Online Appendix: Measuring Swing Voters with a Supervised Machine Learning Ensemble

Christopher Hare[†] Mikayla Kutsuris[‡]

August 12, 2022

Contents

[†]Department of Political Science, University of California, Davis. cdhare@ucdavis.edu
[‡]McGeorge School of Law, University of the Pacific. m_kutsuris@u.pacific.edu

A Variables from the 2012 Cooperative Campaign Analysis Project and VOTER Study Group

The 2012 Cooperative Campaign Analysis Project (Jackman et al., 2012) includes three survey waves (a pre-election baseline, mid-election, and post-election) of 43,998 respondents conducted between December 2011 and December 2012. The economic, social, immigration, and environment scores are estimated using an ordinal IRT model (Imai, Lo and Olmsted, 2016). The remaining scores (religiosity, racial resentment, political knowledge, political interest, ideological inconsistency, and the death penalty issue scores) are estimated using simple summated rating scales.

Our measure of ideological inconsistency is adopted from Federico and Hunt (2013) and combines three standardized indices: the standard deviation of the issue responses (all responses are coded in the same ideological direction), the proportion of issue responses that are ideologically consistent with respondent party affiliation (including leaners), and the average distance between the issue responses and self-placement on the liberal-conservative scale. We use sixteen issue responses to items concerning abortion, affirmative action, climate change, the death penalty, global warming, government regulation of the market, health care, immigration, same-sex marriage, and taxes. We substitute party identification for party and group thermometers, neither of which were included in the 2012 CCAP. We code this index as missing for true independents and use the other two indices to compute their overall scale score. The indices are averaged and split into deciles to form a reliable scale (Cronbach's $\alpha = 0.81$) of ideological inconsistency.

Demographic variables	
Quarter of interview	**wave**
Region ID	**region**
Age	**birthyr**
Female	**gender**
Race ID	**race**
Education	**educ**
Household income	**pp_faminc**
Marital status	**marstat**
Children	**pp_child18**

Personally LGBT	**pp_closegay4**
Religiosity	**pp_pew_churatd**, **pp_pew_prayer**, **pp_pew_religimp**
Religious ID	**pp_religpew**
Gun ownership	**pp_gunown**
Union member	**pp_labunmemb**
Health plan type	**pp_healthdk_0**
Racial resentment	**pp_raceresent_1**, **pp_raceresent_2**, **pp_raceresent_3**, **pp_raceresent_4**

Political variables	
Political knowledge	**pp_pk_HMajL**, **pp_pk_HMinL**, **pp_pk_SCJ**, **pp_pk_SMajL**, **pp_pk_SMinL**, **pp_pk_Speaker**, **pp_pk_VP**, **pp_pk_house**, **pp_pk_ideo**, **pp_pk_senate**
Political interest	**pp_polinterest**, **pp_newsint2**
Party ID	**pp_pid7**
Ideological ID	**pp_ideo5**
Ideological inconsistency	**sd.issues**, **party.consistent**, **libcon.distance**
Economic issue score	**pp_govt_reg**, **pp_healthreformbill**, **pp_taxwealth**, **pp_univhealthcov**
Social issue score	**pp_abortidentity**, **pp_abortview3**, **pp_gaymar2**
Immigration issue score	**pp_immi_contribution**, **pp_immi_makedifficult**, **pp_immi_naturalize**
Environment issue score	**pp_envpoll2**, **pp_envser2**, **pp_envwarm**
Death penalty issue score	**pp_deathpenalty**, **pp_deathpenfreq**
Affirmative action issue score	**pp_affirmact_gen**
Trade issue score	**pp_tradepolicy**
Egotropic economic evaluations	**pp_persfinretro**
Sociotropic economic evaluations	**pp_econtrend**
Right/wrong track assessment	**pp_track**

Issue salience variables	
Salience Iraq	**pp_imiss_a**
Salience economy	**pp_imiss_b**
Salience immigration	**pp_imiss_c**
Salience environment	**pp_imiss_d**
Salience terrorism	**pp_imiss_f**

Salience gay rights	**pp_imiss_g**
Salience education	**pp_imiss_h**
Salience health care	**pp_imiss_j**
Salience Social Security	**pp_imiss_m**
Salience deficit	**pp_imiss_p**
Salience Afghanistan	**pp_imiss_q**
Salience taxes	**pp_imiss_r**
Salience Medicare	**pp_imiss_s**
Salience abortion	**pp_imiss_t**

The Democracy Fund's VOTER Study Group (Democracy Fund Voter Study Group, 2020) is a long-term panel that includes annual surveys between 2016 and 2019. 4,715 respondents from the 2012 CCAP are also panelists in the VOTER Study Group, allowing us to analyze their political attitudes and choices in subsequent election cycles.

2019 VOTER Survey	
2020 presidential vote intention	**vote2020_2019**
Feeling thermometer (Democratic Party)	**Democrats_2019**
Feeling thermometer (Republican Party)	**Republicans_2019**

2016 VOTER Survey	
2016 presidential vote	**presvote16post_2016**

B Component methods and ensemble weights

The learning ensemble includes eight component methods: an additive probit regression model and the seven methods described in Table A.1. The estimated parameters of the probit regression model are presented in Table A.2. We compare the performance of the learning ensemble against a separate spline-based Generalized Additive Model (GAM), estimated and tuned using the **mgcv** and **caret** packages in **R**.[1]

Table A.1: Component methods included in the learning ensemble.

Method	Short Description and Citations
k-Nearest Neighbor Classifier	Classifies observations based on the majority response class of its closest k neighbors using Euclidian distance in the feature space (Hastie, Tibshirani and Friedman, 2009, pp. 463-468). We use the **knn** implementation in base **R** to estimate a k-nearest neighbor classifier.
Group-lasso interaction network	Uses the ℓ_1 (lasso) penalty to perform regularization and identify meaningful pairwise interactions between predictor variables (Lim and Hastie, 2015). We use the **glinternet** implementation in **R** to estimate a group-lasso interaction network (Lim and Hastie, 2018).
Support Vector Machine (SVM) with a radial basis function kernel	Uses a kernel function (often a nonlinear function such as a polynomial or radial basis function) to measure interobservation similarity and define a flexible classification boundary between response classes. The separating boundry seeks to maximize the distance between itself and some subset of observations—the support vectors—closest to it. A tuning parameter (*Cost*) controls model complexity and the proportion of observations used as support vectors (Vapnik, 2000). We use the **svmRadial** implementation in **R** to estimate a support vector machines (with a radial basis function kernel) (Karatzoglou et al., 2004).

Continued on next page

[1] We use ten-fold cross-validation of AUC-ROC values to determine the optimal parameters of the feature selection parameter **select** (FALSE) and the smoothing method parameter **method** (GCV.Cp).

TABLE A.1 – continued from previous page

Method	Short Description and Citations
Neural Network (model averaged feedforward with single hidden layer)	Models the relationship between the predictor and response variables with a layer of hidden units. The hidden units are weighted sums of some or all of the predictor variables, typically transformed with a nonlinear function such as the sigmoid. The output values from the hidden units are then combined to form the model predictions (Bishop, 1995). We use the **avNNet** implementation in **R** to estimate model averaged neural networks (Venables and Ripley, 2002).
Random Forests	Fits a series of decision trees (splitting based on gini impurity) to bootstrapped samples of the original data, whose predictions are then averaged. To reduce correlation between the trees, random forests only consider a random subset m of the P predictor variables before each split in the tree-growing process (Breiman, 2001). We use the **ranger** implementation in **R** to estimate random forests (Wright and Ziegler, 2017).
Extremely Randomized Trees	Extends the random forest approach by randomizing both the subset of variables and the cut-points along the variables when determining the candidate splits at each stage of the tree-growing process (Geurts, Ernst and Wehenkel, 2006). We use the **ranger** implementation in **R** to estimate extremely randomized trees (Wright and Ziegler, 2017).
Stochastic Gradient Boosting	Like random forests, grows and aggregates a series of B base decision trees ($\hat{f}^1, \hat{f}^2, \ldots, \hat{f}^B$). However, boosting sequentially fits decision trees to the model residuals, giving greater weight to previously misclassified observations. At each iteration, the procedure updates its predictions with those from the new tree using a shrinkage parameter (λ) that controls the learning rate: $\hat{f}^b(X) \leftarrow \hat{f}^{b-1}(X) + \lambda \hat{f}^b(X)$, where $0 < \lambda < 1$. In stochastic boosting, only a random proportion (usually 0.5) of the observations are fit at each iteration to reduce variance and improve computational efficiency (Freund and Schapire, 1997; Friedman, 2001). We use the **xgbTree** implementation in **R** to perform boosting (Chen and Guestrin, 2016).

Table A.2: Probit regression model of swing voting from the 2012 Cooperative Campaign Analysis Project

	Swing voter [0,1]
Quarter	−0.08 (0.02)
Region (Northeast)	0.14 (0.06)
Region (South)	0.03 (0.05)
Region (West)	0.05 (0.06)
Age	−0.10 (0.03)
Female	−0.07 (0.02)
Race (Hispanic/Latino)	0.24 (0.10)
Race (Other)	0.24 (0.11)
Race (White)	0.20 (0.08)
Education	0.00 (0.02)
Household income	−0.13 (0.03)
Married	0.04 (0.02)
Children	-0.01 (0.02)
LGBT	−0.06 (0.03)
Religiosity	0.02 (0.03)
Religion (Evangelical Protestant)	−0.02 (0.06)
Religion (Jewish)	−0.12 (0.15)
Religion (Mainline Protestant)	0.05 (0.06)
Religion (Mormon)	−0.07 (0.15)
Religion (Other)	−0.08 (0.09)
Religion (Secular)	−0.05 (0.07)
Own gun	0.00 (0.02)
Union member	0.03 (0.02)
Has health insurance	−0.01 (0.02)
Political knowledge	−0.21 (0.03)
Political interest	−0.01 (0.02)
Partisanship (Weak Democrat)	0.67 (0.07)
Partisanship (Lean Democrat)	0.63 (0.09)
Partisanship (Independent)	0.94 (0.08)
Partisanship (Lean Republican)	0.66 (0.10)
Partisanship (Weak Republican)	0.76 (0.09)
Partisanship (Strong Republican)	0.23 (0.10)
Partisanship (DK/Not sure)	0.61 (0.19)
Ideological identification (Somewhat liberal)	0.03 (0.13)
Ideological identification (Moderate)	0.25 (0.12)
Ideological identification (Somewhat conservative)	0.05 (0.13)
Ideological identification (Very conservative)	−0.06 (0.15)
Ideological identification (DK/Not sure)	0.37 (0.15)

Continued on next page

TABLE A.2 – continued from previous page

	Swing voter [0,1]
Ideological inconsistency	0.34 (0.03)
Economic preferences	0.03 (0.03)
Social preferences	0.05 (0.03)
Immigration preferences	0.02 (0.03)
Environmental preferences	−0.04 (0.03)
Death penalty preferences	−0.06 (0.02)
Affirmative action preferences	0.02 (0.03)
Trade preferences	0.05 (0.02)
Racial resentment	−0.01 (0.03)
Sociotropic economic evaluations	0.09 (0.03)
Egotropic economic evaluations	−0.07 (0.02)
Right/wrong track evaluations	0.14 (0.03)
Salience Iraq	−0.03 (0.03)
Salience economy	0.05 (0.02)
Salience immigration	−0.01 (0.03)
Salience environment	0.03 (0.03)
Salience terrorism	−0.01 (0.03)
Salience gay rights	0.01 (0.03)
Salience education	0.05 (0.02)
Salience health care	−0.06 (0.02)
Salience Social Security	0.03 (0.03)
Salience deficit	0.04 (0.03)
Salience Afghanistan	0.05 (0.03)
Salience taxes	−0.02 (0.02)
Salience Medicare	−0.05 (0.03)
Salience abortion	−0.01 (0.02)
Constant	−2.20 (0.15)
Observations	8,395
Log Likelihood	−2,564.19

Observations from training set. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

The learning ensemble is constructed by combining weighted predictions from each of the eight component methods following the approach developed in Grimmer, Messing and Westwood (2017). Formally, let i index the observations ($i = 1, \ldots, n$) and m index the component methods ($m = 1, \ldots, M$). We first randomly divide the N observations into D ($d = 1, \ldots, D$) folds, and generate predictions for observations in each fold d by fitting the M methods on observations in the remaining $D - 1$ folds. This produces an $N \times M$ matrix of out-of-sample predictions ($\hat{Y}_{im}$)

for each observation across the component methods. We then estimate the component weights (w_m) by fitting the constrained regression problem:

$$Y_i = \sum_{m=1}^{M} w_m \hat{Y}_{im} + \epsilon_i \tag{1}$$

where Y_i is the observed response by respondent i and ϵ_i is a stochastic error term, with the constraints that $\sum_{m=1}^{M} w_m = 1$ and $w_m \geq 0$. Finally, we fit each of the component methods to the complete dataset and weight the predictions using the M-length vector of estimated component weights $\hat{\mathbf{w}}$.[2] The weighted combination of predictions comprise the ensemble estimates: $w = 0.32$ for extremely randomized trees, $w = 0.27$ for the boosted trees, $w = 0.23$ for the group-lasso interaction network, and $w = 0.17$ for the support vector machine classifier.

[2]We add a standard down-sampling step when estimating the ensemble weights to address class imbalance in the swing voter measure (Kuhn and Johnson, 2013, pp. 427-429). Otherwise, the $\hat{\mathbf{w}}$ would overweight methods that correctly predict the 87% of observations who are not classified as swing voters.

C Model results using alternate operationalizations of swing voters

In addition to the operationalization of the swing voter measure presented in the main text, we train three other learning ensembles that replace the response variable with the following measures (all of the predictor variables remain identical):

1. Floating voters: respondents who switched between major parties in their final 2008 and 2012 presidential vote choices (e.g., Key, 1966; Smidt, 2017).
2. Undecided voters: respondents who reported being undecided in the mid-election survey (e.g., Weghorst and Lindberg, 2013).
3. Swing voters (Shaw's (2008) operationalization): respondents who either (1) do not cast consistent partisan votes or (2) abstain at any point over three election cycles (e.g., Shaw, 2008). We use 2008, 2012 (mid-election survey), and 2012 (post-election survey) as our three time periods.

Table A.3 presents the Pearson correlations between the four sets of model predictions on the validation set. All of the correlations are at least 0.8 (and all of those involving the original measure presented in the main text are at least 0.9), suggesting that all four learning ensembles are tapping into a similar swing voter disposition.

Table A.3: Correlations between Model Predictions with Different Operationalizations of the Response Variable

	Original	Floating	Undecided	Shaw's Operationalization
Original	1.00			
Floating	0.90	1.00		
Undecided	0.91	0.80	1.00	
Shaw's Operationalization	0.94	0.86	0.86	1.00

D Model performance metrics for Figure 2

Table A.4 shows the four principal fit metrics obtained when substituting external indicators of swing voting as the response variable, using observations in the out-of-sample validation set (as in Figure 2 in the main text).

Table A.4: Ensemble and GAM fit metrics for external indicators of swing voting using out-of-sample (validation) data.

	AUC-ROC		BSS	
	Ensemble	GAM	Ensemble	GAM
Split-ticket voter (actual)	0.78	0.73	0.09	0.05
Split-ticket voter (self-ID)	0.70	0.69	0.11	0.11
Ambivalent (pre-election)	0.66	0.65	0.07	0.06
Ambivalent (mid-election)	0.75	0.72	0.14	0.09
Switched House vote (2010-2012)	0.79	0.76	0.10	0.07
Switched House vote (2012)	0.76	0.72	0.08	0.05
Undecided about House vote	0.77	0.77	0.14	0.14
GOP support score IQR range (2012)	0.87	0.78	0.41	0.24
Switched presidential vote (2012-2016)	0.76	0.73	0.10	0.06
Switched presidential vote (2016-2020)	0.75	0.69	0.07	0.03
GOP support score IQR range (2020)	0.87	0.82	0.42	0.30
Party feeling thermometer difference ≤ 15 (2019)	0.75	0.73	0.09	0.06

	CE loss		MCC	
	Ensemble	GAM	Ensemble	GAM
Split-ticket voter (actual)	0.25	0.27	0.04	0.00
Split-ticket voter (self-ID)	0.62	0.63	0.28	0.26
Ambivalent (pre-election)	0.61	0.62	0.19	0.14
Ambivalent (mid-election)	0.43	0.45	0.24	0.09
Switched House vote (2010-2012)	0.24	0.26	0.06	0.04
Switched House vote (2012)	0.24	0.25	-0.01	0.00
Undecided about House vote	0.40	0.40	0.19	0.17
GOP support score IQR range (2012)	0.47	0.57	0.59	0.44
Switched presidential vote (2012-2016)	0.29	0.30	0.11	0.00
Switched presidential vote (2016-2020)	0.27	0.29	0.00	0.00
GOP support score IQR range (2020)	0.46	0.53	0.59	0.48
Party feeling thermometer difference ≤ 15 (2019)	0.28	0.29	0.00	0.00

Note: Predictions calibrated using Platt scaling (Platt, 2000).

Figure A.1 reproduces Figure 2 in the main text, but instead shows the difference in proportions between the ensemble and (baseline) GAM prediction deciles for each external indicator. Hence, positive (negative) values indicate higher proportions among observations in the corresponding decile when binning the ensemble (GAM) predictions.

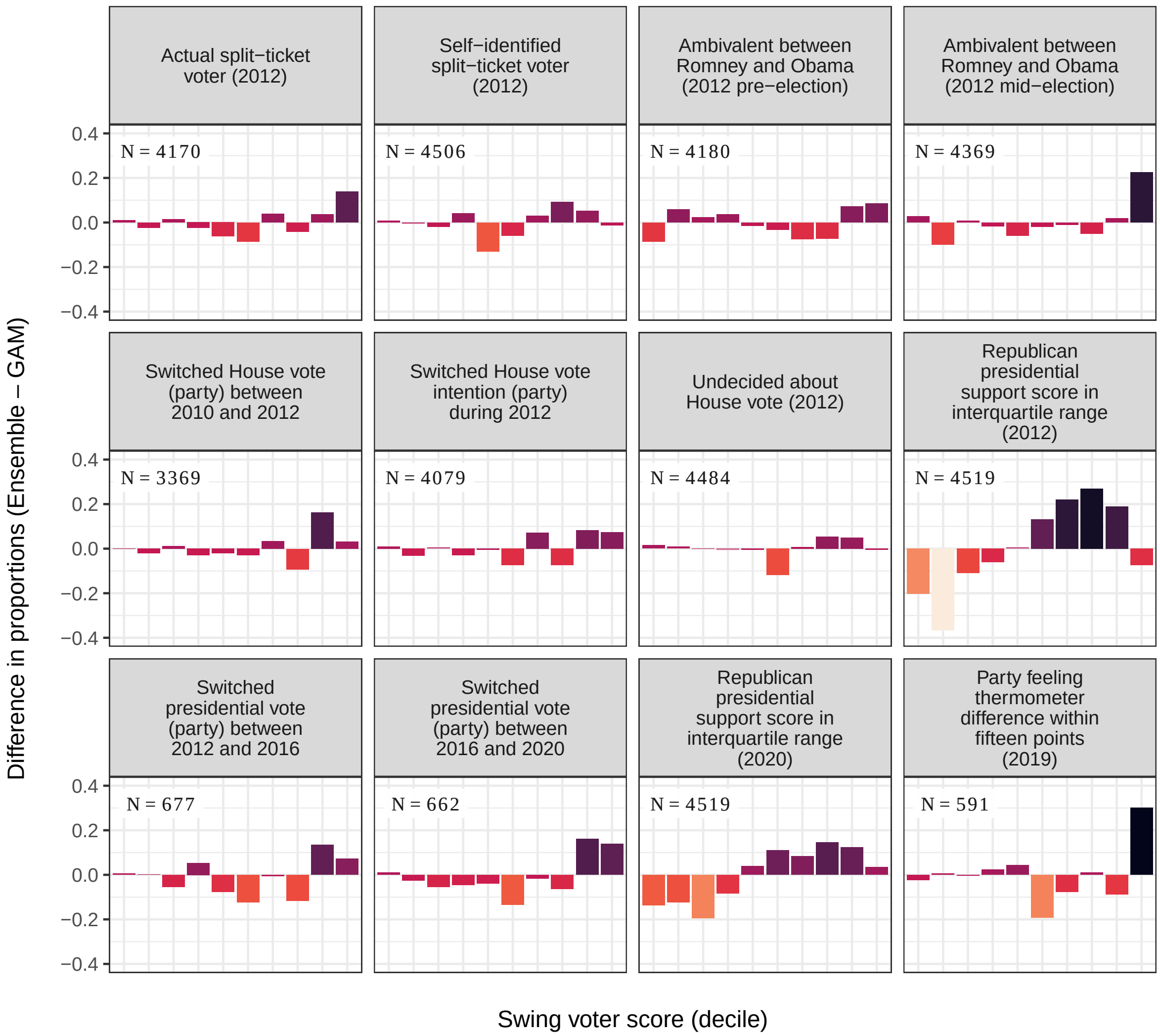

Figure A.1: Difference in predictive performance of ensemble and (baseline) GAM scores on additional indicators of swing voter propensity using out-of-sample (validation) data. Ambivalence defined as placing the candidates within one point of each other on a five-point favorability scale. Republican presidential support scores are calculated by estimating separate ensemble models of 2012 presidential vote choice and 2020 presidential vote intention.

E Republican presidential support scores from the 2012 Cooperative Campaign Analysis Project and the 2019 wave of the VOTER Study Group

The Republican presidential support scores presented in the main text (specifically Figure 2) measure the propensity of the 2012 CCAP respondents to support Mitt Romney over Barack Obama in the 2012 presidential election; and of the 2019 VOTER Study Group respondents to support Donald Trump over the Democratic nominee in the 2020 presidential election. We estimate these scores using the same learning ensemble method used to predict swing voters, but replacing the response variable with 2012 presidential vote choice/2020 presidential vote intention.

For the 2012 Republican presidential support model, the learning ensemble selects four component methods (with weights in parentheses): gradient boosted trees ($w = 0.48$), extremely randomized trees ($w = 0.21$), the SVM classifier ($w = 0.26$), and the random forest ($w = 0.05$). The ensemble achieves a correct classification rate of 97.1$\%$ and an AUC fit statistic of 0.996.

For the 2020 Republican presidential support model, the learning ensemble selects six component methods (with weights in parentheses): the model averaged neural network ($w = 0.40$), group-lasso interaction network ($w = 0.21$), the random forest ($w = 0.15$), the additive probit regression model ($w = 0.14$), the SVM classifier ($w = 0.07$), and k-nearest neighbor classifier ($w = 0.03$). The ensemble achieves a correct classification rate of 94.1$\%$ and an AUC fit statistic of 0.987.

Based on this performance, we believe the predicted probabilities from the model serve as valid measure of respondent propensities to support the Republican presidential candidate in the 2012 and 2020 elections.

F Permutation tests of feature importance

Below we further detail the (unconditional) permutation approach for estimating the feature importance values presented in the main text. Following Breiman (2001), we calculate the importance of each predictor variable X_j by randomly shuffling its values and generating predictions from the learning ensemble while leaving the remainder of the dataset (X_{-j}) unchanged. The difference in predictive accuracy as measured by four fit statistics (AUC-ROC, Brier skill score, cross-entropy loss, and the Matthews correlation coefficient; as recommended by Cook 2007) between the two sets of predictions—those from the original (unpermuted) and permuted datasets—serve as our measure of importance for feature X_j. The feature importance values Ψ_j presented in the main text and below are the mean difference in fit metric when generating predictions from X_j and X_{-j} across 500 permutation trials.

The results shown in the main text (specifically, Figure 3) truncate the number of variables for space purposes. The results for all predictor variables are shown in Figures A.2–A.5. Note that these results are obtained using an *unconditional* permutation scheme. To address the problem of spurious relationships among correlated predictors, Strobl et al. (2008); Debeer and Strobl (2020) develop a conditional permutation method for random forests in which each feature is randomly shuffled conditionally on values of the correlated predictor variable(s).[3] That is, permutation occurs within bins defined by unique values of the correlated variable(s).

Below we develop a generalized version of their conditional permutation scheme and apply it to our learning ensemble. First, for each feature X_j, we identify correlated variables using the maximal information coefficient (MIC) (Reshef et al., 2011; Albanese et al., 2018). The MIC is a nonparametric statistic that captures a wider range of linear and nonlinear relationships between variables than the traditional Pearson's correlation coefficient.[4]

Second, in cases where a feature X_j is correlated with multiple variables, we use cluster analysis

[3]See Chan and Ratkovic (2020) for another exposition of the problem posed by feature dependencies when estimating variable importance.

[4]MIC values are bounded between 0 and 1, with larger values indicating stronger associations between variables. We determine correlated predictors for each feature X_j as those with MIC values greater than 0.1. We note that 69 of the 990 pairwise relationships between the 45 predictor variables meet this criterion.

to define the conditional bins within which values of X_j will be randomly shuffled. Specifically, we apply the k-medoids (or partitioning around medoids [PAM]) clustering algorithm to the correlated predictors and use average silhouette width to determine the appropriate number of clusters (Rousseeuw, 1987; Kaufman and Rousseeuw, 1990). We then randomly permute values of X_j conditional on the cluster assignments (i.e., among observations in each cluster separately). For instance, political knowledge is correlated with political interest and household income. Cluster analysis identifies groupings of observations with similar covariate profiles on these two features. Then, to assess the effect of political knowledge, we randomly permute its values among observations in each of these clusters and calculate the feature importance measure Ψ_j as before. This serves to preserve correlational structure in the original data and provide a more partial estimate of variable importance (Debeer and Strobl, 2020).

The feature importance values from the conditional permutation tests are shown in Figures A.6–A.9. Overall, the conditional and unconditional Ψ_j estimates are correlated at $r = 0.99$ (using the AUC-ROC metric), $r = 0.96$ (Brier skill score), $r = 0.99$ (cross-entropy loss), and $r = 0.95$ (the Matthews correlation coefficient). Among partisan groups, these correlations are 0.97, 0.97, 0.97, and 0.98 among Democrats (for AUC-ROC, Brier skill score, cross-entropy loss, and the Matthews correlation coefficient, respectively). For Republicans, the correlations are 0.99, 0.97, 0.99, and 0.98. For Independents/unaffiliated, the correlations are 0.97, 0.94, 0.91, and 0.94.

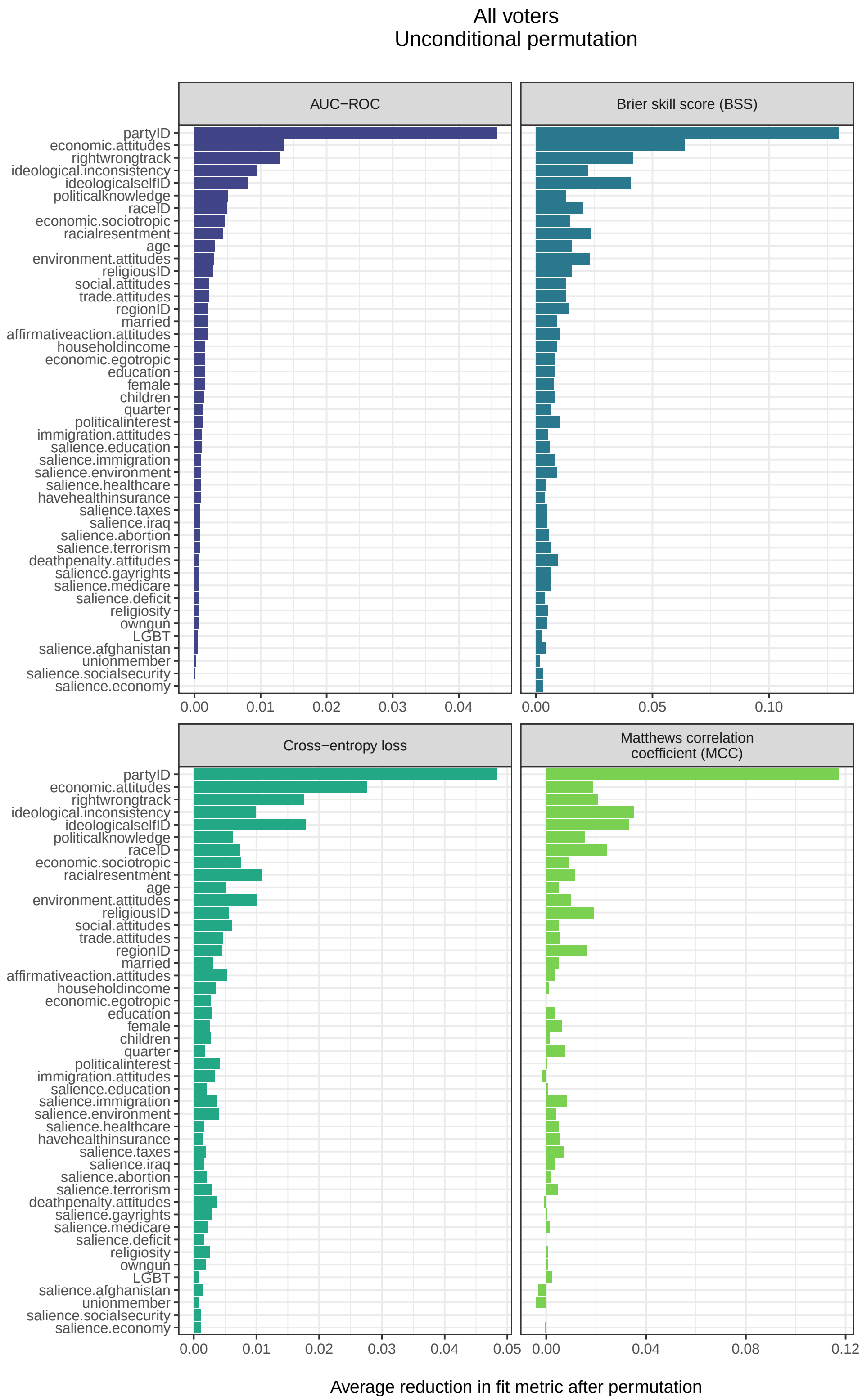

Figure A.2: Feature importance estimates from the learning ensemble using unconditional permutation tests (all respondents).

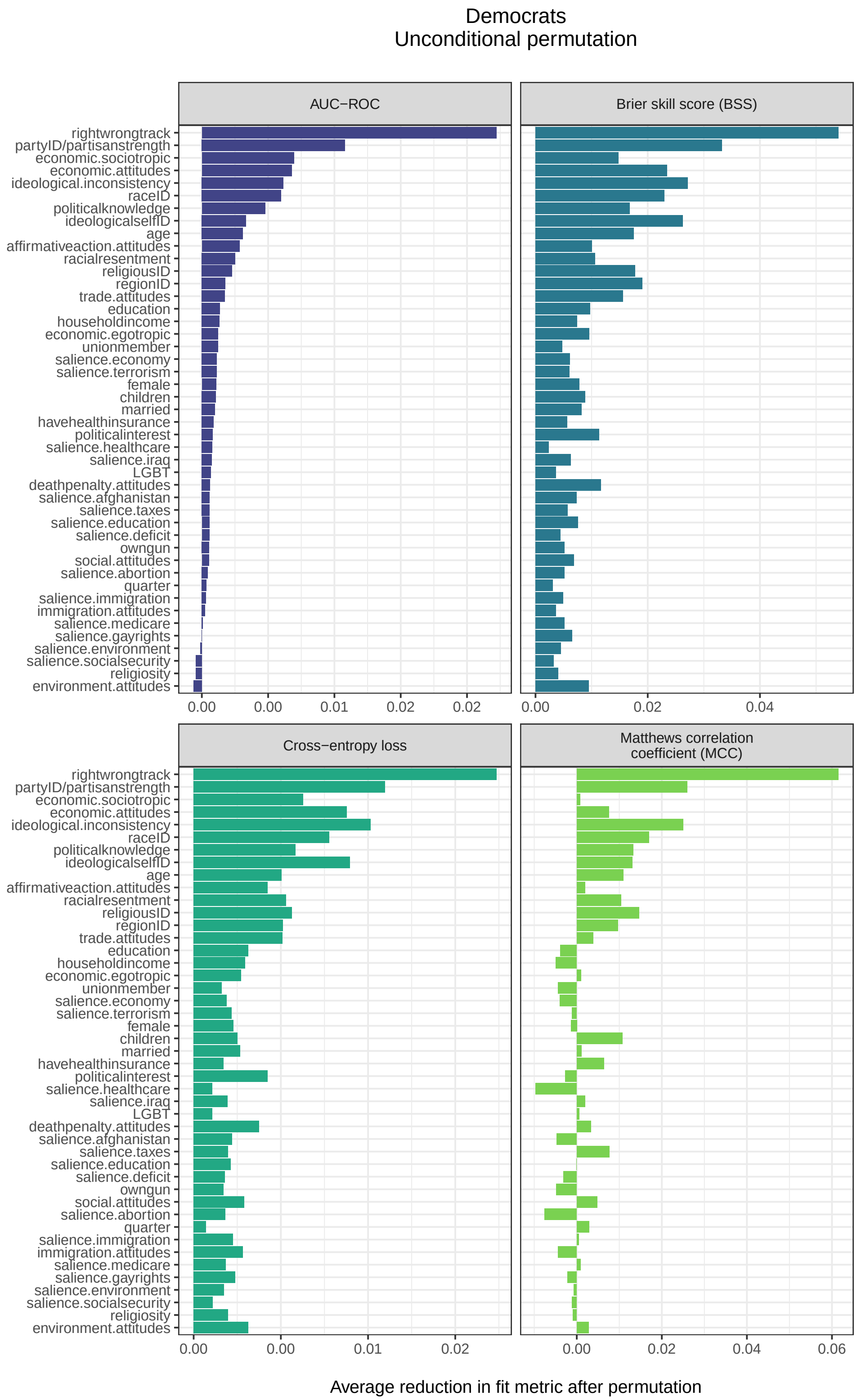

Figure A.3: Feature importance estimates from the learning ensemble using unconditional permutation tests (Democrats).

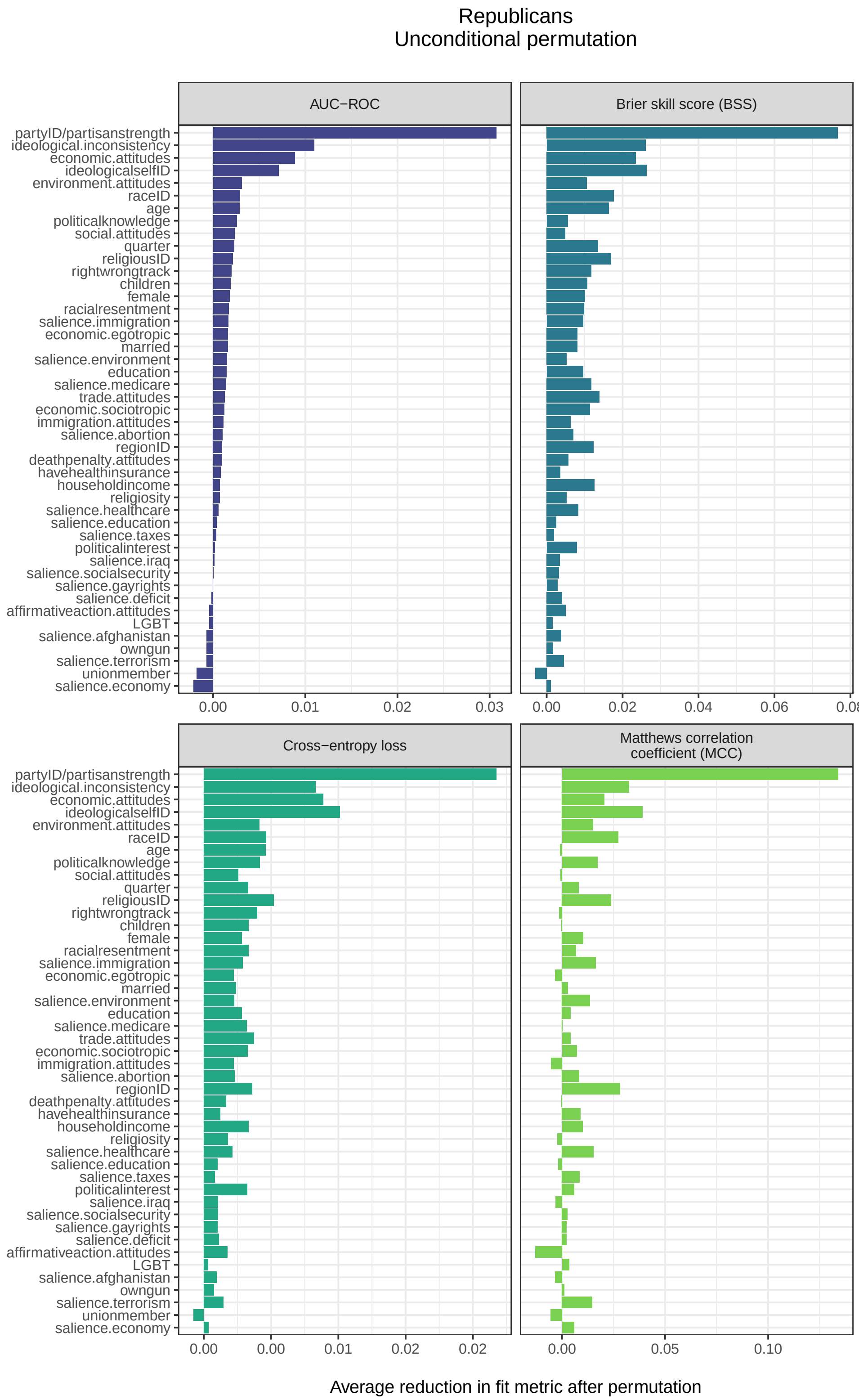

Figure A.4: Feature importance estimates from the learning ensemble using unconditional permutation tests (Republicans).

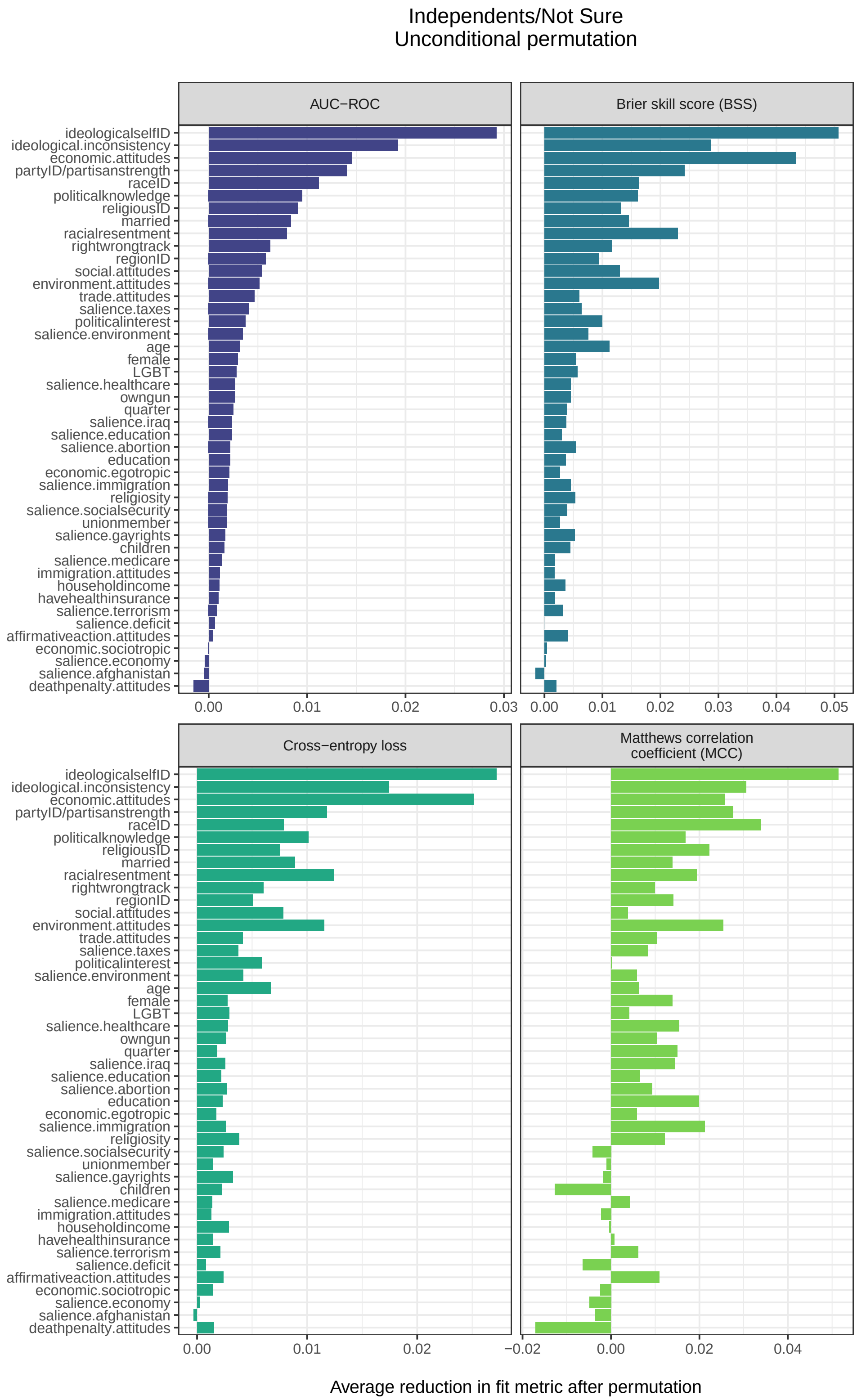

Figure A.5: Feature importance estimates from the learning ensemble using unconditional permutation tests (Independents/Not Sure).

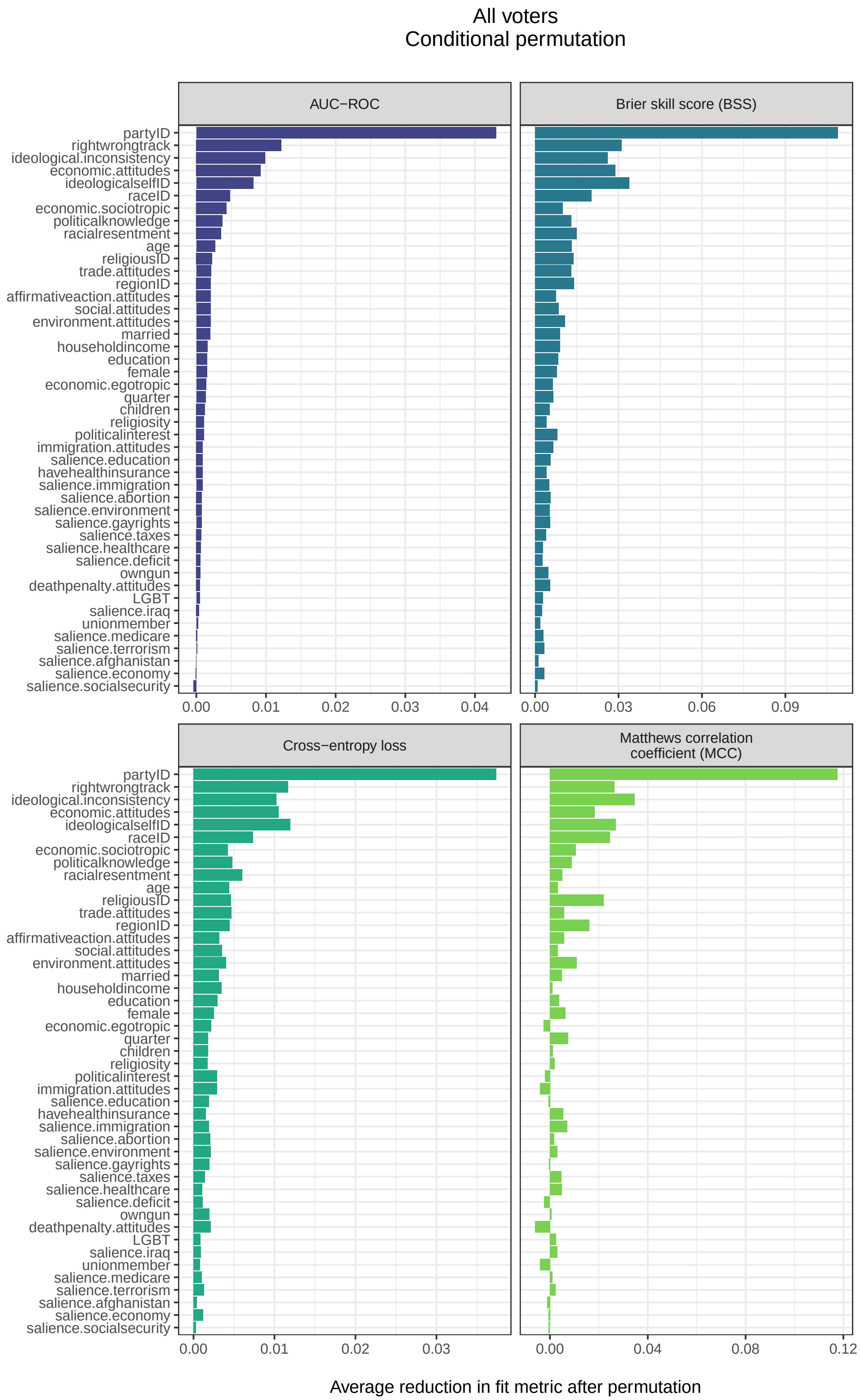

Figure A.6: Feature importance estimates from the learning ensemble using conditional permutation tests (all respondents).

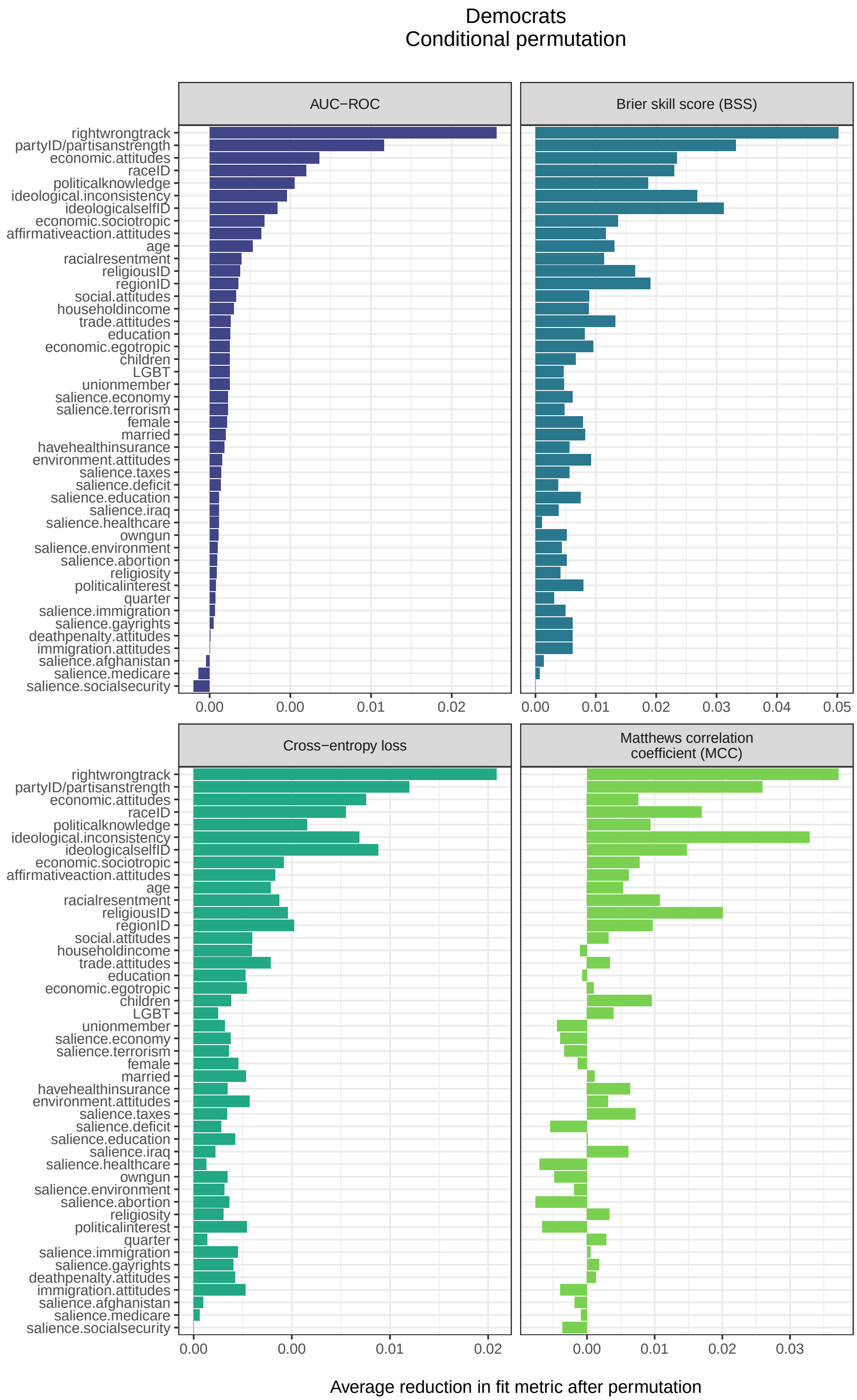

Figure A.7: Feature importance estimates from the learning ensemble using conditional permutation tests (Democrats).

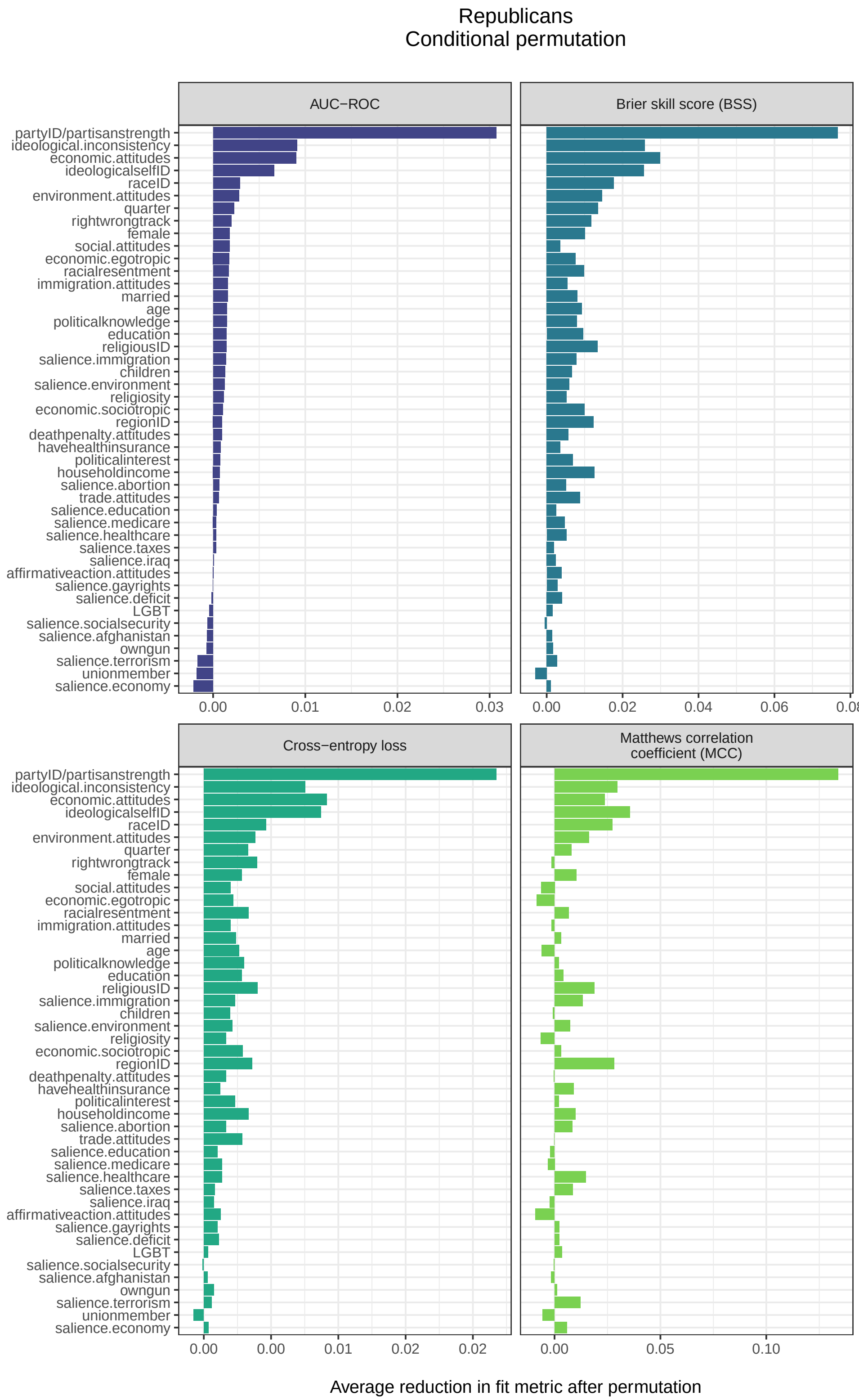

Figure A.8: Feature importance estimates from the learning ensemble using conditional permutation tests (Republicans).

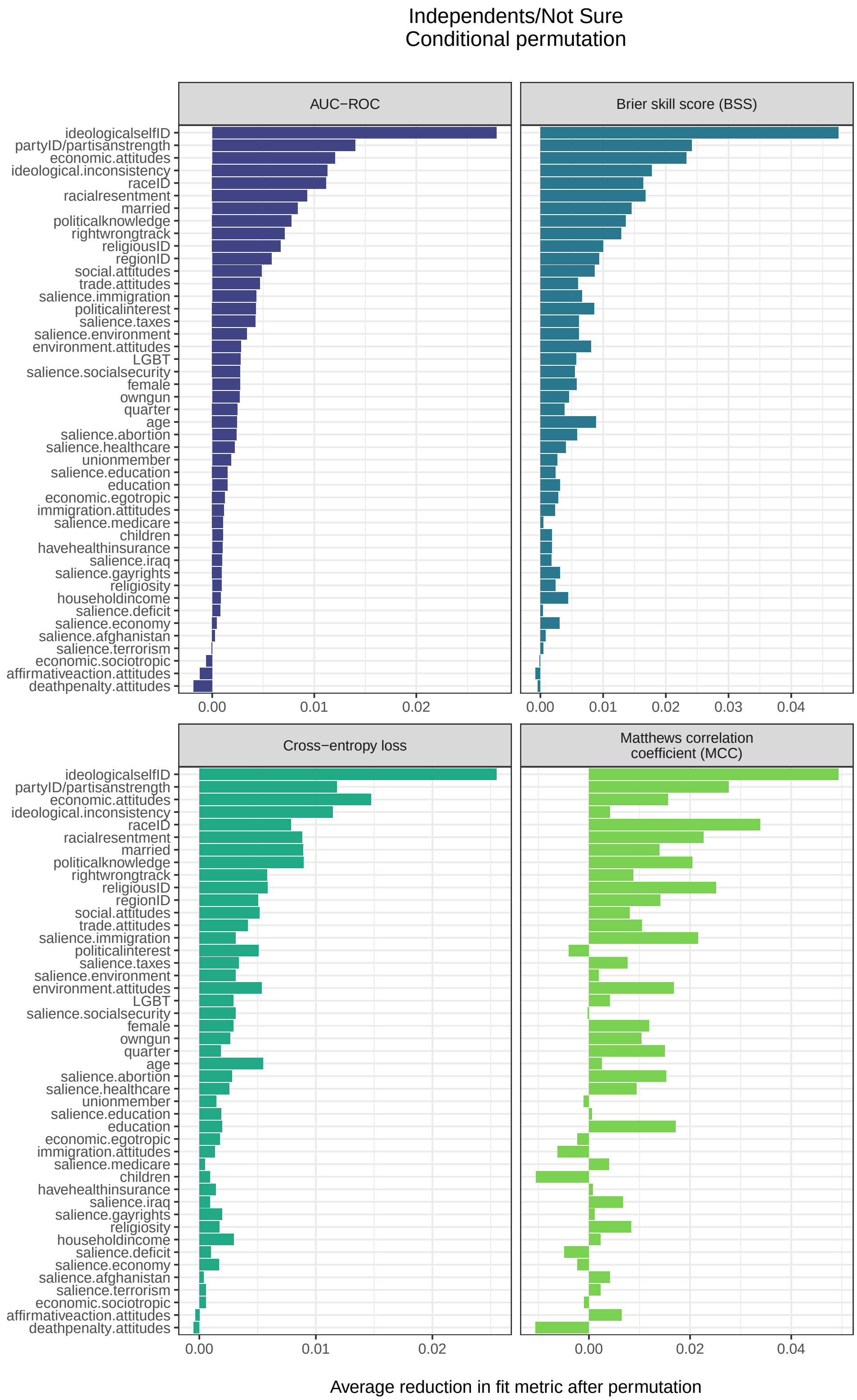

Figure A.9: Feature importance estimates from the learning ensemble using conditional permutation tests (Independents/Not Sure).

G Feature interactions in the learning ensemble

To identify which predictor variables have the largest interactive effects in the learning ensemble, below we use two techniques developed in Friedman and Popescu (2008) using observations from the expanded test set.

The first approach—known as Friedman's H-statistic—works by decomposing the total influence of each variable into its direct and indirect (i.e., conditional) components via the partial dependence function. The partial dependence of a prediction function $f(x)$ on one or more predictor variables x_S provides an estimate of their effect by marginalizing over the empirical distribution of complementary predictor variables x_{-S}. That is,

$$\begin{aligned} f_S(x_S) &= E_{x_{-S}}\left[f(x_S, x_{-S})\right] \\ &= \int f(x_S, x_{-S}) d\mathbb{P}(x_{-S}) \end{aligned} \tag{2}$$

The partial dependence function $f_S(x_S)$ in Equation 2 is usually approximated by using the Monte Carlo method to sample over observations i in $1, \ldots, N$:

$$\hat{f}_S(x_S) = \frac{1}{N} \sum_{i=1}^{N} f(x_S, x_{-S}) \tag{3}$$

For a given predictor variable x_j, Friedman's H-statistic (H_j^2) calculates the difference between (1) the observed prediction function; and (2) an additive version of the prediction function that assumes the effects of x_j and x_{-j} are entirely independent of each other and hence can be expressed as the sum of their respective partial dependence functions:

$$f(x) = f_j(x_j) + f_{-j}(x_{-j}) \tag{4}$$

Friedman's H-statistic expresses this difference as a proportion of the total variance in the prediction function, again using Monte Carlo simulation:

$$H_j^2 = \frac{\sum_{i=1}^{N} \left[f(x^{(i)}) - \hat{f}_j(x_j^{(i)}) - \hat{f}_{-j}(x_-j^{(i)}) \right]^2}{\sum_{i=1}^{N} f^2(x^{(i)})} \tag{5}$$

We estimate Friedman's H-statistic for each of the 45 predictor variables in the learning ensemble using the **iml** package in **R** (Molnar, Bischl and Casalicchio, 2018) on observations in the out-of-sample validation set ($N = 4,519$). The results, shown in Figure A.10, indicate that a large percentage of the total effects of party identification (34$\%$), ideological inconsistency (20$\%$), economic issue attitudes (17$\%$), and right/wrong track evaluations (16$\%$) are interactive in nature.

Table A.5 lists the two-way interactions with the largest Friedman's H-statistic values for all voters and partisan groups separately. One conditional relationship that looks especially promising is between ideological inconsistency and right/wrong track evaluations among Democrats. Running an additive linear regression model (with an interaction term between these two features) for out-of-sample Democratic voters yields a significant coefficient, with the results shown in Figure A.11.

The second approach we use to identify meaningful interactive relationships is to fit prediction rule ensembles (also from Friedman and Popescu [2008]) to our estimated swing voter propensity scores. Prediction rule ensembles (also referred to as the RuleFit algorithm) first grow 500 boosted decision trees fit to an outcome (i.e., the swing voter predictions) and generate a large number of interpretable IF/THEN rules based on the conditional paths from the root nodes of each tree. A lasso penalty (determined by cross-validation) is then used to reduce the number of rules to prevent overfitting and estimate coefficients for the remaining rules.

We estimate separate prediction rule ensembles for all voters and partisan groups in the validation set using default values from the **pre** package in **R** (Fokkema, 2020). The most important rules identified are shown in Table A.6. The rules provide several promising candidates for future research, including the role of environmental attitudes among Republicans and the influence of religiosity among non-Black Democrats.

Table A.5: Pairwise Interaction Strength for Ensemble Swing Voter Predictions

Interaction pair	Friedman's H-statistic
All voters	
Region : Religious identification	0.51
Death penalty issue attitudes : Religious identification	0.36
Female : Region	0.35
Social issue attitudes : Economic issue attitudes	0.34
Salience terrorism : Economic issue attitudes	0.30
Environmental issue attitudes : Economic issue attitudes	0.27
Salience environment : Region	0.26
Married : Household income	0.26
Racial resentment : Economic issue attitudes	0.25
Economic issue attitudes : Right/wrong track evaluations	0.25
Salience immigration : Economic issue attitudes	0.24
Democratic voters	
Political knowledge : Right/wrong track evaluations	0.16
Affirmative action issue attitudes : Ideological identification	0.15
Racial resentment : Right/wrong track evaluations	0.15
Ideological inconsistency : Right/wrong track evaluations	0.13
Economic issue attitudes : Ideological identification	0.12
Republican voters	
Gun ownership : Right/wrong track evaluations	0.23
Immigration issue attitudes : Right/wrong track evaluations	0.21
Quarter : ideological identification	0.15
Salience terrorism : Ideological identification	0.14
Religious identification : Ideological identification	0.13
Independent/DK voters	
Social issue attitudes : Economic issue attitudes	0.34
Racial resentment : Economic issue attitudes	0.30
Salience environment : Ideological identification	0.26
Environmental issue attitudes : Economic issue attitudes	0.25
Salience terrorism : Economic issue attitudes	0.24

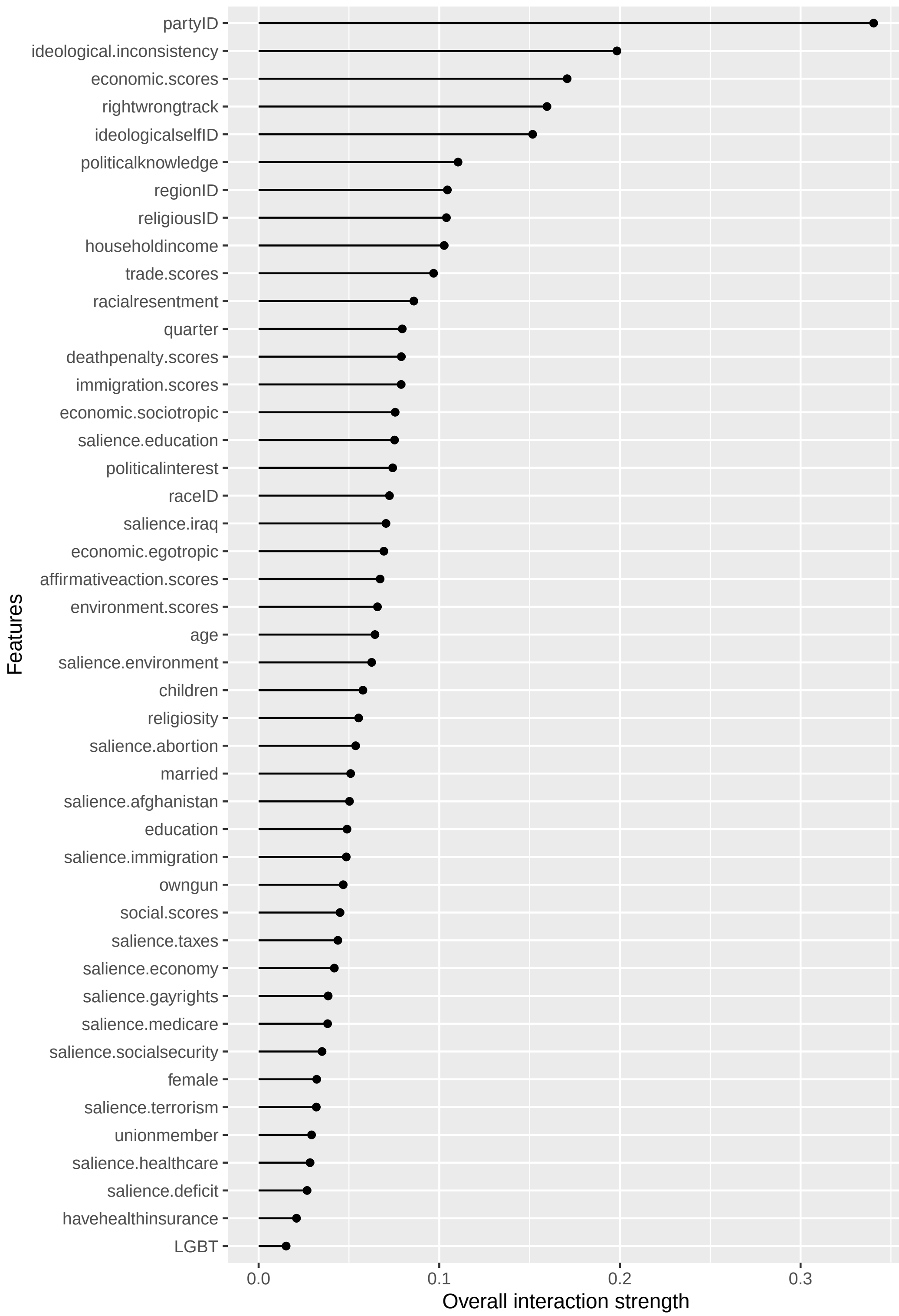

Figure A.10: Friedman's H-statistic Values for Predictor Variables in the Learning Ensemble

Higher values indicate that a larger proportion of the variable's influence is conditional on other variables in the model.

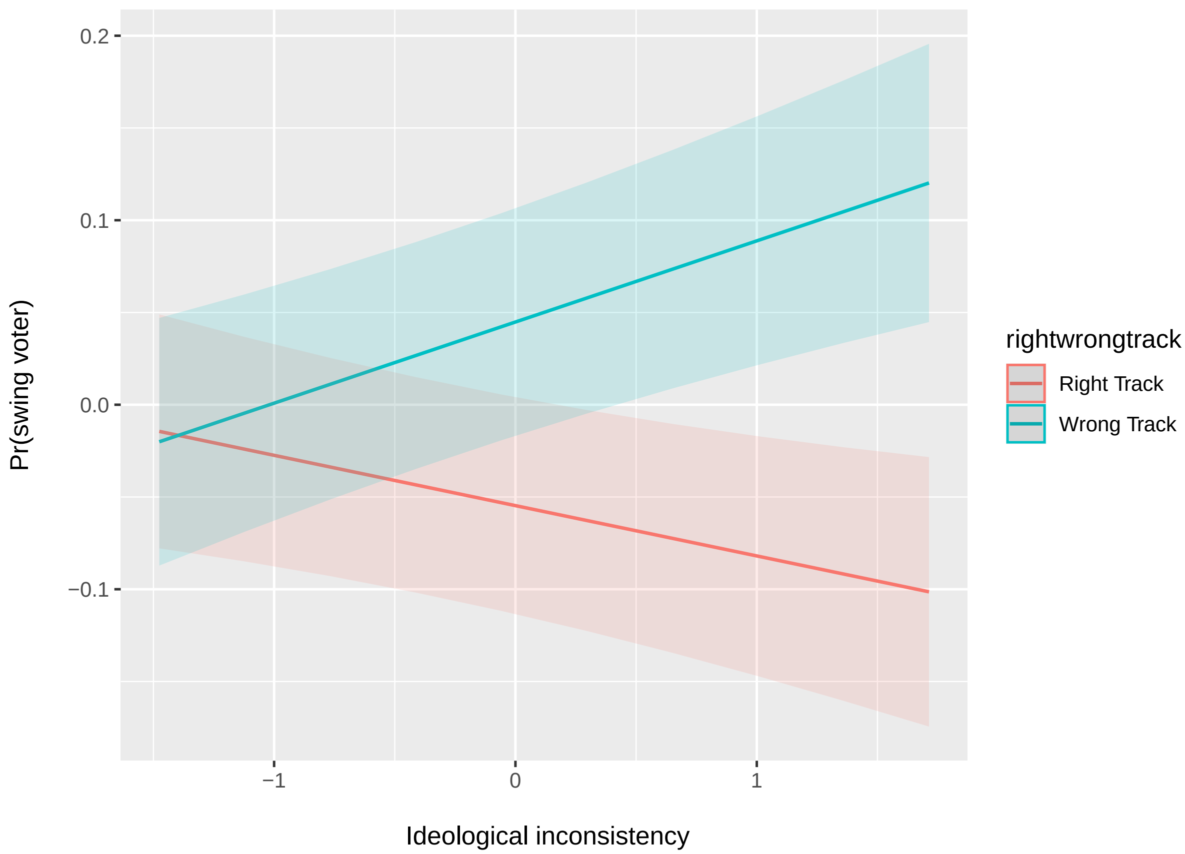

Figure A.11: Swing Voter Predictions by Ideological Inconsistency and Right/Wrong Track Evaluations (Democratic Voters)

Table A.6: Interactions Identified by the Prediction Rule Ensembles

Rule	Coefficient
All voters	
Party identification ∈ {Strong Democrat, Lean Democrat, Lean Republican} & Race ∈ {Black}	-0.47
Ideological inconsistency $>$ -0.77 & Household income $\leq$ \$20k & Right/wrong track evaluations ∈ {Wrong track}	0.41
Right/wrong track evaluations ∈ {Right track, Not sure} & Economic issue attitudes $\leq$ -0.61	-0.39
Political knowledge $\leq$ 0.34 & Right/wrong track evaluations ∈ {Right track, Not sure} & Party identification ∈ {Strong Democrat, Weak Democrat, Lean Democrat}	-0.37
Ideological inconsistency $\leq$ -0.06 & Ideological inconsistency $>$ -0.77 & Trade issue attitudes ∈ {Not sure, Oppose increased trade}	0.37
Democratic voters	
Right/wrong track evaluations ∈ {Right track, Not sure} & Age $>$ 34 & Salience environment ∈ {Very important, Somewhat important}	-0.61
Right/wrong track evaluations ∈ {Wrong track} & Political knowledge $\leq$ 0.34 & Salience terrorism ∈ {Very important}	0.49
Right/wrong track evaluations ∈ {Right track, Not sure} & Party identification ∈ {Strong Democrat, Lean Democrat} & Age $>$ 44	-0.35
Economic issue attitudes $\leq$ 0.05 & Sociotropic economic evaluations ∈ {Same, Worse} & Party identification ∈ {Weak Democrat, Lean Democrat}	0.34
Ideological inconsistency $\leq$ -0.06 & Party identification ∈ {Strong Democrat} & Salience taxes ∈ {Not very important, Somewhat important, Very important}	-0.33
Economic issue attitudes $>$ -1.26 & Race ∈ {Hispanic/Latino, Other, White} & Religiosity $>$ -0.03	0.22
Republican voters	
Economic issue attitudes $>$ 0.71 & Ideological identification ∈ {Somewhat liberal, Somewhat conservative, Very conservative} & Salience immigration ∈ {Somewhat important, Very important}	-0.59
Economic issue attitudes $>$ 0.71 & Ideological identification ∈ {Moderate, Somewhat conservative, Very conservative, DK/not sure}	-0.55
Political knowledge $>$ -0.80 & Environmental issue attitudes $>$ -0.54 & Salience immigration ∈ {Somewhat important, Very important}	-0.39
Party identification ∈ {Strong Republican} & Ideological identification ∈ {Moderate, Somewhat conservative, Very conservative, DK/not sure} & Race ∈ {Hispanic/Latino, Other, White}	-0.24
Ideological inconsistency $\leq$ 0.30 & Racial resentment $>$ -0.58	-0.13

Independent/DK voters	
Ideological inconsistency $\leq$ 1.00 & Ideological identification $\in$ {Very liberal, Somewhat liberal, Moderate, Somewhat conservative, Very conservative}	-0.52
Political interest $\in$ {Medium, High} & Trade issue attitudes $\in$ {Not sure, Oppose increased trade}	-0.22
Trade issue attitudes $\in$ {Favor increased trade} & Racial resentment $>$ 0.10	0.14
Political interest $\in$ {High} & Salience healthcare $\in$ {Very important} & Political knowledge $>$ -1.18	-0.09
Ideological inconsistency $\leq$ 1.00 & Trade issue attitudes $\in$ {Not sure, Favor increased trade}	-0.07

Issue attitudes standardized and coded such that higher values indicate more conservative positions.

H Descriptive statistics for swing voter predictions

Tables A.7–A.9 show the mean and variances of the swing voter scores (using survey weights) from the learning ensemble by state, political knowledge, and ideological inconsistency (Federico and Hunt, 2013). Tables A.10–A.11 flip the variables and show mean political knowledge and ideological inconsistency values by swing voter score deciles. Both are based on predictions for all out-of-sample observations ($N = 19,599$).

The results show monotonic or near-monotonic relationships between the swing voter predictions and political knowledge (in a negative direction) and ideological inconsistency (in a positive direction). They also reveal that North Dakota, Connecticut, and New Hampshire are the states whose electorates have the largest mean swing voting propensity.[5] Of course, swing states—those most likely to switch between parties in national or subnational elections—will not necessarily be those with the largest proportion of swing voters, as partisan balance also matters.

Table A.7: Ensemble swing voter predictions by state (sorted by weighted mean score), 2012 Cooperative Campaign Analysis Project.

State	N	Weighted mean SV score	Weighted σ^2
North Dakota	53	0.36	0.05
Connecticut	211	0.35	0.05
New Hampshire	114	0.34	0.04
West Virginia	161	0.33	0.04
Rhode Island	56	0.33	0.04
New Jersey	513	0.32	0.04
Kansas	176	0.32	0.06
Idaho	126	0.31	0.04
Arkansas	191	0.31	0.04
Pennsylvania	956	0.31	0.04
South Dakota	41	0.30	0.05
Ohio	730	0.29	0.05
Missouri	400	0.29	0.04
Indiana	390	0.29	0.04
Oklahoma	215	0.29	0.04

[5] In a 2012 post, Nate Silver found that Rhode Island, New Hampshire, and Maine were the most politically "elastic" states. These states also rank high in our estimates. See Nate Silver, "Swing Voters and Elastic States," 21 May 2012, available from: `https://fivethirtyeight.com/features/swing-voters-and-elastic-states/`

Florida	1574	0.29	0.04
Tennessee	343	0.29	0.04
Oregon	329	0.28	0.05
Hawaii	62	0.28	0.03
New York	1070	0.28	0.04
Michigan	610	0.27	0.04
Massachusetts	350	0.27	0.05
Kentucky	282	0.27	0.04
Wyoming	40	0.27	0.05
Georgia	625	0.26	0.04
Maine	120	0.26	0.04
Montana	54	0.26	0.04
Virginia	479	0.26	0.03
Nevada	265	0.26	0.03
North Carolina	592	0.26	0.04
Colorado	427	0.26	0.03
Wisconsin	377	0.26	0.04
Arizona	513	0.26	0.04
Texas	1361	0.26	0.04
Mississippi	167	0.25	0.04
Minnesota	297	0.25	0.05
Maryland	387	0.25	0.03
Alabama	239	0.24	0.03
Illinois	825	0.24	0.04
California	1986	0.24	0.03
Utah	208	0.24	0.02
Louisiana	183	0.23	0.04
Nebraska	128	0.23	0.04
Iowa	172	0.23	0.04
South Carolina	307	0.22	0.03
Washington	537	0.22	0.03
New Mexico	140	0.22	0.04
Delaware	79	0.21	0.04
Vermont	37	0.19	0.04
Alaska	36	0.17	0.02
District of Columbia	65	0.16	0.02

Table A.8: Ensemble swing voter predictions by political knowledge, 2012 Cooperative Campaign Analysis Project.

Knowledge decile	N	Weighted mean SV score	Weighted σ^2
1: Lowest Knowledge	1264	0.43	0.04
2	1502	0.39	0.03
3	2110	0.36	0.04
4	2249	0.33	0.04
5	1970	0.30	0.04
6	1645	0.27	0.04
7	1387	0.24	0.03
8	1620	0.20	0.03
9	2141	0.16	0.02
10: Highest Knowledge	3711	0.13	0.02

Table A.9: Ensemble swing voter predictions by ideological inconsistency, 2012 Cooperative Campaign Analysis Project.

Ideological inconsistency decile	N	Weighted mean SV score	Weighted σ^2
10: Most Inconsistent	2002	0.43	0.03
9	2069	0.37	0.04
8	1902	0.35	0.04
6	1944	0.33	0.04
7	1997	0.29	0.03
4	1919	0.24	0.04
5	1798	0.20	0.03
3	2119	0.19	0.03
1: Least Inconsistent	1930	0.15	0.02
2	1919	0.13	0.02

Table A.10: Political knowledge by ensemble swing voter predictions, 2012 Cooperative Campaign Analysis Project.

Swing voter score decile	N	Weighted Mean Knowledge	Weighted σ^2
1: Lowest SV Score	1960	8.76	3.65
2	1960	8.12	4.97
3	1960	7.22	6.20
4	1960	6.72	7.68
5	1960	5.74	7.98
6	1960	5.33	6.99
7	1961	4.95	7.38
8	1959	4.63	7.14
9	1960	4.04	5.88
10: Highest SV Score	1959	3.82	5.65

Table A.11: Ideological inconsistency by ensemble swing voter predictions, 2012 Cooperative Campaign Analysis Project.

Swing Voter Score Decile	N	Weighted Mean Inconsistency	Weighted σ^2
10: Highest SV Score	1959	7.33	5.75
9	1960	7.06	5.85
8	1959	6.71	6.96
7	1961	6.19	7.20
6	1960	5.74	7.15
5	1960	5.39	7.37
4	1960	5.00	7.33
3	1960	4.35	6.39
2	1960	3.63	4.35
1: Lowest SV Score	1960	2.87	3.17

I Uncertainty measures for the results

Following Grimmer, Messing and Westwood (2017), we do not present uncertainty estimates for the results in the main text. Developing procedures to estimate measures of parameter uncertainty with known asymptotic properties from supervised machine learning algorithms (particularly regularization, tree-based, and ensemble methods) remains an ongoing area of research (see, e.g., Chatterjee and Lahiri, 2011; Wager, Hastie and Efron, 2014; Coyle and van der Laan, 2018; Das, Gregory and Lahiri, 2019).

Below we present two alternative strategies to assess the stability of the model performance metrics and feature importance estimates. Though we caution against an asymptotic interpretation of these results, we think they are nonetheless useful ways to test the learning ensemble's general sensitivity to error processes.

The first approach is a sensitivity analysis that randomly perturbs the model predictions (by adding random error of varying magnitude from a uniform distribution) and assesses the dropoff in predictive performance using the four principal fit metrics (AUC-ROC, Brier skill score, cross-entropy loss, and the Matthews correlation coefficient). Our use of perturbation tests draws from Altman, Gill and McDonald (2004, chap. 4) and the simulation extrapolation (SIMEX) method of Cook and Stefanski (1994), which uses a Monte Carlo approach to simulate the influence of varying levels of measurement error on coefficient estimates (see also Carroll et al., 1996).

We proceed using the same out-of-sample validation set ($N = 4,519$) from the main text. In each of 500 stimulations, we take a random draw from a uniform distribution with set lower/upper bounds and add it to the ensemble and GAM predictions (truncating both sets of predictions to range between 0 and 1). We vary the lower/upper bounds of the uniform distribution between -0.01/0.01 and -0.21/0.21, and calculate the resulting fit metrics for the learning ensemble and GAM.[6] The results are presented in Figure A.12, and suggest that the ensemble continues to outperform the original GAM predictions across error levels for all but cross-entropy loss, a metric on which

[6]The standard deviation of the ensemble predictions is 0.21 ($\sigma = 0.17$ for the GAM predictions). The mean difference between the two sets of predictions is 0.1.

the ensemble deteriorates more quickly and begins to overlap with the original GAM value when the error bounds reach -0.06/0.06. However, even in this case, the ensemble is less sensitive to random error (i.e., declines more slowly as the magnitude of the error increases) than the GAM.

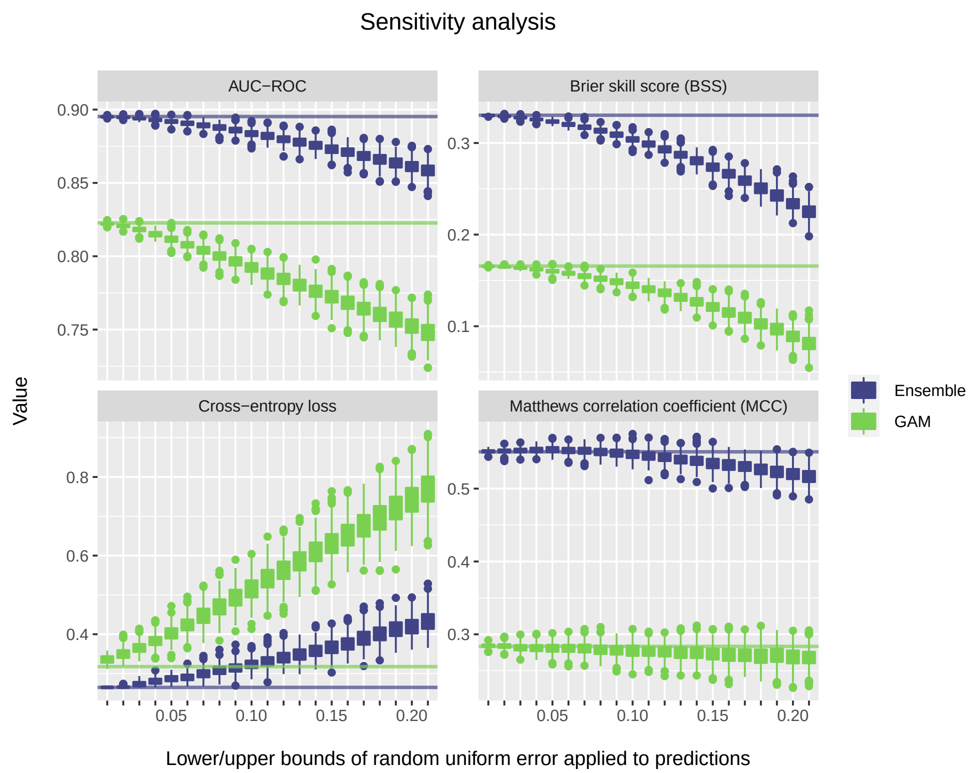

Figure A.12: Sensitivity analysis of model fit metrics using out-of-sample (validation) data. Horizontal lines indicate the original value of the fit metric from the unperturbed predictions. Trials repeated 500 times.

The second approach involves generating uncertainty measures for the feature importance estimates presented in Figure 3 in the main text. Ishwaran and Lu (2019) advocate use of the delete-d jackknife (Shao and Wu, 1989) for estimating the variance of permutation-based variable importance measures. The delete-d jackknife operates like the standard jackknife (or leave-one-out) estimator, except that it operates on random subsamples of size (n-d). We draw 500 subsamples, setting $d = \sqrt{n}$ (Efron and Tibshirani, 1993, p. 149). Again using the out-of-sample validation set, this means $d = 67$ (and $d = 49$ for Democrats, $d = 41$ for Republicans, and $d = 21$ for independents/not sure). The results for the unconditional and conditional permutation tests (for all respondents and by partisanship) are shown in Figures A.13-A.20.

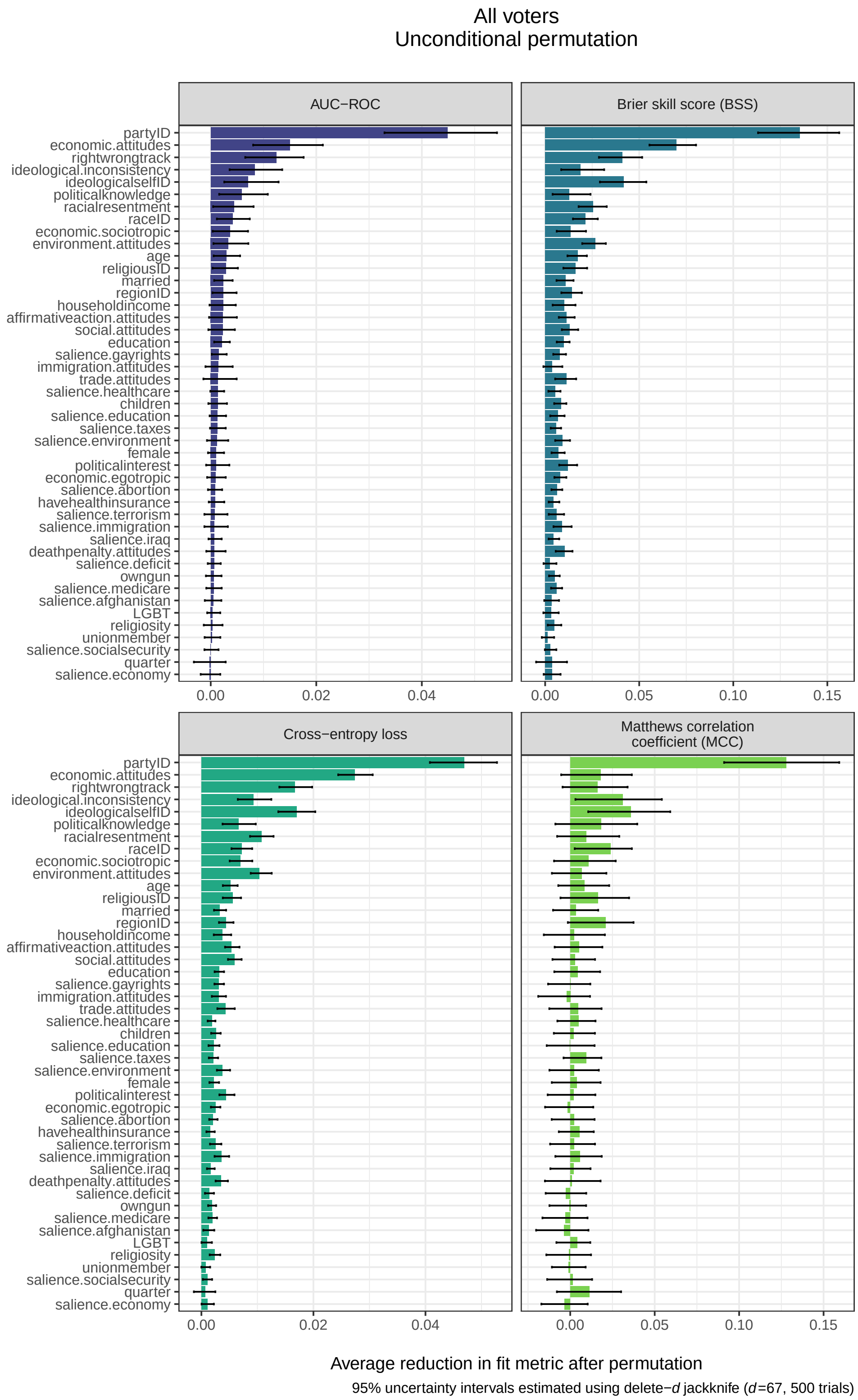

Figure A.13: Feature importance estimates from the learning ensemble using unconditional permutation tests (all respondents, with uncertainty intervals).

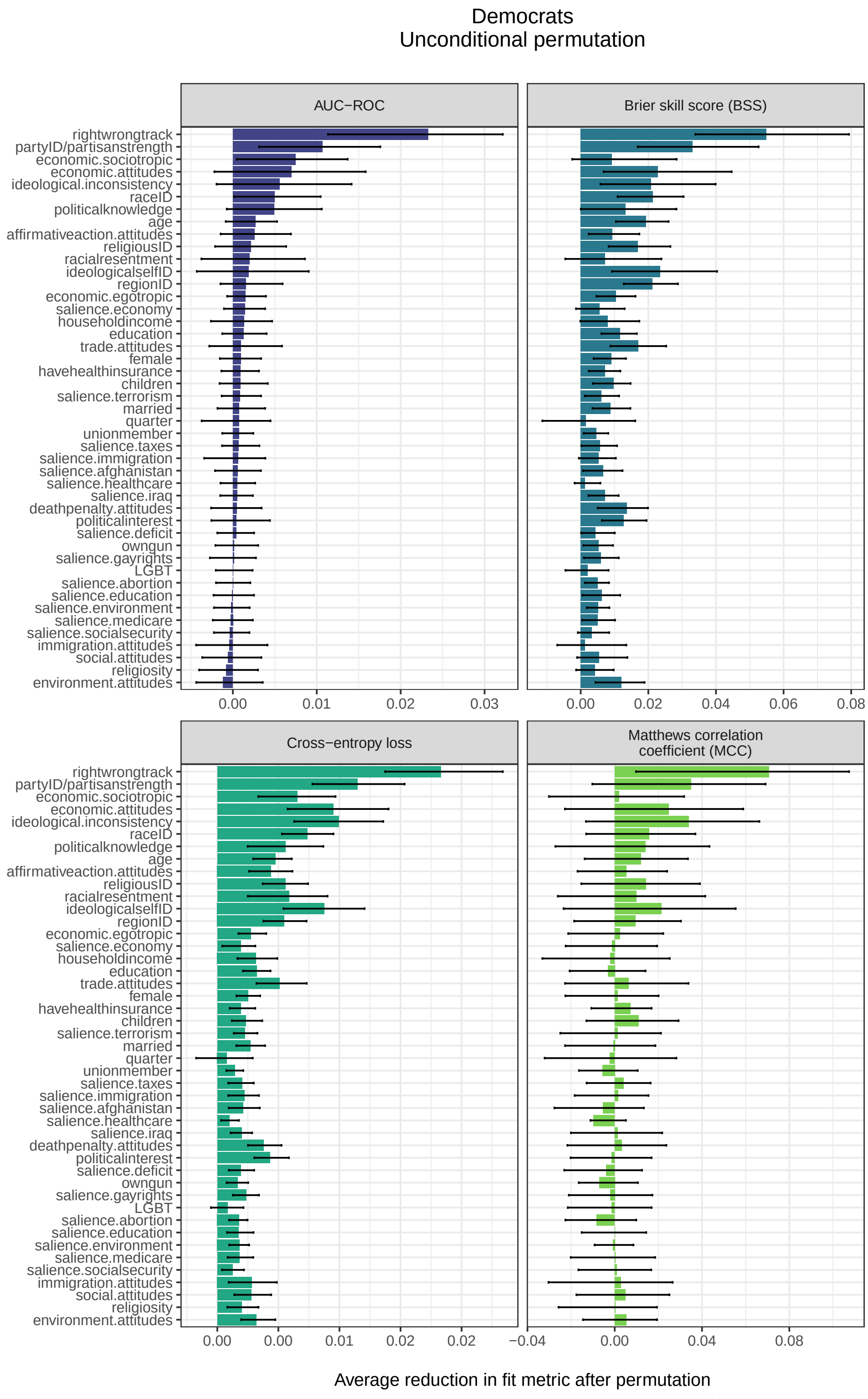

Figure A.14: Feature importance estimates from the learning ensemble using unconditional permutation tests (Democrats, with uncertainty intervals).

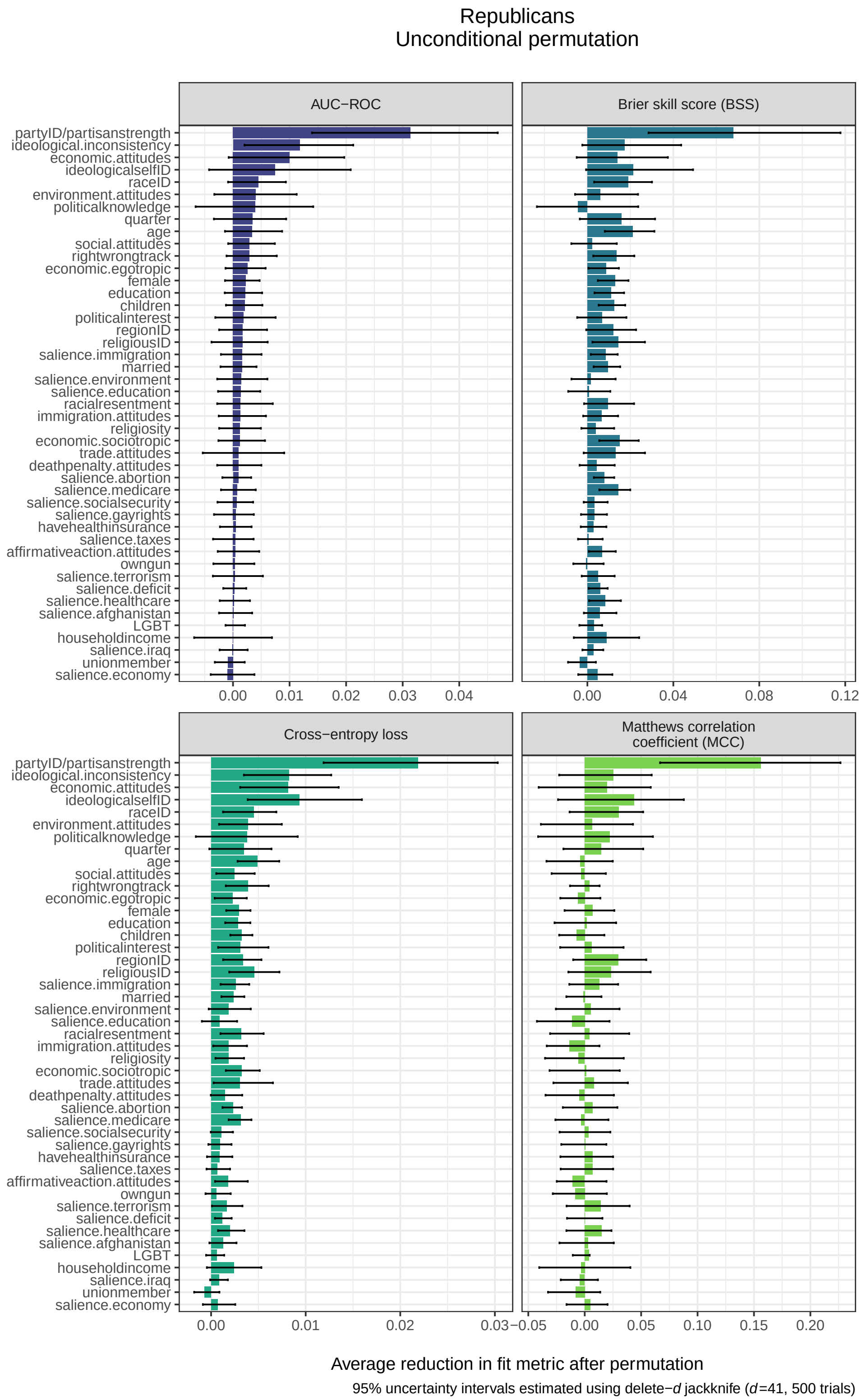

Figure A.15: Feature importance estimates from the learning ensemble using unconditional permutation tests (Republicans, with uncertainty intervals).

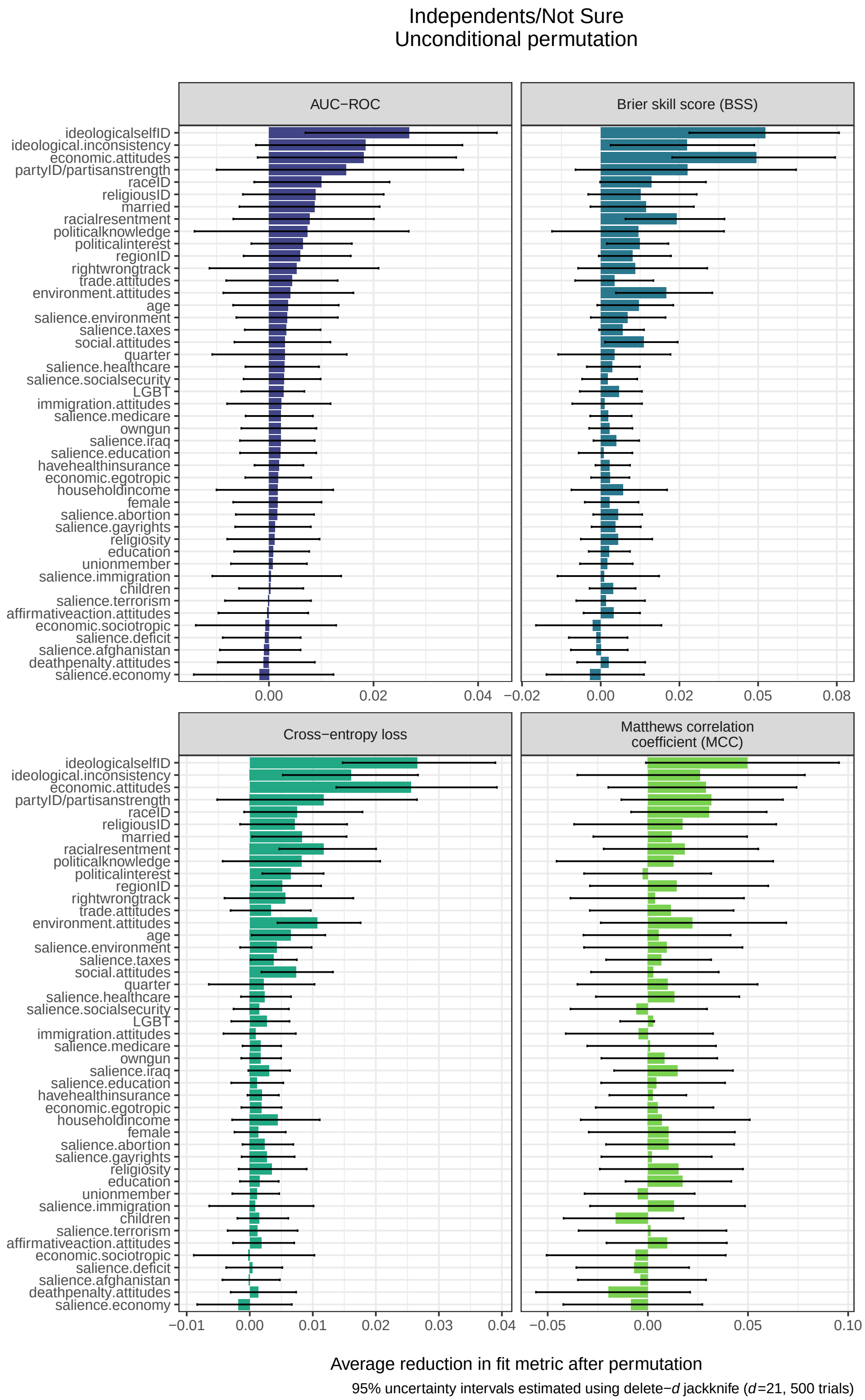

Figure A.16: Feature importance estimates from the learning ensemble using unconditional permutation tests (Independents/Not Sure, with uncertainty intervals).

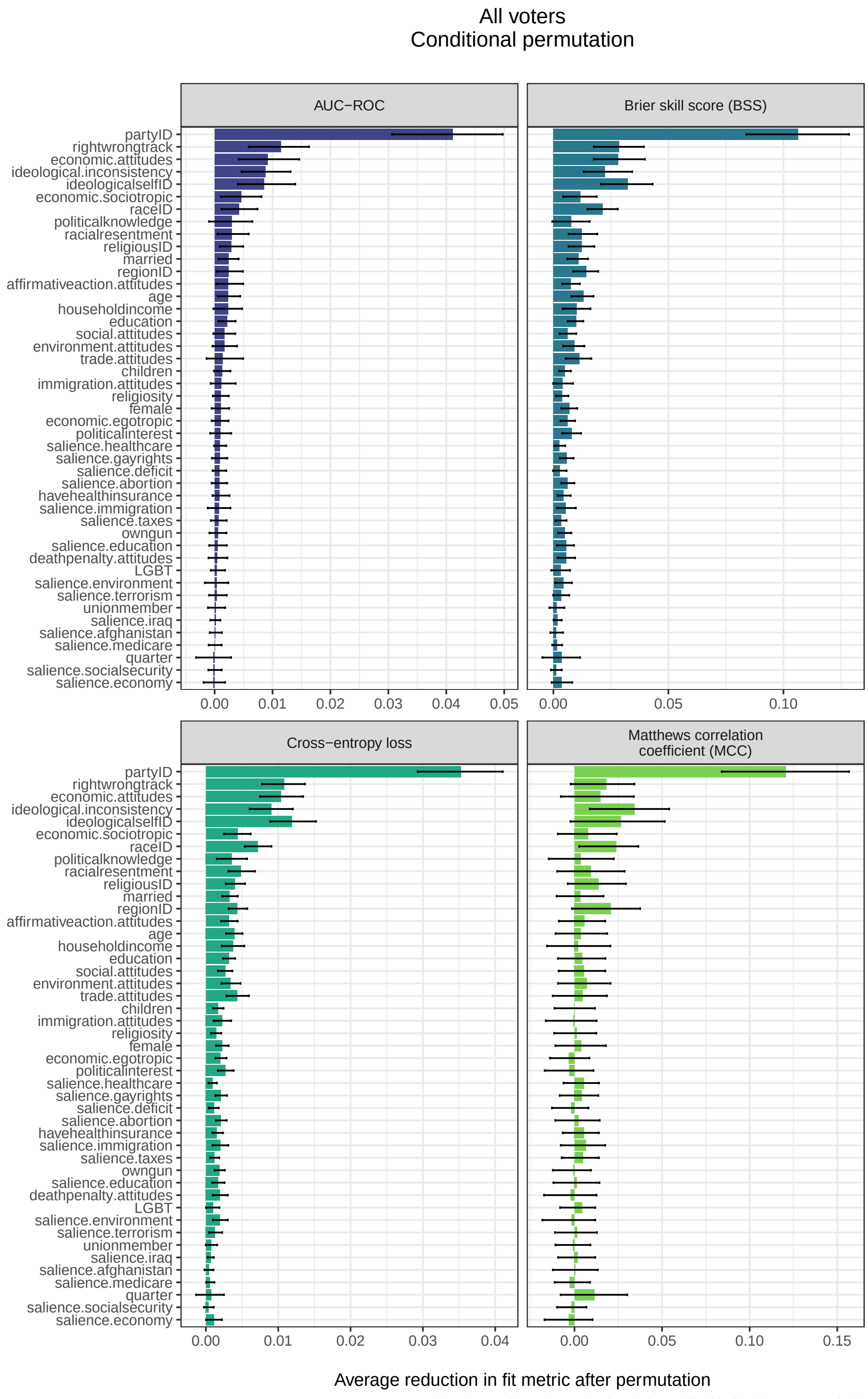

Figure A.17: Feature importance estimates from the learning ensemble using conditional permutation tests (all respondents, with uncertainty intervals).

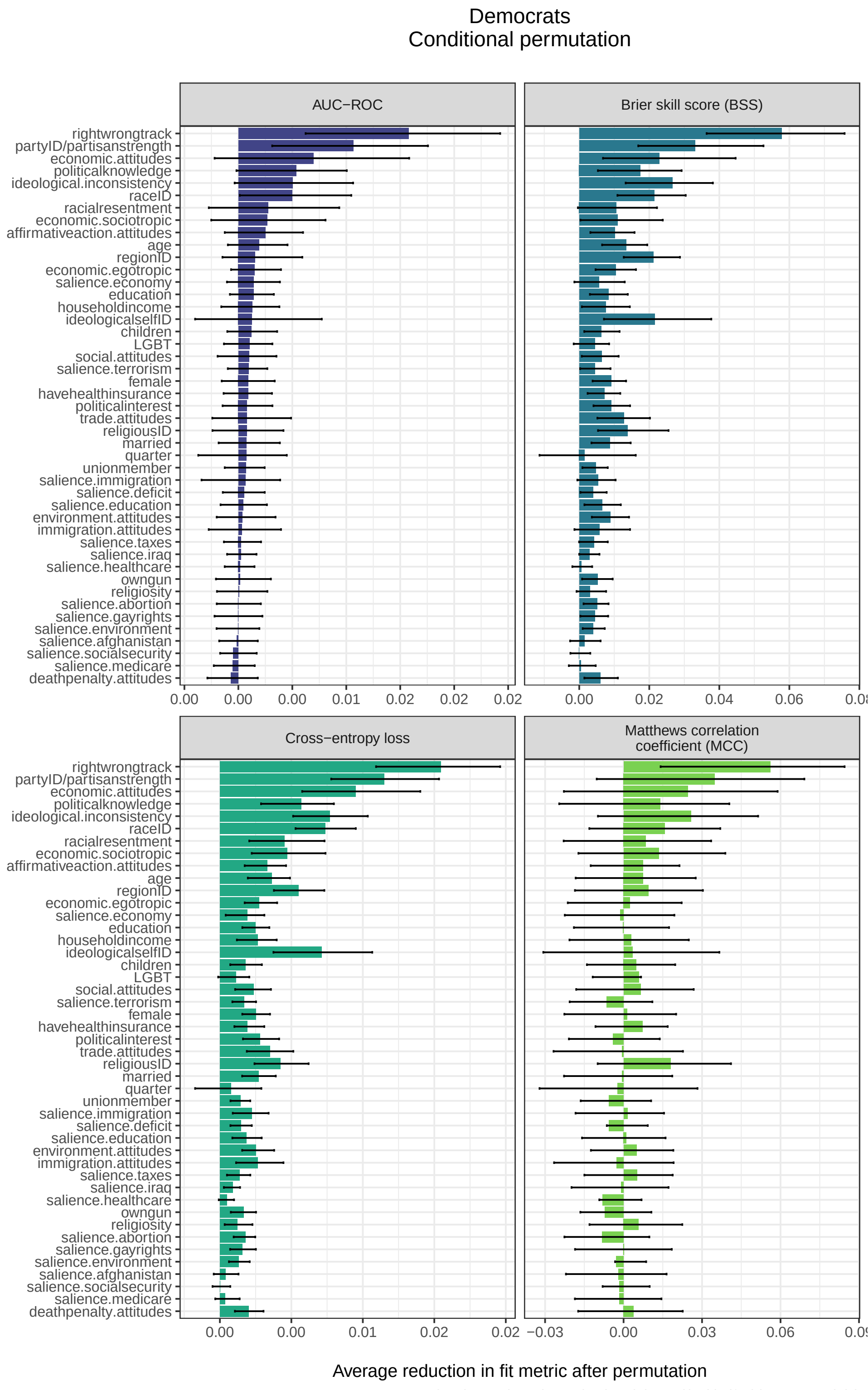

Figure A.18: Feature importance estimates from the learning ensemble using conditional permutation tests (Democrats, with uncertainty intervals).

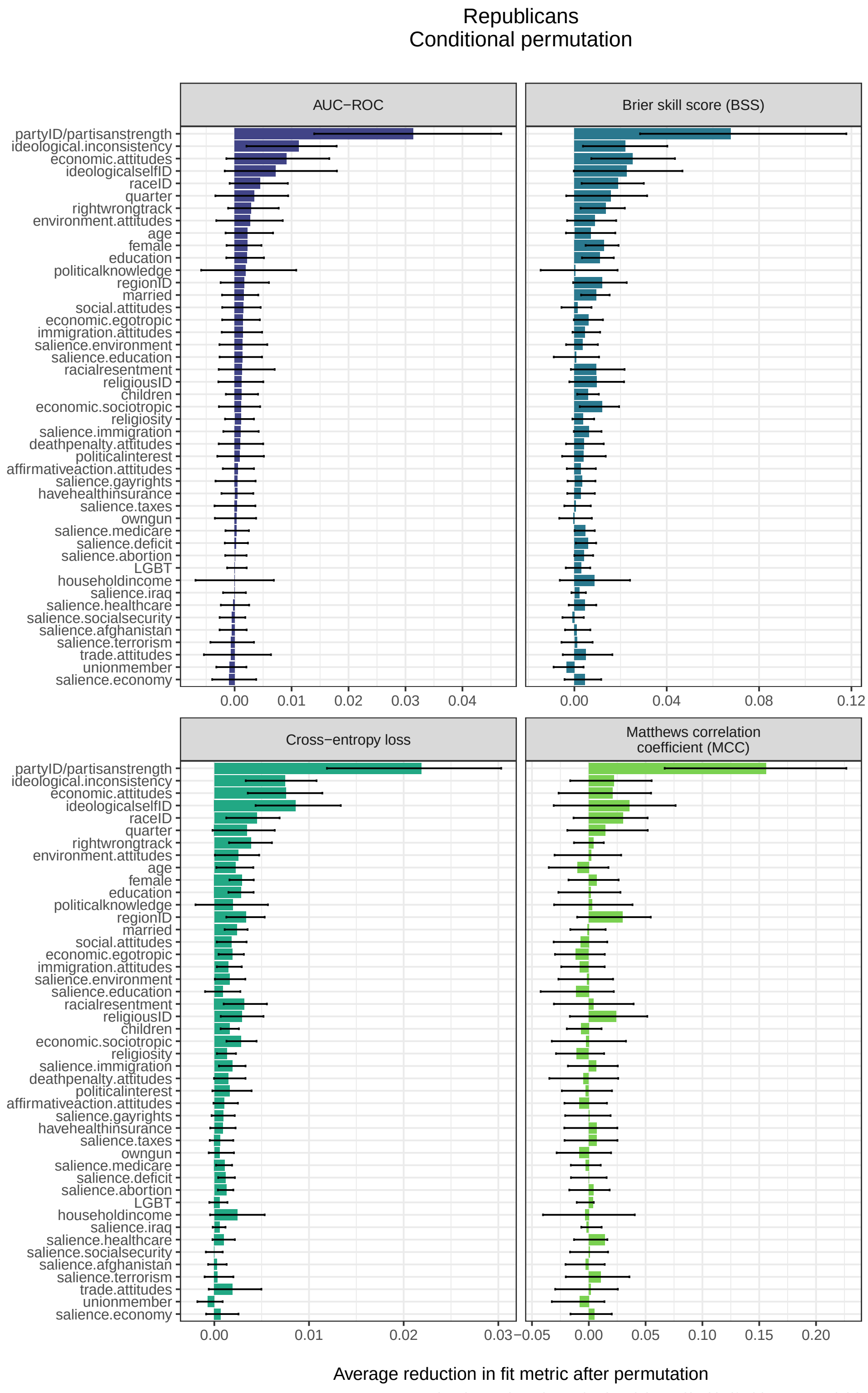

Figure A.19: Feature importance estimates from the learning ensemble using conditional permutation tests (Republicans, with uncertainty intervals).

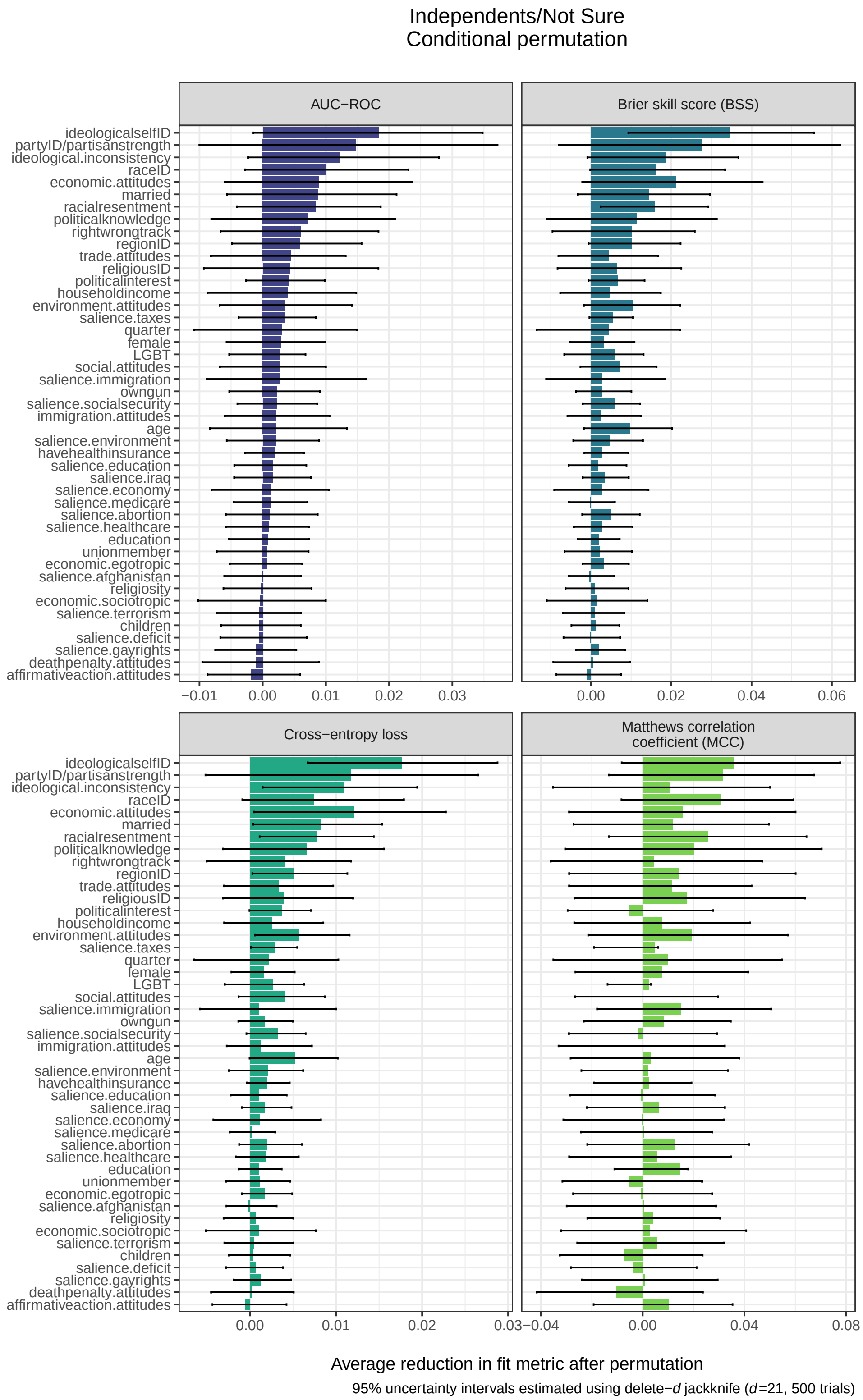

Figure A.20: Feature importance estimates from the learning ensemble using conditional permutation tests (Independents/Not Sure, with uncertainty intervals).

J Results from imputed data

In the main analysis, we use $12,914$ observations from the 2012 Cooperative Campaign Analysis Project with complete profiles across the swing voter measure and 45 predictor variables. These observations are split into a training set ($N = 8,395$) and a validation set ($N = 4,519$).

Below, we present results using observations with missing values on at least one of the 45 predictor variables. Specifically, we perform imputation using bagged decision trees with the **step_impute_bag** function from the **tidymodels** package in **R** (Kuhn and Johnson, 2013; Kuhn and Wickham, 2020). Because we do not impute values for the target swing voter variable, this provides $4,258$ observations with complete profiles after imputation on the predictors. Figure A.21 illustrates our partitioning of the data.

Figure A.21: Structure of the 2012 Cooperative Campaign Analysis Project data.

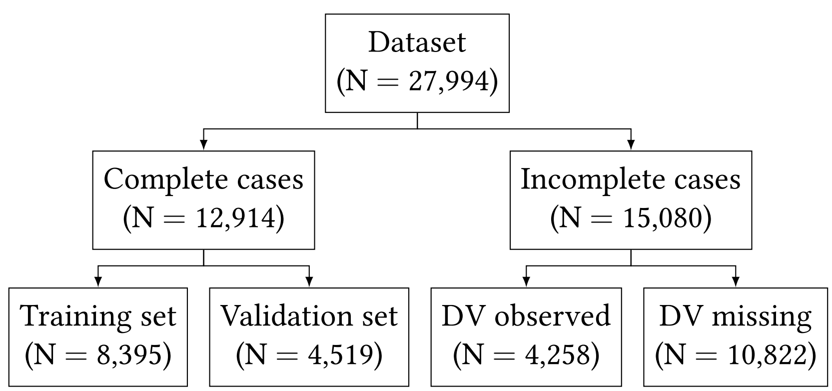

Figure A.22 is an analogue of Figure 2 in the main text, using the $4,258$ observations from the imputed dataset. Figures A.23–A.30 provide the results of unconditional and conditional permutation tests (analogues of Figure 3 in the main text and the figures in Appendix Section D) for the imputed observations. Both analyses produce substantively similar findings to those obtained with the original validation set (i.e., the results presented in the main text).

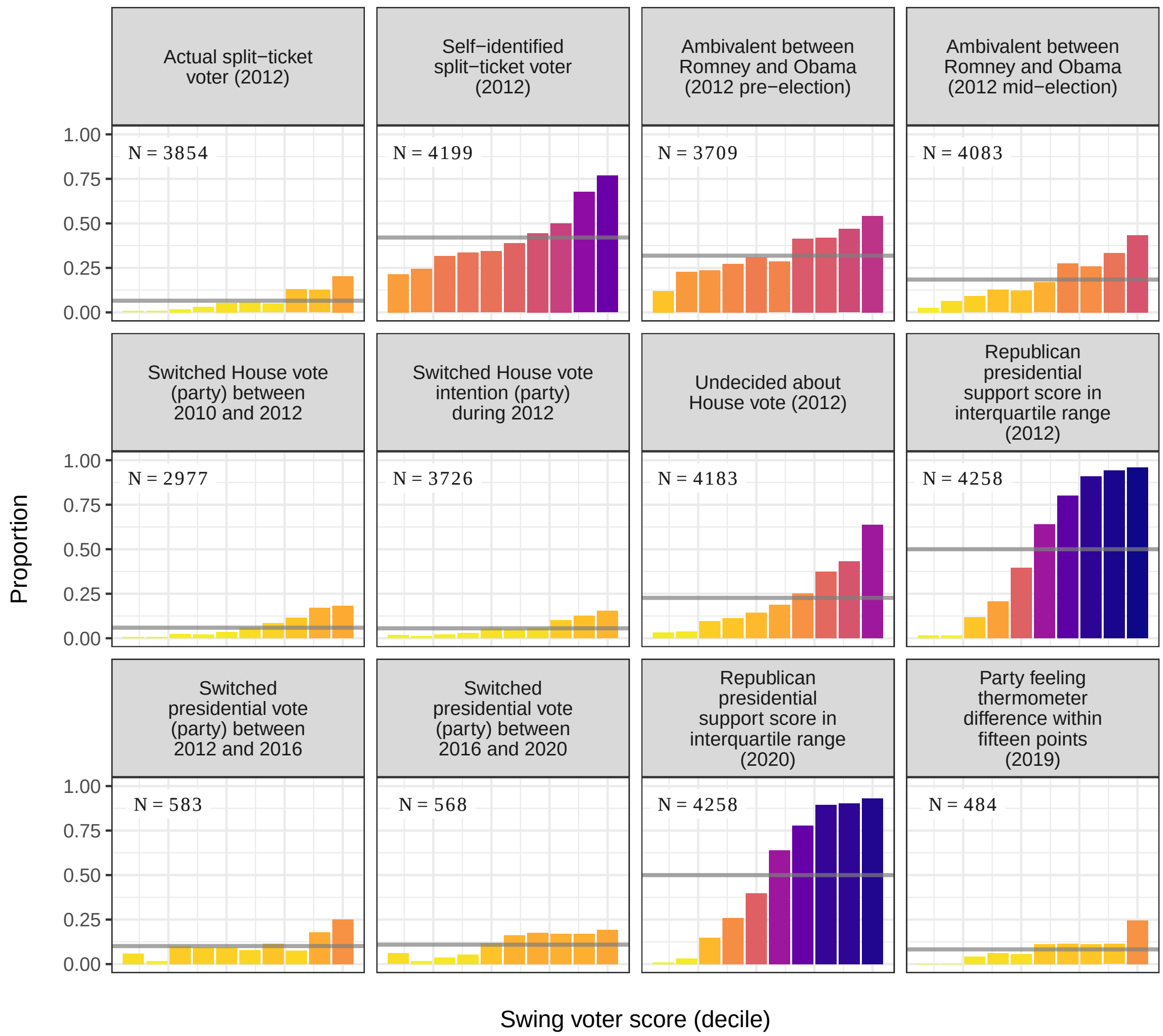

Figure A.22: Predictive performance on additional indicators of swing voter propensity using imputed out-of-sample data. Horizontal bars show overall proportion of respondents satisfying the corresponding indicator. Ambivalence defined as placing the candidates within one point of each other on a five-point favorability scale. Republican presidential support scores are calculated by estimating separate ensemble models of 2012 presidential vote choice and 2020 presidential vote intention.

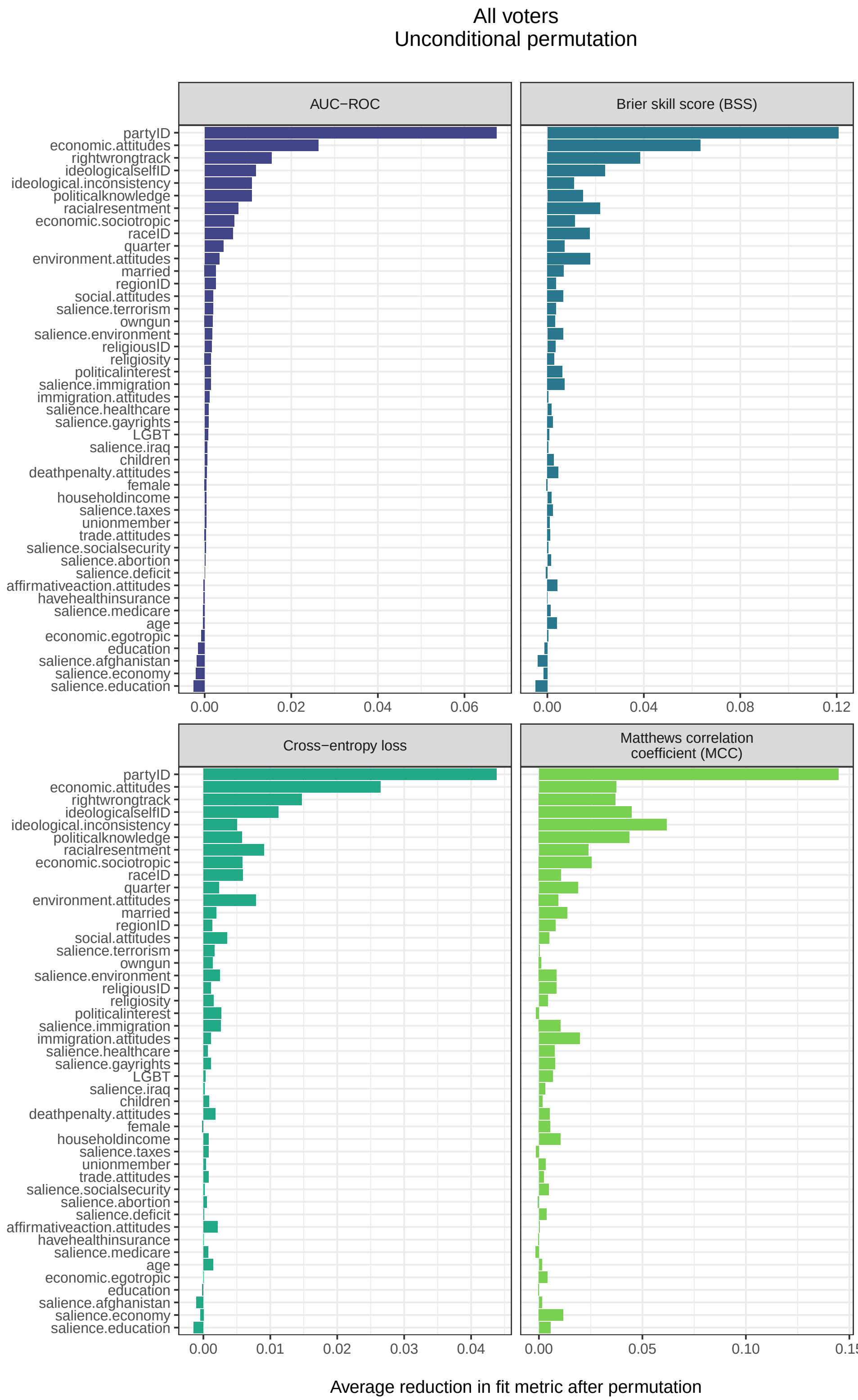

Figure A.23: Feature importance estimates from the learning ensemble using unconditional permutation tests (all respondents, imputed out-of-sample data).

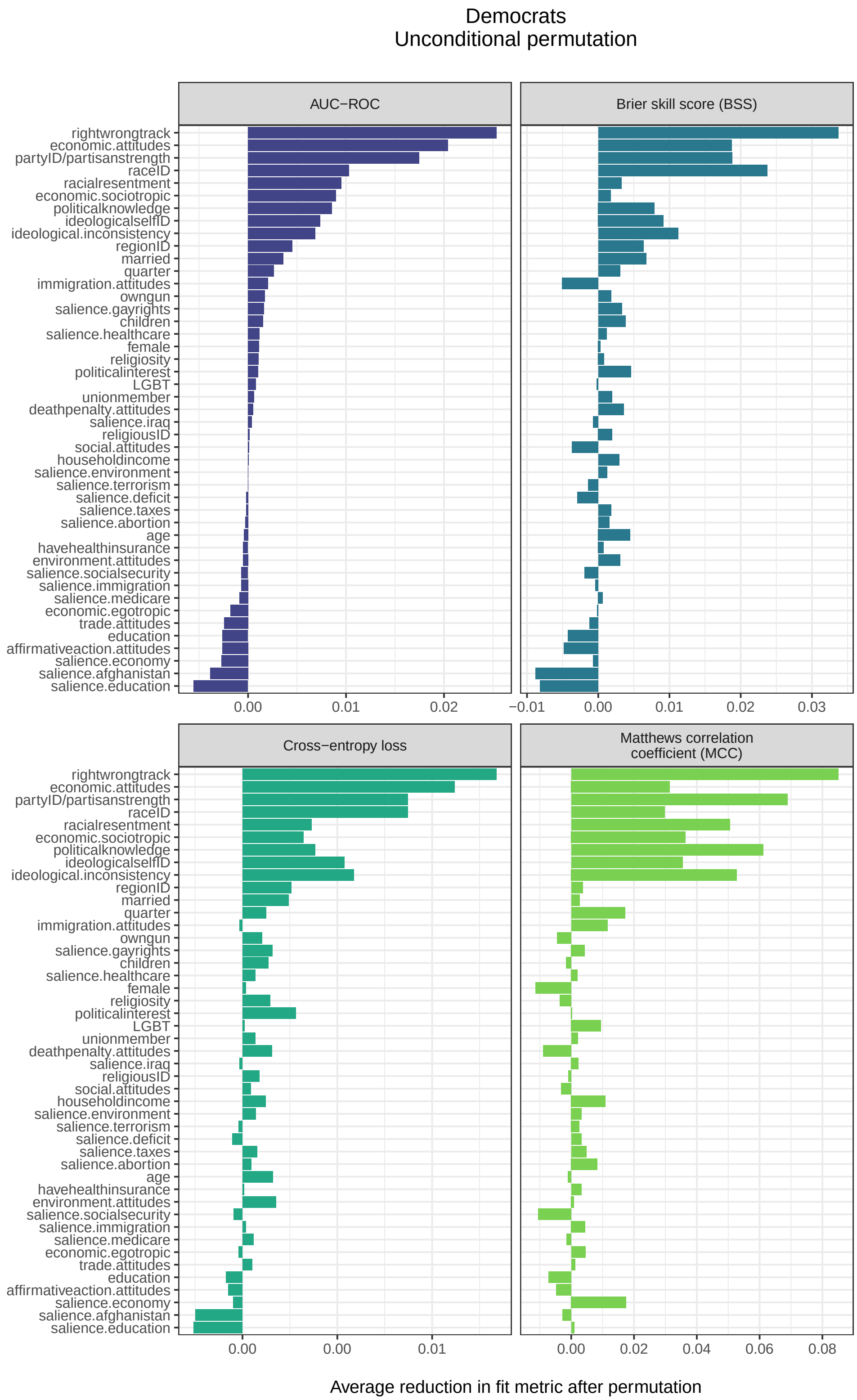

Figure A.24: Feature importance estimates from the learning ensemble using unconditional permutation tests (Democrats, imputed out-of-sample data).

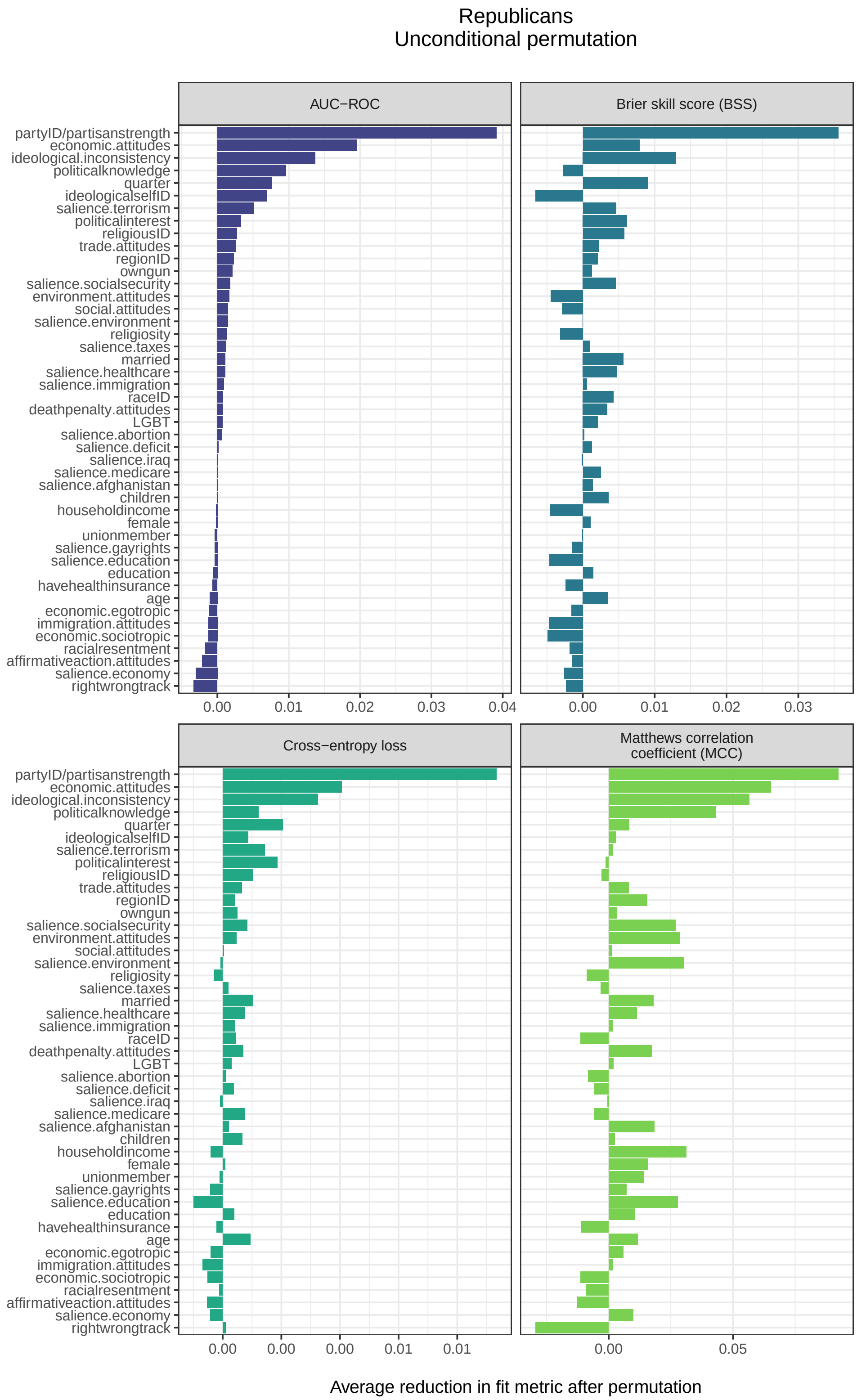

Figure A.25: Feature importance estimates from the learning ensemble using unconditional permutation tests (Republicans, imputed out-of-sample data).

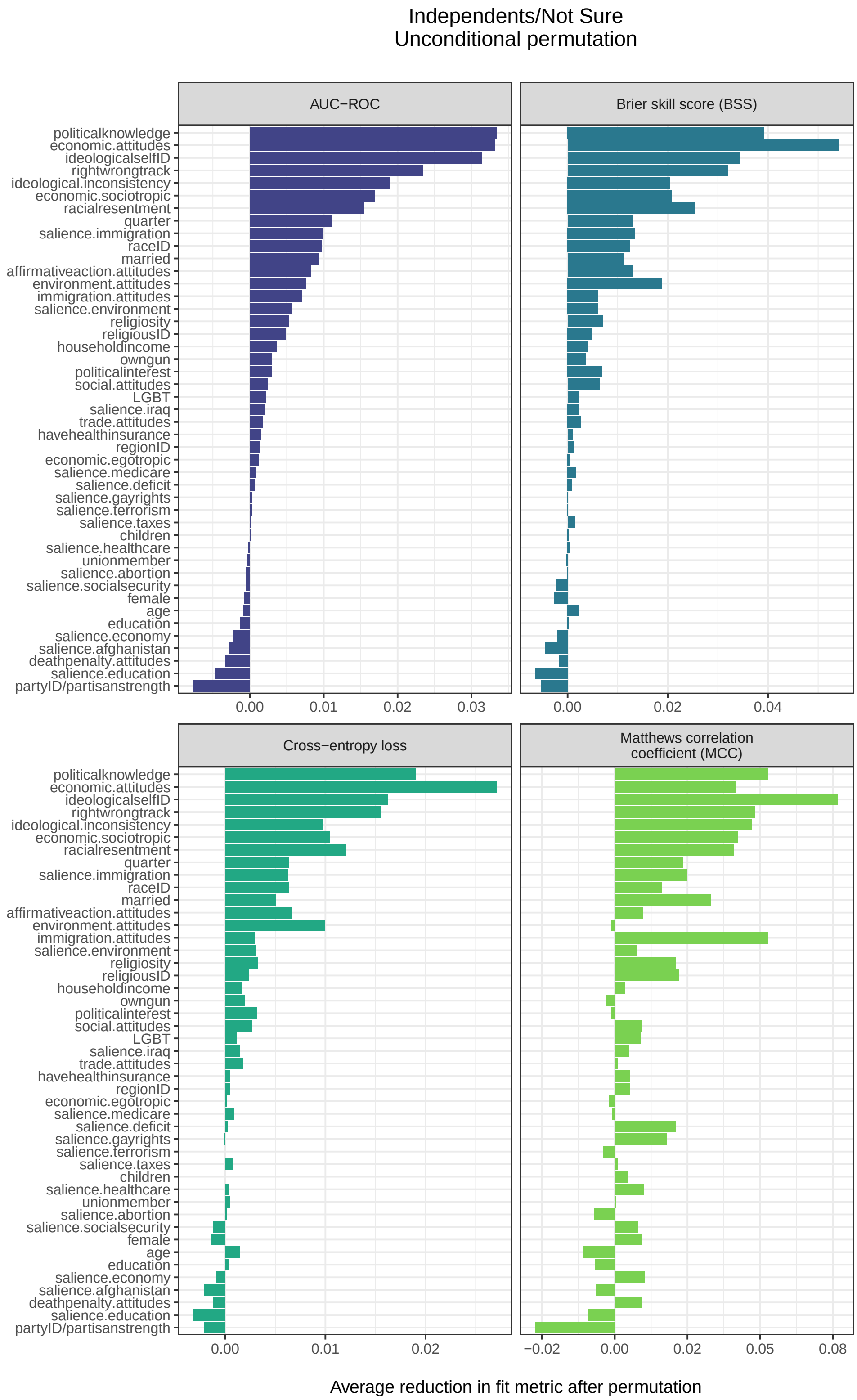

Figure A.26: Feature importance estimates from the learning ensemble using unconditional permutation tests (Independents/Not Sure, imputed out-of-sample data).

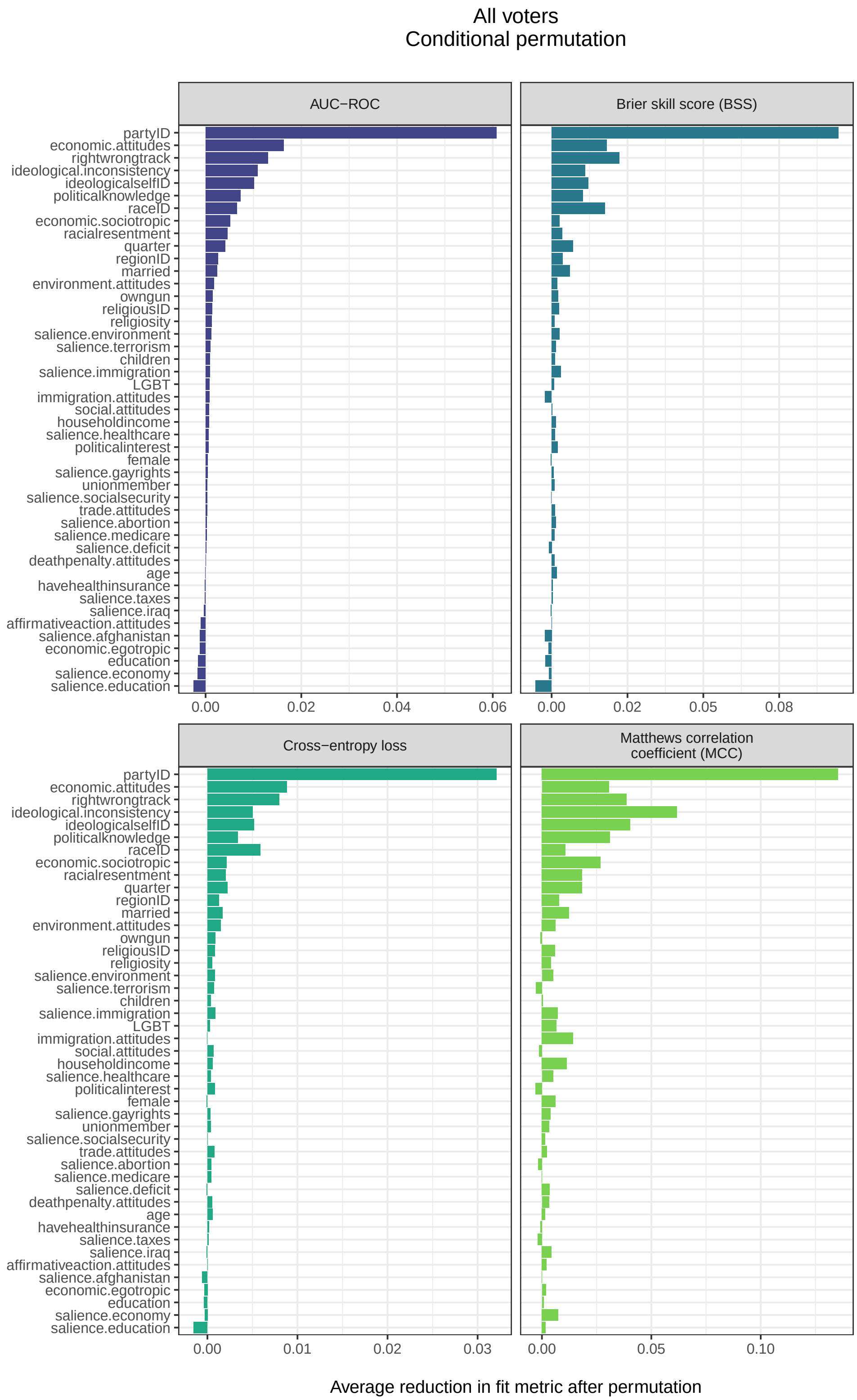

Figure A.27: Feature importance estimates from the learning ensemble using conditional permutation tests (all respondents, imputed out-of-sample data).

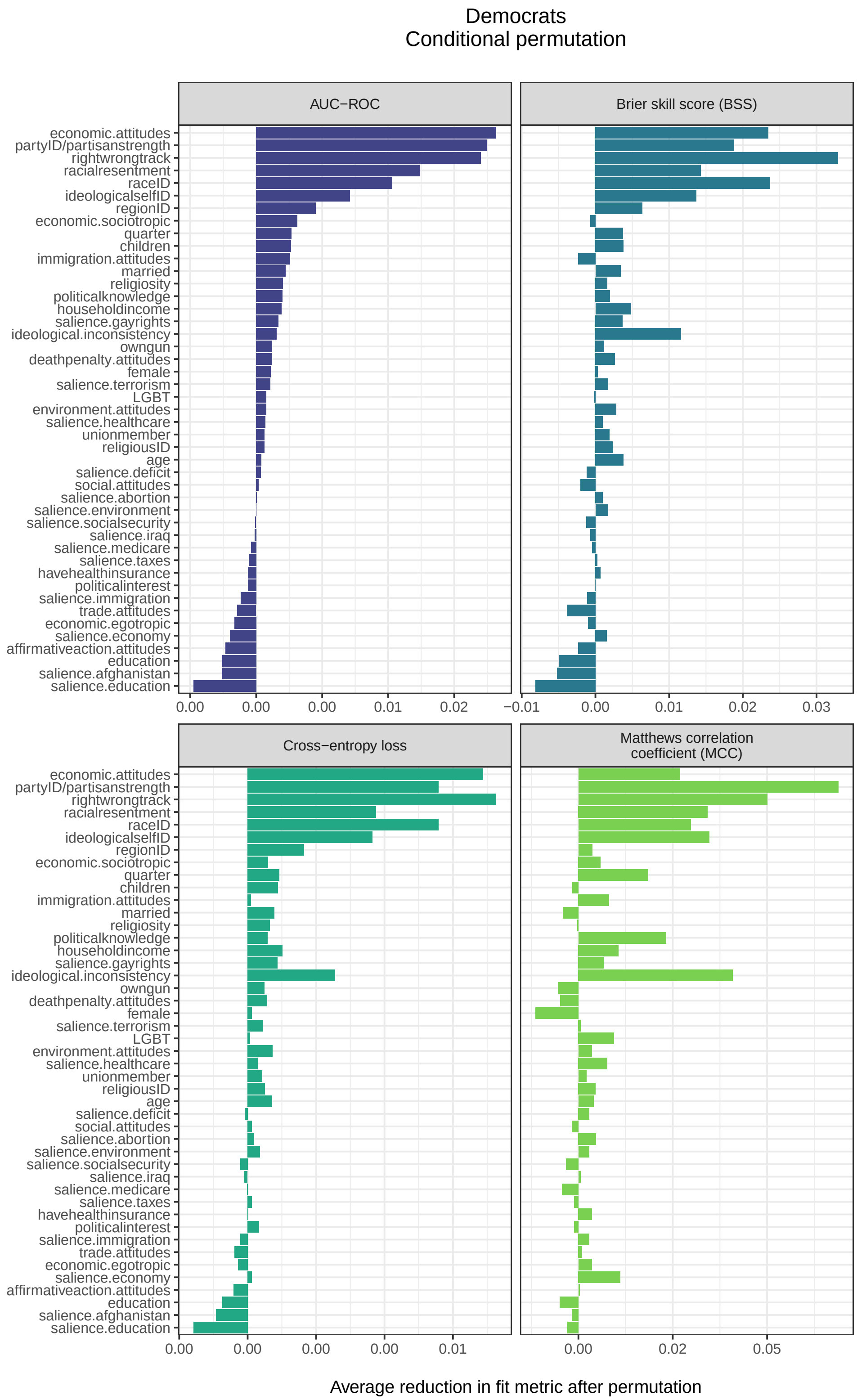

Figure A.28: Feature importance estimates from the learning ensemble using conditional permutation tests (Democrats, imputed out-of-sample data).

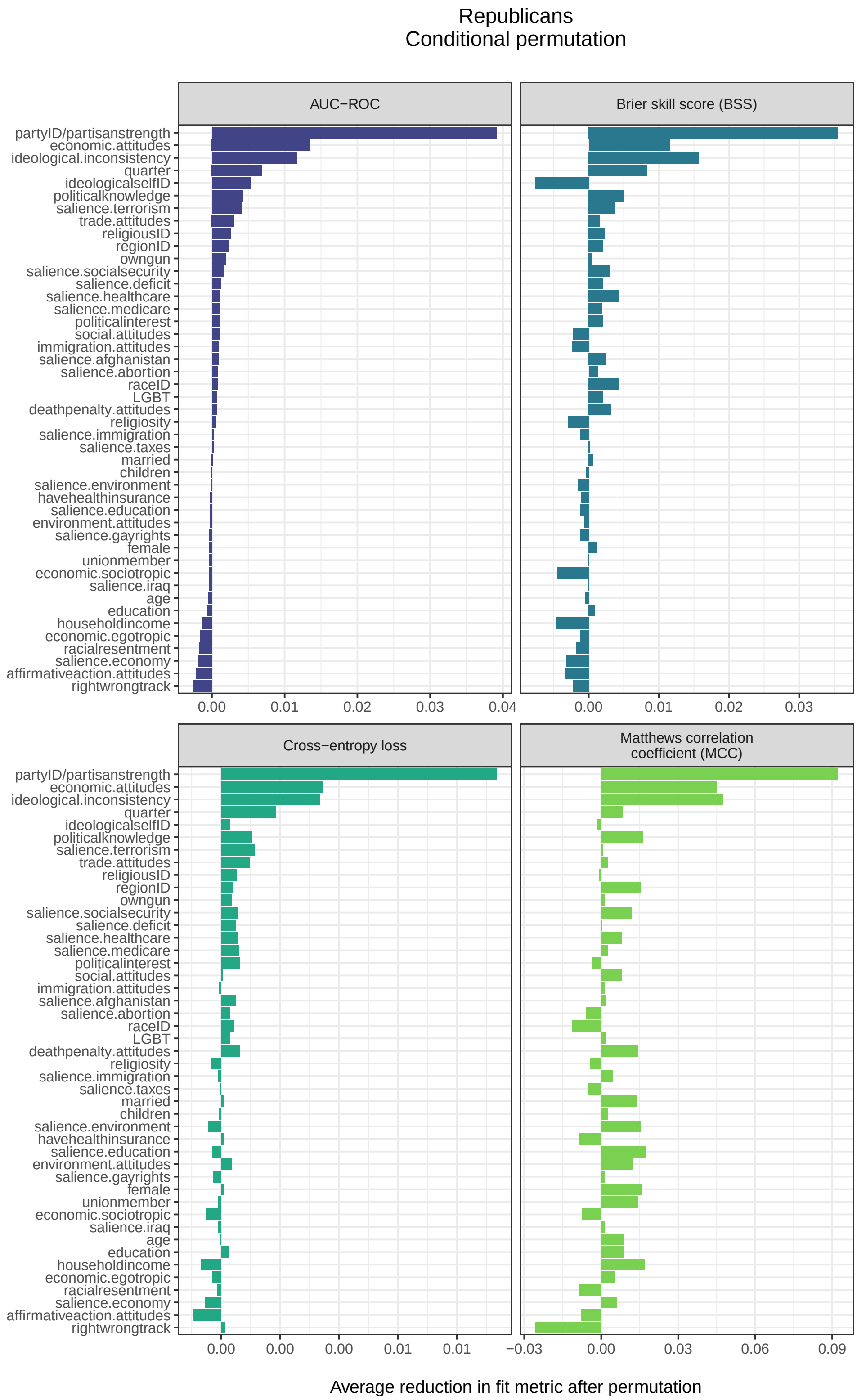

Figure A.29: Feature importance estimates from the learning ensemble using conditional permutation tests (Republicans, imputed out-of-sample data).

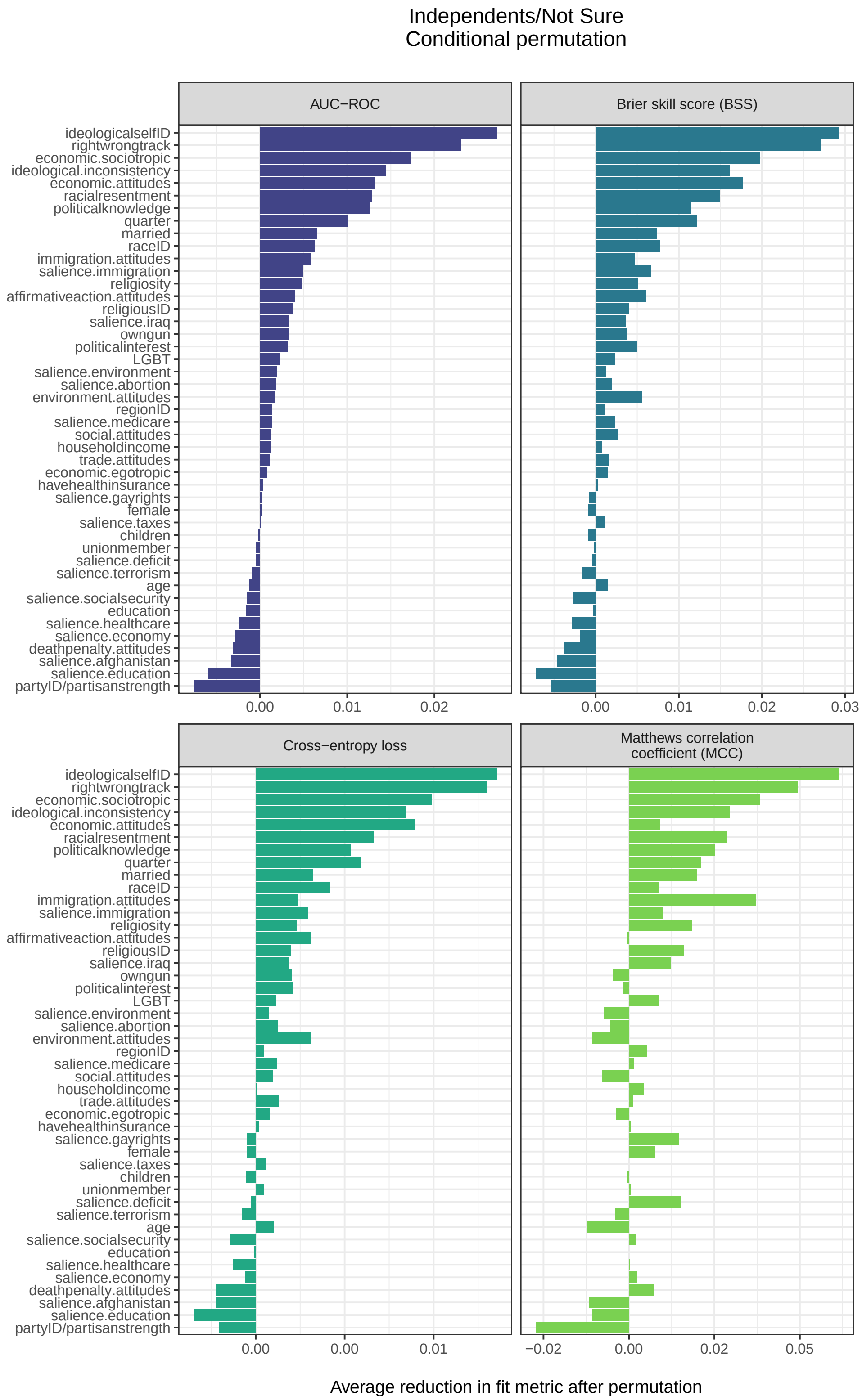

Figure A.30: Feature importance estimates from the learning ensemble using conditional permutation tests (Independents/Not Sure, imputed out-of-sample data).

References

Albanese, Davide, Samantha Riccadonna, Claudio Donati and Pietro Franceschi. 2018. "A Practical Tool for Maximal Information Coefficient Analysis." *GigaScience* 7(4):1–8.

Altman, Micah, Jeff Gill and Michael P. McDonald. 2004. *Numerical Issues in Statistical Computing for the Social Scientist.* New York: Wiley.

Bishop, Christopher M. 1995. *Neural Networks for Pattern Recognition.* New York: Oxford University Press.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45(1):5–32.

Carroll, Raymond J, Helmut Küchenhoff, F Lombard and Leonard A Stefanski. 1996. "Asymptotics for the SIMEX estimator in nonlinear measurement error models." *Journal of the American Statistical Association* 91(433):242–250.

Chan, Zenobia and Marc Ratkovic. 2020. "Improving Variable Importance Measures." Presented at the 2020 Annual Meeting of the Society for Political Methodology (PolMeth XXXVII).

Chatterjee, Arindam and Soumendra Nath Lahiri. 2011. "Bootstrapping Lasso Estimators." *Journal of the American Statistical Association* 106(494):608–625.

Chen, Tianqi and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. pp. 785–794.

Cook, John R. and Leonard A. Stefanski. 1994. "Simulation-Extrapolation Estimation in Parametric Measurement Error Models." *Journal of the American Statistical Association* 89(428):1314–1328.

Cook, Nancy R. 2007. "Use and Misuse of the Receiver Operating Characteristic Curve in Risk Prediction." *Circulation* 115(7):928–935.

Coyle, Jeremy and Mark J. van der Laan. 2018. Targeted Bootstrap. In *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies*, ed. Mark J. van der Laan and Sherri Rose. New York: Springer pp. 523–539.

Das, Debraj, Karl Gregory and S.N. Lahiri. 2019. "Perturbation Bootstrap in Adaptive Lasso." *Annals of Statistics* 47(4):2080–2116.

Debeer, Dries and Carolin Strobl. 2020. "Conditional Permutation Importance Revisited." *BMC Bioinformatics* 21(1):1–30.

Democracy Fund Voter Study Group. 2020. Views of the Electorate Research Survey, January 2019. Release 1: January 2020. Washington, DC: Democracy Fund.
URL: *https://www.voterstudygroup.org/*

Efron, Bradley and Robert J. Tibshirani. 1993. *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Federico, Christopher M. and Corrie V. Hunt. 2013. "Political Information, Political Involvement, and Reliance on Ideology in Political Evaluation." *Political Behavior* 35(1):89–112.

Fokkema, Marjolein. 2020. "Fitting Prediction Rule Ensembles with R Package pre." *Journal of Statistical Software* 92(12):1–30.

Freund, Yoav and Robert E. Schapire. 1997. "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting." *Journal of Computer and System Sciences* 55(1):119–139.

Friedman, Jerome H. 2001. "Greedy Function Approximation: A Gradient Boosting Machine." *Annals of Statistics* 29(5):1189–1232.

Friedman, Jerome H. and Bogdan E. Popescu. 2008. "Predictive Learning via Rule Ensembles." *The Annals of Applied Statistics* 2(3):916–954.

Geurts, Pierre, Damien Ernst and Louis Wehenkel. 2006. "Extremely Randomized Trees." *Machine Learning* 63(1):3–42.

Grimmer, Justin, Solomon Messing and Sean J. Westwood. 2017. “Estimating Heterogeneous Treatment Effects and the Effects of Heterogeneous Treatments with Ensemble Methods.” *Political Analysis* 25(4):413–434.

Hastie, Trevor, Robert Tibshirani and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* 2nd ed. New York: Springer.

Imai, Kosuke, James Lo and Jonathan Olmsted. 2016. “Fast Estimation of Ideal Points with Massive Data.” *American Political Science Review* 110(4):631–656.

Ishwaran, Hemant and Min Lu. 2019. “Standard Errors and Confidence Intervals for Variable Importance in Random Forest Regression, Classification, and Survival.” *Statistics in Medicine* 38(4):558–582.

Jackman, Simon, John Sides, John Tesler and Lynn Vavreck. 2012. The 2012 Cooperative Campaign Analysis Project. Palo Alto, CA: YouGov/Polimetrix.

Karatzoglou, Alexandros, Alex Smola, Kurt Hornik and Achim Zeileis. 2004. “kernlab – An S4 Package for Kernel Methods in R.” *Journal of Statistical Software* 11(9):1–20.

Kaufman, Leonard and Peter J. Rousseeuw. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis.* New York: Wiley.

Key, Jr., V.O. 1966. *The Responsible Electorate.* New York: Vintage.

Kuhn, Max and Hadley Wickham. 2020. *tidymodels: A Collection of Packages for Modeling and Machine Learning using tidyverse Principles.*
URL: *https://www.tidymodels.org*

Kuhn, Max and Kjell Johnson. 2013. *Applied Predictive Modeling.* New York: Springer.

Lim, Michael and Trevor Hastie. 2015. “Learning Interactions via Hierarchical Group-Lasso Regularization.” *Journal of Computational and Graphical Statistics* 24(3):627–654.

Lim, Michael and Trevor Hastie. 2018. *glinternet: Learning Interactions via Hierarchical Group-Lasso Regularization.* R package version 1.0.8.
URL: *cran.r-project.org/package=glinternet*

Molnar, Christoph, Bernd Bischl and Giuseppe Casalicchio. 2018. "iml: An R package for Interpretable Machine Learning." *Journal of Open Source Software* 3(26):786–787.

Platt, John C. 2000. Probabilistic Outputs for Support Vector Machines. In *Advances in Large Margin Classifiers*, ed. Alexander J. Smola, Peter L. Bartlett, Bernhard Schöelkopf and Dale Schuurmans. Cambridge, MA: MIT Press pp. 61–74.

Reshef, David N., Yakir A. Reshef, Hilary K. Finucane, Sharon R. Grossman, Gilean McVean, Peter J. Turnbaugh, Eric S. Lander, Michael Mitzenmacher and Pardis C. Sabeti. 2011. "Detecting Novel Associations in Large Data Sets." *Science* 334(6062):1518–1524.

Rousseeuw, Peter J. 1987. "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis." *Journal of Computational and Applied Mathematics* 20:53–65.

Shao, Jun and C.F.J Wu. 1989. "A General Theory for Jackknife Variance Estimation." *Annals of Statistics* 17(3):1176–1197.

Shaw, Daron R. 2008. Swing Voting and U.S. Presidential Elections. In *The Swing Voter in American Politics*, ed. William G. Mayer. Washington, DC: Brookings Institution Press pp. 75–101.

Smidt, Corwin D. 2017. "Polarization and the Decline of the American Floating Voter." *American Journal of Political Science* 61(2):365–381.

Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin and Achim Zeileis. 2008. "Conditional Variable Importance for Random Forests." *BMC Bioinformatics* 9(1):307.

Vapnik, Vladimir. 2000. *The Nature of Statistical Learning Theory.* 2nd ed. New York: Springer.

Venables, W. N. and B. D. Ripley. 2002. *Modern Applied Statistics with S.* Fourth ed. New York: Springer.

Wager, Stefan, Trevor Hastie and Bradley Efron. 2014. "Confidence Intervals for Random Forests: The Jackknife and the Infinitesimal Jackknife." *Journal of Machine Learning Research* 15(1):1625–1651.

Weghorst, Keith R. and Staffan I. Lindberg. 2013. "What Drives the Swing Voter in Africa?" *American Journal of Political Science* 57(3):717–734.

Wright, Marvin N. and Andreas Ziegler. 2017. "ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R." *Journal of Statistical Software* 77(1):1–17.