and Electoral Independence*. Cambridge, Mass.: Harvard University Press.

Carey, John M. and Matthew Soberg Shugart. 1995. "Incentives to cultivate a personal vote: A rank ordering of electoral formulas." *Electoral Studies* 14(4):417–439.

Coleman, John J and Paul F Manna. 2000. Congressional Campaign Spending and the Quality of Democracy. Technical Report 3.

Cox, Gary W. 1997. *Making Votes Count: Strategic Coordination in the World's Electoral Systems*. New York, NY: Cambridge University Press.

Eisel, Stephan and Jutta Graf. 2002. "Bundestagswahl 2002 – Die umstrittenen Wahlkreise.".

Fiorina, Morris P. 1981. *Retrospective Voting in American National Elections*. New Haven: Yale University Press.

Frazer, Elizabeth and Kenneth Macdonald. 2003. "Sex Differences in Political Knowledge in Britain." *Political Studies* 51(1):67–83.

Giebler, Heiko and Bernhard Weßels. 2017. "If You Don't Know Me by Now: Explaining Local Candidate Recognition." *German Politics* 26(1):146–169.

GLES. 2019*a*. "Vorwahl-Querschnitt (GLES 2009)." GESIS Datenarchiv, Köln. ZA5300 Datenfile Version 5.0.2, https://doi.org/10.4232/1.13228.

GLES. 2019*b*. "Vorwahl-Querschnitt (GLES 2013)." GESIS Datenarchiv, Köln. ZA5700 Datenfile Version 2.0.2, https://doi.org/10.4232/1.13231.

GLES. 2019*c*. "Vorwahl-Querschnitt (GLES 2017)." GESIS Datenarchiv, Köln. ZA6800 Datenfile Version 5.0.1, https://doi.org/10.4232/1.13234.

Grimmer, Justin, Margaret E. Roberts and Brandon M. Stewart. 2021. "Machine Learning for Social Science: An Agnostic Approach." *Annual Review of Political Science* 24:395–419.

Grönlund, Kimmo and Henry Milner. 2006. "The Determinants of Political Knowledge in Comparative Perspective." *Scandinavian Political Studies* 29(4):386–406.

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 Michael F. Meffert. 2017. Strategic Voting. In *The SAGE Handbook of Electoral Behaviour*, ed. Kai Arzheimer, Jocelyn Evans and Michael S. Lewis-Beck. Los Angeles: Sage chapter 16, pp. 339–366.

Gschwend, Thomas and Thomas Zittel. 2015. "Do constituency candidates matter in German Federal Elections? The personal vote as an interactive process." *Electoral Studies* 39:338–349.

Kam, Cindy D. and Elizabeth J. Zechmeister. 2013. "Name Recognition and Candidate Support." *American Journal of Political Science* 57(4):971–986.

Key, V. O. 1964. *Politics, parties and pressure groups*. New York, NY: Thomas Y. Crowell.

Kim, Seo-young Silvia and Jan Zilinsky. 2022. "Division Does Not Imply Predictability: Demographics Continue to Reveal Little About Voting and Partisanship." *Political Behavior* pp. 1–21.

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.

Kraft, Patrick W. and Kathleen Dolan. 2022. "Asking the Right Questions: A Framework for Developing Gender-Balanced Political Knowledge Batteries." *Political Research Quarterly* .

Kramer, Gerald H. 1971. "Short-Term Fluctuations in Us Voting Behavior, 1896-1964." *American Political Science Review* 65(1):131–143.

Laakso, Markku and Rein Taagepera. 1979. ""Effective" Number of Parties." *Comparative Political Studies* 12(1):3–27.

Lang, Michel, Martin Binder, Jakob Richter, Patrick Schratz, Florian Pfisterer, Stefan Coors, Quay Au, Giuseppe Casalicchio, Lars Kotthoff and Bernd Bischl. 2019. "mlr3: A modern object-oriented machine learning framework in R." *Journal of Open Source Software* 4(44):1903.

Lundberg, Thomas Carl. 2006. "Second-class representatives? Mixed-member proportional representation in Britain." *Parliamentary Affairs* 59(1):60–77.

Lupu, Noam and Zach Warner. 2022. "Why are the affluent better represented around the world?" *European Journal of Political Research* 61(1):67–85.

Manow, Philip. 2015. *Mixed Rules, Mixed Strategies: Parties and Candidates in Germany's Electoral System.* ECPR Press.

Molina, Mario and Filiz Garip. 2019. "Machine Learning for Sociology." *Annual Review of Sociology* 45:27–45.

Montgomery, Jacob M. and Santiago Olivella. 2018. "Tree-Based Models for Political Science Data." *American Journal of Political Science* 62(3):729–744.

Neunhoeffer, Marcel and Sebastian Sternberg. 2019. "How Cross-Validation Can Go Wrong and What To Do About It." *Political Analysis* 27(1):101–106.

Parker, Glenn R. 1981. "Interpreting Candidate Awareness in U. S. Congressional Elections." *Legislative Studies Quarterly* 6(2):219–233.

Rheault, Ludovic, Andre Blais, John H Aldrich and Thomas Gschwend. 2020. "Understanding people's choice when they have two votes." *Journal of Elections, Public Opinion and Parties* 30(4):466–483.

Shugart, Matthew Soberg, Melody Ellis Valdini and Kati Suominen. 2005. "Looking for Locals: Voter Information Demands and Personal Vote-Earning Attributes of Legislators under Proportional Representation." *American Journal of Political Science* 49(2):437–449.

Sohnius, Marie-Lou, Thomas Gschwend and Oliver Rittmann. 2022. "Welche Auswirkungen haben größere Wahlkreise auf das politische Verhalten? Ein empirischer Beitrag zur Wahlrechtsreform." *Politische Vierteljahresschrift, forthcoming* .

Stratmann, Thomas and Martin Baur. 2002. "Plurality Rule, Proportional Representation, and the German Bundestag: How Incentives to Pork-Barrel Differ across Electoral Systems." *American Journal of Political Science* 46(3):506–514.

Zittel, Thomas and Thomas Gschwend. 2008. "Individualised Constituency Campaigns in Mixed-Member Electoral Systems: Candidates in the 2005 German Elections." *West European Politics* 31(5):978–1003.

Online Appendices:

When Do Voters Know Their Candidates?
A Data-Driven Approach to the German Federal Election

A GLES Question Wording and Variable Codes

Here, we list all items from the German Longitudinal Election Studies in 2009, 2013, and 2017 that were used in the analysis. For variables with multiple answer categories that were recoded to a binary variable, underlined numbers (e.g. (02)) indicate the category coded as 1. Categories with italic numbers are coded as missing values (e.g. *(98)*).

Candidate Knowledge

2009: `q81m1 - q81m5`:

Kennen Sie den Namen von einem oder vielleicht sogar mehreren der hiesigen Wahlkreiskandidaten oder -kandidatinnen und können Sie mir sagen, für welche Partei diese bei der Bundestagswahl am 27. September 2009 antreten? Bitte nennen Sie mir den Namen und die Partei der Kandidatinnen bzw. Kandidaten.

2013: `q82a - q82e`:

Kennen Sie den Namen von einem oder vielleicht sogar mehreren der hiesigen Wahlkreiskandidaten und können Sie mir sagen, für welche Partei diese bei der Bundestagswahl am 22. September 2013 antreten? Bitte nennen Sie mir den Namen und die Partei der Kandidaten.

2017: `q77a1 - q77f1`:

Kennen Sie den Namen von einem oder vielleicht sogar mehreren der hiesigen Wahlkreiskandidaten und können Sie mir sagen, für welche Partei diese bei der Bundestagswahl am 24. September 2017 antreten? Bitte nennen Sie mir den Namen und die Partei der Kandidaten.

⇒ Coded 1 if respondent named the candidate and the party, 0 otherwise.

Party Identification

2009: `q139a`:

Und jetzt noch einmal kurz zu den politischen Parteien. In Deutschland neigen viele Leute längere Zeit einer bestimmten politischen Partei zu, obwohl sie auch ab und zu eine andere Partei wählen. Wie ist das bei Ihnen: Neigen Sie - ganz allgemein gesprochen - einer bestimmten Partei zu? Und wenn ja, welcher?

2013: `q119a`:

Und nun noch einmal kurz zu den politischen Parteien. In Deutschland neigen viele Leute längere Zeit einer bestimmten politischen Partei zu, obwohl sie auch ab und

zu eine andere Partei wählen. Wie ist das bei Ihnen: Neigen Sie - ganz allgemein gesprochen - einer bestimmten Partei zu? Und wenn ja, welcher?

2017: `q99a`:

Und nun noch einmal kurz zu den politischen Parteien. In Deutschland neigen viele Leute längere Zeit einer bestimmten politischen Partei zu, obwohl sie auch ab und zu eine andere Partei wählen. Wie ist das bei Ihnen: Neigen Sie - ganz allgemein gesprochen - einer bestimmten Partei zu? Und wenn ja, welcher?

⇒ Coded 1 if respondent indicated to identify with the party of the candidate, 0 otherwise.

Voted for Candidate

2009: `q11aa`:

Bei der Bundestagswahl können Sie ja zwei Stimmen vergeben. Die Erststimme für einen Kandidaten aus Ihrem Wahlkreis und die Zweitstimme für eine Partei. Hier ist ein Musterstimmzettel, ähnlich wie Sie ihn bei der Bundestagswahl erhalten. Was werden Sie auf Ihrem Stimmzettel ankreuzen? Bitte nennen Sie mir jeweils die Kennziffer für Ihre Erst- und Zweitstimme.

(A) jetzt bitte für die Erststimme

2013: `q11aa`:

Bei der Bundestagswahl können Sie ja zwei Stimmen vergeben. Die Erststimme für einen Kandidaten aus Ihrem Wahlkreis und die Zweitstimme für eine Partei. Hier ist ein Musterstimmzettel, ähnlich wie Sie ihn bei der Bundestagswahl erhalten. Was werden Sie auf Ihrem Stimmzettel ankreuzen? Bitte nennen Sie mir jeweils die Kennziffer für Ihre Erst- und Zweitstimme.

(A) Erststimme

2017: `q11a`:

Bei der Bundestagswahl können Sie ja zwei Stimmen vergeben. Die Erststimme für einen Kandidaten aus Ihrem Wahlkreis und die Zweitstimme für eine Partei. Hier ist ein Musterstimmzettel, ähnlich wie Sie ihn bei der Bundestagswahl erhalten. Was werden Sie auf Ihrem Stimmzettel ankreuzen? Bitte nennen Sie mir jeweils die Kennziffer für Ihre Erst- und Zweitstimme. Jetzt bitte für die Erststimme.

⇒ Coded 1 if respondent indicated to vote for the candidate, 0 otherwise. If respondent already voted via mail, the mail vote was taken instead of intended vote choice.

Political Knowledge

2009: q6:

Bei der Bundestagswahl haben Sie ja zwei Stimmen, eine Erststimme und eine Zweitstimme. Wie ist das eigentlich, welche der beiden Stimmen ist ausschlaggebend für die Sitzverteilung im Bundestag?

(01) die Erststimme
(02) die Zweitstimme
(03) beide sind gleich wichtig
(98) weiß nicht
(99) keine Angabe

2013: q7:

Bei der Bundestagswahl haben Sie ja zwei Stimmen, eine Erststimme und eine Zweitstimme. Wie ist das eigentlich, welche der beiden Stimmen ist ausschlaggebend für die Sitzverteilung im Bundestag?

(01) die Erststimme
(02) die Zweitstimme
(03) beide sind gleich wichtig
(98) weiß nicht
(99) keine Angabe

2017: q7:

Bei der Bundestagswahl haben Sie ja zwei Stimmen, eine Erststimme und eine Zweitstimme. Wie ist das eigentlich, welche der beiden Stimmen ist ausschlaggebend für die Sitzverteilung im Bundestag?

(01) die Erststimme
(02) die Zweitstimme
(03) beide sind gleich wichtig
(98) weiß nicht
(99) keine Angabe

Local Representation Important

2009: q84b:

Es gibt unterschiedliche Auffassungen darüber, wen ein Abgeordneter repräsentieren

soll. Wie wichtig ist Ihnen das Folgende. Bitte sagen Sie mir den zutreffenden Wert auf dieser Skala. Der Abgeordnete sollte alle Bürger im Wahlkreis repräsentieren.

(01) überhaupt nicht wichtig

(02)

(03)

(04)

(05) sehr wichtig

(98) weiß nicht

(99) keine Angabe

⇒ Recoded to match the scale in 2013 and 2019.

2013: `q95b`:

Es gibt unterschiedliche Auffassungen darüber, wen ein Abgeordneter repräsentieren soll. Wie wichtig ist Ihnen das Folgende. Bitte sagen Sie mir den zutreffenden Wert auf dieser Skala. Der Abgeordnete sollte alle Bürger im Wahlkreis repräsentieren.

(1) sehr wichtig

(2) wichtig

(3) mittelmäßig

(4) nicht so wichtig

(5) überhaupt nicht wichtig

(-98) weiß nicht

(-99) keine Angabe

2017: `q90b`:

Es gibt unterschiedliche Auffassungen darüber, wen ein Abgeordneter repräsentieren soll. Wie wichtig ist Ihnen das Folgende. Bitte sagen Sie mir den zutreffenden Wert auf dieser Skala. Der Abgeordnete sollte alle Bürger im Wahlkreis repräsentieren.

(01) überhaupt nicht wichtig

(02)

(03)

(04)

(05) sehr wichtig

(98) weiß nicht

(99) keine Angabe

Political Interest

2009: q2:

Einmal ganz allgemein gesprochen: Wie stark interessieren Sie sich für Politik: sehr stark, ziemlich stark, mittelmäßig, weniger stark oder überhaupt nicht?

(01) sehr stark
(02) ziemlich stark
(03) mittelmäßig
(04) weniger stark
(05) überhaupt nicht
(98) weiß nicht
(99) keine Angabe

2013: q3:

Einmal ganz allgemein gesprochen: Wie stark interessieren Sie sich für Politik: sehr stark, ziemlich stark, mittelmäßig, weniger stark oder überhaupt nicht?

(01) sehr stark
(02) ziemlich stark
(03) mittelmäßig
(04) weniger stark
(05) überhaupt nicht
(98) weiß nicht
(99) keine Angabe

2017: q3:

Einmal ganz allgemein gesprochen: Wie stark interessieren Sie sich für Politik: sehr stark, ziemlich stark, mittelmäßig, weniger stark oder überhaupt nicht?

(01) sehr stark
(02) ziemlich stark
(03) mittelmäßig
(04) weniger stark
(05) überhaupt nicht
(98) weiß nicht
(99) keine Angabe

Age

2009: `q1a`

Sagen Sie mir bitte, wie alt Sie sind.

2013: `q2a`

Würden Sie mir bitte Ihr Geburtsdatum nennen?

2017: `q2a`

Würden Sie mir bitte sagen, in welchem Jahr Sie geboren wurden? Und in welchem Monat? Und an welchem Tag?

Subjective Economic Situation

2009: `q18`:

Und nun zu Ihrer wirtschaftlichen Lage. Wie beurteilen Sie Ihre derzeitige eigene wirtschaftliche Lage? Bitte sagen Sie es mir anhand dieser Liste.

(01) sehr gut
(02) gut
(03) teils/teils
(04) schlecht
(05) sehr schlecht
(98) weiß nicht
(99) keine Angabe

2013: `q17`:

Und nun zu Ihrer wirtschaftlichen Lage. Wie beurteilen Sie Ihre derzeitige eigene wirtschaftliche Lage? Bitte sagen Sie es mir anhand dieser Liste.

(01) sehr gut
(02) gut
(03) teils/teils
(04) schlecht
(05) sehr schlecht
(98) weiß nicht
(99) keine Angabe

2017: `q15`:

Und nun zu Ihrer wirtschaftlichen Lage. Wie beurteilen Sie Ihre derzeitige eigene wirtschaftliche Lage? Bitte sagen Sie es mir anhand dieser Liste.

(01) sehr gut

(02) gut

(03) teils/teils

(04) schlecht

(05) sehr schlecht

(98) weiß nicht

(99) keine Angabe

Female

2009: `q1-1`:

Ist die Zielperson männlich oder weiblich?

(01) männlich

(02) weiblich

2013: `q1-1`:

Ist die Zielperson männlich oder weiblich?

(01) männlich

(02) weiblich

2017: `q1`:

Ist die Zielperson männlich oder weiblich?

(01) männlich

(02) weiblich

Turnout Intention

2009: `q9`:

Wenn Wahlen stattfinden, geben viele Leute ihre Stimme ab, andere kommen nicht dazu, ihre Stimme abzugeben oder nehmen aus anderen Gründen nicht an der Wahl teil. Einmal angenommen, Sie wären schon wahlberechtigt: Wie wahrscheinlich würden Sie dann am 27. September an der Bundestagswahl teilnehmen?

(01) bestimmt zur Wahl gehen

(02) wahrscheinlich zur Wahl gehen

(03) vielleicht zur Wahl gehen

(04) wahrscheinlich nicht zur Wahl gehen

(05) bestimmt nicht zur Wahl gehen

(98) weiß nicht

(99) keine Angabe

(100) trifft nicht zu

2013: `q10`:

Wenn Wahlen stattfinden, geben viele Leute ihre Stimme ab, andere kommen nicht dazu, ihre Stimme abzugeben oder nehmen aus anderen Gründen nicht an der Wahl teil. Einmal angenommen, Sie wären schon wahlberechtigt: Wie wahrscheinlich würden Sie dann .. am 22. September zur Bundestagwahl gehen.

(1) bestimmt zur Wahl gehen

(2) wahrscheinlich zur Wahl gehen

(3) vielleicht zur Wahl gehen

(4) wahrscheinlich nicht zur Wahl gehen

(5) bestimmt nicht zur Wahl gehen

(6) habe bereits per Briefwahl meine Stimme abgegeben

(-97) trifft nicht zu

(-98) weiß nicht

(-99) keine Angabe

2017: `q10`:

Wenn Wahlen stattfinden, geben viele Leute ihre Stimme ab, andere kommen nicht dazu, ihre Stimme abzugeben oder nehmen aus anderen Gründen nicht an der Wahl teil. Einmal angenommen, Sie wären schon wahlberechtigt: Wie wahrscheinlich würden Sie dann am 24. September an der Bundestagswahl teilnehmen?

(1) bestimmt zur Wahl gehen

(2) wahrscheinlich zur Wahl gehen

(3) vielleicht zur Wahl gehen

(4) wahrscheinlich nicht zur Wahl gehen

(5) bestimmt nicht zur Wahl gehen

(-97) trifft nicht zu

(-98) weiß nicht

(-99) keine Angabe

⇒ Vote intention was also coded as 1 if respondent indicated to have already vote via mail.

Highschool

2009: `d206`:

Welchen allgemeinbildenden Schulabschluss haben Sie?

(01) Schule beendet ohne Abschluss

(02) Hauptschulabschluss, Volksschulabschluss, Abschluss der polytechnischen Oberschule 8. oder 9. Klasse

(03) Realschulabschluss, Mittlere Reife, Fachschulreife oder Abschluss der polytechnischen Oberschule 10. Klasse

(04) Fachhochschulreife (Abschluss einer Fachoberschule etc.)

(05) Abitur bzw. erweiterte Oberschule mit Abschluss 12. Klasse (Hochschulreife)

(06) anderen Schulabschluss, und zwar:

(09) bin noch Schüler

(98) weiß nicht

(99) keine Angabe

2013: `q163`:

Welchen allgemeinbildenden Schulabschluss haben Sie?

(01) Schule beendet ohne Abschluss

(02) Hauptschulabschluss, Volksschulabschluss, Abschluss der polytechnischen Oberschule 8. oder 9. Klasse

(03) Realschulabschluss, Mittlere Reife, Fachschulreife oder Abschluss der polytechnischen Oberschule 10. Klasse

(04) Fachhochschulreife (Abschluss einer Fachoberschule etc.)

(05) Abitur bzw. erweiterte Oberschule mit Abschluss 12. Klasse (Hochschulreife)

(06) anderen Schulabschluss, und zwar:

(09) bin noch Schüler

(98) weiß nicht

(99) keine Angabe

2017: `q136`:

Welchen allgemeinbildenden Schulabschluss haben Sie?

(01) Schule beendet ohne Abschluss

(02) Hauptschulabschluss, Volksschulabschluss, Abschluss der polytechnischen Oberschule 8. oder 9. Klasse

(03) Realschulabschluss, Mittlere Reife, Fachschulreife oder Abschluss der polytechnischen Oberschule 10. Klasse

(04) Fachhochschulreife (Abschluss einer Fachoberschule etc.)

(05) Abitur bzw. erweiterte Oberschule mit Abschluss 12. Klasse (Hochschulreife)

(06) anderen Schulabschluss, und zwar:

(09) bin noch Schüler

(98) weiß nicht

(99) keine Angabe

Satisfaction with Democracy

2009: q5:

Wie zufrieden oder unzufrieden sind Sie - alles in allem - mit der Demokratie, so wie sie in Deutschland besteht? Sind Sie ... sehr zufrieden, ziemlich zufrieden, teils/teils, ziemlich unzufrieden oder sehr unzufrieden?

(01) sehr zufrieden

(02) ziemlich zufrieden

(03) teils/teils

(04) ziemlich unzufrieden

(05) sehr unzufrieden

(98) weiß nicht

(99) keine Angabe

2013: q6:

Wie zufrieden oder unzufrieden sind Sie - alles in allem - mit der Demokratie, so wie sie in Deutschland besteht? Sind Sie ... sehr zufrieden, ziemlich zufrieden, teils/teils, ziemlich unzufrieden oder sehr unzufrieden?

(01) sehr zufrieden

(02) ziemlich zufrieden

(03) teils/teils

(04) ziemlich unzufrieden

(05) sehr unzufrieden

(98) weiß nicht

(99) keine Angabe

2017: q6:

Wie zufrieden oder unzufrieden sind Sie - alles in allem - mit der Demokratie, so wie sie in Deutschland besteht? Sind Sie ... sehr zufrieden, ziemlich zufrieden, teils/teils, ziemlich unzufrieden oder sehr unzufrieden?

(01) sehr zufrieden

(02) ziemlich zufrieden

(03) teils/teils

(04) ziemlich unzufrieden

(05) sehr unzufrieden

(98) weiß nicht

(99) keine Angabe