Empirical

Predicting Current Voting Intentions by Big Five Personality Domains, Facets, and Nuances – A Random Forest Analysis Approach in a German Sample

Cornelia Sindermann[1] , René Mõttus[2] , Dmitri Rozgonjuk[1,3] ,

Christian Montag[1]

[1] *Department of Molecular Psychology, Ulm University, Ulm, Germany.* **[2]** *Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom.* **[3]** *Institute of Mathematics and Statistics, University of Tartu, Tartu, Estonia.*

Abstract

To understand what was driving individual differences in voting intentions in a large German sample, we investigated the predictability of voting intentions from the Big Five personality domains, facets, and nuances, thereby tackling shortcomings of previous studies. Using random forest analyses in a dataset of N = 4,286 individuals (46.01% men), separate models were trained to predict intentions to 1) not vote versus to vote, 2) vote for a specific party, and 3) vote for a left-versus right-from-the-center party from either the Big Five personality domains, facets, or nuances (represented by individual items). Except for intentions to not vote versus to vote, balanced accuracies to predict voting intentions marginally exceeded those achieved by a baseline learner always predicting the majority class. Using nuances over facets and domains slightly increased balanced accuracies. Results indicate that additional variables should be considered to accurately predict voting intentions, at least in German samples.

Keywords

Big Five, personality, voting intentions, voting, random forest

> **Relevance Statement**
>
> The present work aimed to overcome limitations of previous research by 1) implementing a prediction design and 2) investigating Big Five personality facets and nuances besides domains to compare predictive accuracy of models across levels of the trait hierarchy.
>
> **Key Insights**
> - Predictability of voting intentions from personality was investigated.
> - Big Five personality domains, facets, and nuances were examined.
> - Random forest analyses were implemented.
> - Voting intentions could hardly be predicted from self-reported personality.

Voting is a civil right in democracies around the world. By voting, citizens have the power to influence political developments. Therefore, understanding what underlies voting decisions and political preferences is an important aim in several fields of research like political science and psychology. With regard to personality psychology, previous findings indicate weak associations between the Big Five personality domains and, for example, political left- versus right-positioning and party preferences, accordingly (Krieger et al., 2019). However, it is unclear whether it is possible to accurately *predict* one's current voting intentions and preferences for a specific party from broad personality traits. Moreover, it is unclear whether narrower personality traits, facets and nuances, can enhance predictive accuracy over Big Five personality domains. Investigating these questions is especially interesting in countries like Germany, where the party system is more complex than in, for example, the USA or the UK. Therefore, the present study aimed to investigate whether one can predict Germans' voting intentions from self-reported Big Five personality domains, facets, and nuances by means of random forest analyses.

The Personality Trait Hierarchy

Personality traits constitute a pool of hierarchically organized characteristics, with the Big Five personality domains of Openness (to Experience), Conscientiousness, Extraversion, Agreeableness, and Neuroticism (Costa & McCrae, 1992; Goldberg, 1990; Tupes & Christal, 1992) at one of the highest levels. High scores in Openness describe individuals who are open for aesthetics, like to make new experiences, and/or are intellectually curious. Individuals scoring high in Conscientiousness carry out their duties carefully, are orderly, ambitious, and/or disciplined. Highly extraverted individuals are socially outgoing, tend to feel positive emotions, and/or are active and assertive. High scores in Agreeableness indicate that a person is helpful, altruistic, forgiving, and/or considerate. High Neuroticism describes individuals who are anxious, hostile, impulsive, and/or might show depressive tendencies (Rammstedt & Danner, 2017).

PsychOpen GOLD

Each of the Big Five personality domains can be split into different narrower aspects also called facets, although there is no consensus yet regarding an exhaustive set of facets (Costa & McCrae, 1992; DeYoung et al., 2007; John et al., 1991; Rammstedt & Danner, 2017). It has been demonstrated that facets are not just different ways of expressing the Big Five personality domains but reflect unique personality characteristics (Jang et al., 1998) and that they often provide incremental predictive value over the Big Five personality domains (Judge et al., 2013; Paunonen & Ashton, 2001). Below facets, many individual questionnaire items also contain reliable and valid trait information beyond what they share with other items (Mõttus et al., 2020). That is, many items represent unique personality traits, often called nuances, that are narrower than facets and domains (McCrae, 2015; McCrae & Mõttus, 2019; Mõttus, Kandler, et al., 2017) and show unique predictive validity (Seeboth & Mõttus, 2018).

As a result, it seems advisable to compare different levels of the personality trait hierarchy in terms of predictive accuracy when exploring the links of personality with other variables, such as voting intentions. In some cases, the Big Five personality domains may provide just as accurate predictions as lower levels of the trait hierarchy, but the reverse is more likely (Mõttus, Bates, et al., 2017; Mõttus et al., 2020).

Personality Traits and Political Attitudes

When investigating voting intentions, a first step includes examination of intentions to not vote versus to vote. The few previous studies investigating associations between Big Five personality domains and non-voting versus voting intentions as well as non-voting versus voting behavior/self-reported voter turnout have reported inconclusive results. The mixed findings might be due to differences in voting settings (e.g., countries and years) as well as specific measures used. Hence, the results might only be applicable to the specific samples under investigation (Dawkins, 2017; Mondak & Halperin, 2008; Mondak et al., 2010; Sindermann et al., 2020).

In contrast, results on associations between Big Five personality domains and political ideology (which is likely to be associated with voting [intentions] for a specific party) are more conclusive, especially regarding the Openness and Conscientiousness domains. While Openness has been positively associated with a more left-leaning and liberal ideological self-positioning, Conscientiousness has been positively associated with a more right-leaning and conservative self-positioning (Cooper et al., 2013; Gerber et al., 2011; Hirsh et al., 2010; Krieger et al., 2019; Mondak & Halperin, 2008; Sibley et al., 2012). The effects of the Extraversion, Agreeableness, and Neuroticism domains are usually smaller, are not as often replicated, and the sizes and directions of effects vary between studies (Cooper et al., 2013; Hirsh et al., 2010; Krieger et al., 2019; Mondak & Halperin, 2008).

Results of two meta-analyses on German samples support the associations of political ideology with the Openness and Conscientiousness domains. Moreover, these studies

report smaller effects of Agreeableness (in one of the meta-analyses) and Neuroticism (in both meta-analyses), which were both related to higher left-leaning and lower conservative ideological self-positioning (Krieger et al., 2019; Sibley et al., 2012). These findings support the existence of associations between personality traits, especially the Openness and Conscientiousness domains, and political ideology in Germany and lead to the assumption that personality traits might also be associated with voting intentions for specific parties.

In Germany, there are currently six parties/party alliances included in the *Bundestag*, the German federal parliament. Three of them are seen as right-from-the-center parties/party alliances, namely the Alternative for Germany (AfD), the party alliance of the Christian Democratic Union of Germany and the Christian Social Union in Bavaria (CDU/CSU), and the Free Democratic Party (FDP). The other three, Bündnis 90/Die Grünen (Alliance 90/The Greens), Social Democratic Party of Germany (SPD), and DIE LINKE (The Left) are described as left-from-the-center parties/party alliances (Volkens et al., 2020). More specifically, the AfD is seen as a right-wing populist party and shows critical attitudes on immigration related topics (Berbuir et al., 2015; Lewandowsky, 2015). The CDU/CSU is seen as a conservative, center-right party alliance and – at the time of data collection – the chancellor, Angela Merkel, was a member of the CDU. The FDP supports economic liberal positions plus restrictive attitudes on refugee and European politics, which makes it a party voted by many individuals with higher incomes. Bündnis 90/Die Grünen was long time mostly associated with environmental politics. The SPD pronounces the values of freedom, justice, and solidarity. Finally, DIE LINKE has been described as left-wing populist party (see Bundeszentrale für politische Bildung, 2017; Expatica, 2020; Schleunes et al., 2020; Volkens et al., 2020 for information on German parties).

Associations between the Big Five personality domains and voting intentions and attitudes towards specific parties in the German context have been investigated before. One study found highest scores in Openness in individuals who would vote for Bündnis 90/Die Grünen in comparison to putative voters of the other major German parties (Sindermann et al., 2020). Lowest scores in Openness were found in individuals who stated that they would vote for the CDU/CSU; only individuals who indicated they would not vote showed lower scores. At the same time, putative voters of the CDU/CSU showed highest scores in Conscientiousness. Lowest scores in Conscientiousness were observed in individuals who stated they would vote for DIE LINKE. However, not all of the differences between groups were statistically significant (Sindermann et al., 2020). Another study found that Conscientiousness was negatively related to voting for left-from-the-center parties (specifically, the SPD and Bündnis 90/Die Grünen) and positively to voting for right-from-the-center parties (specifically, the CDU/CSU). Openness showed the opposite pattern of associations (Vecchione et al., 2011). Findings from a third study among others support a positive correlation between Openness and attitudes towards

left-from-the-center German parties; specifically, the SPD, Bündnis 90/Die Grünen, and the PDS (a precursor party of DIE LINKE). At the same time, Openness was negatively related to attitudes towards the CDU/CSU and the FDP, hence, right-from-the-center parties. Conscientiousness showed the opposite pattern of associations, except the link with attitudes towards the FDP, which was non-significant (Schoen & Schumann, 2007). The associations between party preferences and other Big Five personality domains were either mixed or non-significant in the cited studies.

The Current Study

The Big Five personality domains appear to be *descriptively* associated with non-voting versus voting and voting intentions/preferences for specific parties in the German context: while associations with non-voting versus voting are inconclusive, generally speaking, Openness seems to be associated with preferences for left-from-the-center parties and Conscientiousness seems to be associated with preferences for right-from-the-center parties. However, since effect sizes are often rather small, it remains unknown whether it is possible to *predict* one's voting intentions (non-voting vs. voting and voting for a specific party) from the Big Five personality domains. Descriptions and predictions are not the same (Mõttus et al., 2020). While descriptive modeling documents associations between variables in a given sample, predictive research constructs models with the aim to predict observations in an independent sample: that is, prediction models are "trained" and tested in independent (sub)samples. Successful predictions require that the associations are not overfitted and are sufficiently robust to generalize beyond idiosyncrasies of particular samples and their compositions (Mõttus et al., 2020; Yarkoni & Westfall, 2017). Results relying on predictive analysis might resolve some problems of previous research, such as lack of generalizability, and might help to overcome non-replicable findings. Moreover, the predictive associations between lower levels of the personality trait hierarchy and voting intentions in the German multiparty context have not yet been explored. It is possible that constructing prediction models based on Big Five personality facets and nuances will yield more accurate predictions than those based on the Big Five personality domains. However, it is important that such comparative studies account for model overfitting, because more complex models automatically account for more variance in their outcome than comparatively simpler models: training and validating models in independent samples is a particularly effective way of achieving this (Mõttus & Rozgonjuk, 2021; Seeboth & Mõttus, 2018; Yarkoni & Westfall, 2017).

For the present work, we chose to implement random forest analyses. This approach differs from, for example, different kinds of regression analyses for classification problems (e.g., binomial/binary logistic, multinomial logistic) in many ways. Of importance for the present work, the non-parametric random forest algorithm 1) does not assume linear relations between predictor and criterion variables, 2) can deal with a large number of intercorrelated predictors, 3) is robust against outliers, and 4) combinations of

predictor variables are automatically considered, thus, also combinations (i.e., statistical interactions) of variables not expected by the researchers are taken into account (Buskirk & Kolenikov, 2015; Lingjun et al., 2019; Mendez et al., 2008; Stachl et al., 2020).

In conclusion, the present study aimed at investigating the aforementioned associations by means of random forest analyses in a German sample. We had no predefined hypotheses with regards to predictions of intentions to not vote versus to vote. Regarding voting intentions for a specific party (leaning), Openness and Conscientiousness were expected to be the most important predictors. No hypotheses were built for Big Five personality facets or nuances given the lack of existing literature.

Method

Sample

By means of two online surveys, we recruited a convenience sample of N = 4,286 individuals (n = 1,972 men, n = 2,307 women, n = 7 "third gender") eligible to be included in the present analyses (see Supplementary Materials [D] for details on data collection and data cleaning).

Self-Report Measures

Big Five Inventory

The Big Five personality traits were assessed by applying the Big Five Inventory (BFI; John et al., 1991) in German language (Rammstedt & Danner, 2017). It consists of 45 items but the 45th item is unique to the German version. It was not included in our analyses based on previous studies in the German context including the original validation study of the German BFI (Rammstedt & Danner, 2017; Sindermann et al., 2020) and to enable closer comparability with studies applying the English version of the BFI. All items of the questionnaire are answered on a 5-point Likert-scale from 1 = "*very inapplicable*" to 5 = "*very applicable*". Next to the broad Big Five personality domains, two facets of each Big Five personality domain can be calculated. These facets are labeled Openness for aesthetics, Openness for ideas (Openness); Order, Self-discipline (Conscientiousness); Assertiveness, Activity (Extraversion); Altruism, Compliance (Agreeableness); Anxiety, Depression (Neuroticism). Psychometric properties are detailed in Supplementary Materials (D).

Current Voting Intentions

Individuals stated which party they would vote for if general elections were held the following Sunday. Response options were CDU/CSU (n = 586, 13.67%), SPD (n = 331, 7.72%), Bündnis 90/Die Grünen (n = 1,769, 41.27%), FDP (n = 292, 6.81%), DIE LINKE (n = 412, 9.61%), AfD (n = 187, 4.36%), "others" (n = 464, 10.83%) – indicating voting for one

of the smaller parties not currently included in the German federal parliament – and "I would not vote" (*n* = 245, 5.72%); percentages do not add up to 100% due to rounding.

Statistical Analyses

Main Analyses

The statistical software R (Version 4.1.0; R Core Team, 2018) and RStudio (Version 1.4.1106; RStudio Team, 2020) were used for all analyses. Packages used for the analyses are listed in Supplementary Materials (D). First, we calculated descriptive statistics of the BFI on domain, facet, and item level (see Supplementary Materials [D]).

Next, we trained random forest models to predict intentions 1) to not vote versus to vote and 2) to vote for specific parties (within putative voters) from either Big Five i) domains, ii) facets, or iii) items (6 models in total). More specifically, we implemented a 10-times repeated 10-fold cross-validation procedure. Thus, we split the total sample into ten different folds of equal size (and with equal distributions of the respective criterion variable) ten times; thus, a total of 100 different folds were built across repeats. Moreover, we needed to account for class imbalance in the criterion variables. Prominent methods to deal with imbalanced data are weighting and over-/under-sampling (Chen et al., 2004). We applied a weighting procedure to put heavier costs on misclassifications in the minority class and decrease prediction errors in the minority class, accordingly. Specifically, weights were set indirectly proportional to the class sizes by applying the following formula: class weight of class x = *n*(majority class) / *n*(class x) (Breiman & Cutler, n.d.). We used the default settings for the number of trees (*n* = 500), splitting rule ("gini"), and the numbers of variables randomly sampled as candidates at each split (square root of number of predictor variables) in all models.

The results across folds and repeats were averaged. More specifically, we extracted mean accuracies across folds and repeats and compared these to the No Information Rate (NIR), which is the accuracy derived from always predicting the majority class (i.e., percentage of observations in the majority class). Additionally, we computed the mean balanced accuracies ([sensitivity + specificity] / 2) from the confusion matrices derived from the 10-times repeated 10-fold cross-validations. Balanced accuracies are of importance in this specific work given the imbalanced class distributions and because the balanced accuracy weights performance of the model for each class equally. Next, we calculated the respective misclassification errors from the confusion matrices across repeated cross-validations. Finally, variable importance scores derived from the final models were extracted and are presented in Supplementary Materials (D).

The procedure was applied to predict the criterion variables – intentions to not vote versus to vote and intentions to vote for specific parties – by either Big Five personality domains, facets, or nuances in separate models.

Additional Analyses

We additionally ran the same analyses to predict intentions to vote for left- versus right-from-the-center parties (in individuals indicating that they would vote for one of the major German parties). Grouping of parties into left and right was implemented according to Volkens et al. (2020): left: DIE LINKE, SPD, Bündnis 90/Die Grünen; right: FDP, CDU/CSU, AfD. This investigation was not planned from the beginning but was data driven. Specifically, it was based on the imbalance in the aforementioned criterion variables. By splitting voting intentions into two groups with intentions to vote for a left-versus right-from-the-center party, we aimed at receiving more balanced classes and to counteract problems of majority class voting. We still applied the weighting procedure (detailed above) for consistency with other analyses.

Finally, results of binomial/binary logistic regression analyses to predict intentions to 1) not vote versus to vote, 2) vote for a specific party (one-vs.-all approach), and 3) vote for left- versus right-from-the-center parties are presented in Supplementary Materials (E) for easier interpretation of effects of single predictors.

Results

Predicting Intentions to Not Vote Versus to Vote

The mean prediction accuracies across the 10-times repeated 10-fold cross-validations to predict intentions to not vote versus to vote were 92.94% (SD = 0.01) for the models comprising Big Five personality domains, 93.97% (SD = 0.00) for the models comprising facets, and 94.25% (SD = 0.00) for the models comprising items. Thus, all mean accuracies were lower than the NIR, which was 94.28% in the total sample (thus, roughly the same in each fold).

Mean balanced accuracies across the 10-times repeated 10-fold cross-validations were 49.87%, 49.97%, and 49.98% for the models comprising Big Five personality domains, facets, or items, respectively. Thus, all mean balanced accuracies were lower than the balanced accuracy of 50.00% which would be achieved by a model always predicting the majority class.

Table 1 presents confusion matrices across 10-times repeated 10-fold cross-validations and misclassification errors for models based on Big Five personality domains, facets, or items.

Table 1

Confusion Matrices and Misclassification Errors When Predicting Intentions to Not Vote Versus to Vote From Big Five Personality Domains, Facets, or Items Across 10-Times Repeated 10-Fold Cross-Validations

	Confusion matrices and misclassification errors	
	True score	
Predicted	**Non-voting**	**Voting**
Domains		
Non-voting	30	604
Voting	2,420	39,806
Facets		
Non-voting	7	141
Voting	2,443	40,269
Items		
Non-voting	0	13
Voting	2,450	40,397
Total *N* (per class and confusion matrix)	2,450	40,410
Misclassification errors		
Misclassification errors - domains	98.78%	1.49%
Misclassification errors - facets	99.71%	0.35%
Misclassification errors - items	100.00%	0.03%

Note. The distribution of current voting intentions in the total sample was: *n* = 245 (5.72%) non-voters, *n* = 4,041 (94.28%) voters (because of the 10-times repeated 10-fold cross-validations, these numbers are multiplied by 10 in the matrices). Misclassification errors are defined as the proportion of individuals in a given class who have been incorrectly classified; misclassification errors in the class of individuals who actually intended to not vote: 1-Sensitivity; misclassification errors in the class of individuals who actually intended to vote: 1-Specificity.

Predicting Intentions to Vote for a Specific Party Within Putative Voters

Mean prediction accuracies across the 10-times repeated 10-fold cross-validations to predict intentions to vote for a specific party were 23.40% (*SD* = 0.02), 27.39% (*SD* = 0.02), and 34.63% (*SD* = 0.02) when including Big Five personality domains, facets, or items as predictors in the models, respectively. Thus, all accuracies were below the NIR of 43.78%.

Mean balanced accuracies across the 10-times repeated 10-fold cross-validations and across all classes were 51.48% for the models built based on Big Five personality domains, 51.96% for the models comprising facets, and 52.71% for the models comprising items. For each individual class and across the 10-times repeated 10-fold cross-validations, the balanced accuracies lied between 49.08% ("others") and 53.48% ("CDU/CSU"), between 48.58% ("SPD") and 54.16% ("DIE LINKE"), and between 50.20% ("FDP") and 55.74% ("CDU/CSU") for the models comprising Big Five personality domains, facets, or items,

respectively. Thus, balanced accuracies of the predictions of voting intentions for most parties as well as mean balanced accuracies across all parties exceeded the balanced accuracy of 50.00%, which would be achieved by a model always predicting the majority class. More specifically, this was true for voting intentions for all parties except the SPD and "other" parties when predictions were based on domains, for all parties except the SPD when predictions were based on facets, and for all parties when predictions were based on items.

Table 2 comprises confusion matrices and misclassification errors for predictions of intentions to vote for a specific party from either Big Five personality domains, facets, or items across repeated cross-validations.

Table 2

Confusion Matrices and Misclassification Errors When Predicting Intentions to Vote for Specific Parties From Big Five Personality Domains, Facets, or Items Across the 10-Times Repeated 10-Fold Cross-Validations

Confusion matrices and misclassification errors							
	True score						
Predicted	**DIE LINKE**	**SPD**	**Bündnis 90/ Die Grünen**	**FDP**	**CDU/CSU**	**AfD**	**"others"**
Domains							
DIE LINKE	718	269	2,370	342	613	224	724
SPD	302	335	1,958	301	601	172	525
Bündnis 90/Die Grünen	1,337	1,173	5,985	725	1,614	346	1,378
FDP	288	283	1,485	312	543	232	405
CDU/CSU	608	550	2,978	588	1,411	428	760
AfD	193	210	845	270	385	184	338
"others"	674	490	2,069	382	693	284	510
Facets							
DIE LINKE	808	351	2,282	260	438	142	623
SPD	246	119	1,183	184	389	44	340
Bündnis 90/Die Grünen	1,625	1,404	7,644	1,019	2,166	527	1,748
FDP	197	251	1,056	252	399	175	342
CDU/CSU	564	706	3,115	631	1,507	567	788
AfD	174	106	715	163	423	189	249
"others"	506	373	1,695	411	538	226	550
Items							
DIE LINKE	604	309	1,663	264	346	115	546
SPD	71	99	373	56	168	30	49
Bündnis 90/Die Grünen	2,341	1,880	10,832	1,445	2,838	669	2,500
FDP	88	84	444	99	204	114	190
CDU/CSU	524	630	2,895	701	1,722	642	798

Confusion matrices and misclassification errors

Predicted	DIE LINKE	SPD	Bündnis 90/ Die Grünen	FDP	CDU/CSU	AfD	"others"
			True score				
AfD	65	90	308	109	231	162	79
"others"	427	218	1,175	246	351	138	478
Total N (per class and confusion matrix)	4,120	3,310	17,690	2,920	5,860	1,870	4,640
Misclassification errors							
Misclassification errors - domains	82.57%	89.88%	66.17%	89.32%	75.92%	90.16%	89.01%
Misclassification errors - facets	80.39%	96.40%	56.79%	91.37%	74.28%	89.89%	88.15%
Misclassification errors - items	85.34%	97.01%	38.77%	96.61%	70.61%	91.34%	89.70%

Note. DIE LINKE = The Left; SPD = Social Democratic Party of Germany; Bündnis 90/Die Grünen = Alliance 90/The Greens; FDP = Free Democratic Party; CDU/CSU = party alliance of the Christian Democratic Union of Germany and the Christian Social Union in Bavaria; AfD = Alternative for Germany. The distribution of current voting intentions in the total sample used for these analyses was as follows: n = 412 (10.20%) DIE LINKE, n = 331 (8.19%) SPD, n = 1,769 (43.78%) Bündnis 90/Die Grünen, n = 292 (7.23%) FDP, n = 586 (14.50%) CDU/CSU, n = 187 (4.63%) AfD, n = 464 (11.48%) "others" (because of the 10-times repeated 10-fold cross-validations, these numbers are multiplied by 10 in the matrices). Individuals stating they would not vote were not included in the present analyses; percentages do not add up to 100% due to rounding. Misclassification errors are defined as the proportion of individuals in a given class who have been incorrectly classified; misclassification errors = 1-Sensitivity.

Predicting Intentions to Vote for Left- Versus Right-From-the-Center Parties Within Putative Voters of One of the Six Major German Parties

When predicting intentions to vote for left- versus right-from-the-center parties, mean prediction accuracies across 10-times repeated 10-fold cross-validations were 64.15% (*SD* = 0.02), 65.16% (*SD* = 0.02), and 68.85% (*SD* = 0.02) when using Big Five personality domains, facets, or items in the prediction models, respectively. These accuracies were all lower than the NIR of 70.23%.

Balanced accuracies across the 10-times repeated 10-fold cross-validations were 56.40% for predictions from Big Five personality domains, 55.87% for predictions from facets, and 57.70% for predictions from items. Thus, balanced accuracies of these models exceeded the balanced accuracy of 50.00% achieved by a model always predicting the majority class.

Table 3 presents confusion matrices and misclassification errors when predicting intentions to vote for left- versus right-from-the-center parties from either Big Five personality domains, facets, or items across repeated cross-validations.

Table 3

Confusion Matrices and Misclassification Errors When Predicting Intentions to Vote for Left- Versus Right-From-the-Center Parties From Big Five Personality Domains, Facets, or Items Across 10-Times Repeated 10-Fold Cross-Validations

	Confusion matrices and misclassification errors	
	True score	
Predicted	**Left**	**Right**
Domains		
Left	18,982	6,685
Right	6,138	3,965
Facets		
Left	19,804	7,146
Right	5,316	3,504
Items		
Left	21,421	7,442
Right	3,699	3,208
Total *N* (per class and confusion matrix)	25,120	10,650
Misclassification errors		
Misclassification errors - domains	24.43%	62.77%
Misclassification errors - facets	21.16%	67.10%
Misclassification errors - items	14.73%	69.88%

Note. The distribution of current voting intentions in the total sample included in these analyses was: *n* = 2,512 (70.23%) intending to vote for a left-from-the-center party, *n* = 1,065 (29.77%) intending to vote for a right-from-the-center party (because of the 10-times repeated 10-fold cross-validations, these numbers are multiplied by 10 in the matrices). Individuals stating they would vote for "another" party and individuals stating they would not vote were not included in the analyses. Misclassification errors are defined as the proportion of individuals in a given class who have been incorrectly classified; misclassification errors in the class of individuals who actually intended to vote for a left-from-the-center party: 1-Sensitivity; misclassification errors in the class of individuals who actually intended to vote for a right-from-the-center party: 1-Specificity.

Discussion

The present study aimed to investigate the predictability of voting intentions of German individuals from Big Five personality domains, facets, and nuances (indexed by individual items).

Results showed that intentions to not vote versus to vote could not be predicted better than by a baseline learner/model always predicting the majority class (i.e., "voter").

Regarding the prediction of intentions to vote for a specific party within the group of putative voters, mean prediction accuracies across repeated cross-validations ranged from 23.40% for the models comprising Big Five personality domains to 34.63% for the models comprising nuances. All of these accuracies were below the NIR. In this realm, it is important to note that accurate predictions of individuals' voting intentions in multi-party systems are difficult to achieve; this has also been observed in other studies. One study using political Facebook likes of individuals, thus, variables much closer related to voting intentions than personality traits, reports a prediction accuracy of around 60% (Kristensen et al., 2017). Nevertheless, mean balanced accuracies for predictions of intentions to vote for a specific party observed in the present work were – in parts – higher than the mean balanced accuracy of 50.00%; a balanced accuracy of 50.00% is achieved by a model always predicting the majority class. Due to the class imbalance in the present sample, balanced accuracies are of special interest because the performance of the model in predicting each class is weighted equally. Voting for the Christian religion oriented and conservative party alliance of CDU/CSU (from Big Five personality domains and nuances) and for the left-wing party DIE LINKE (from Big Five personality facets) showed highest balanced accuracies exceeding 50.00% (Bundeszentrale für politische Bildung, 2017; Volkens et al., 2020). Moreover, the mean balanced accuracies across classes and across 10-times repeated 10-fold cross-validations were exceeding 50.00% for predictions from Big Five personality domains, facets, and nuances. This indicates that voting for specific parties can – to a certain degree – be predicted by personality at different levels of the trait hierarchy when weighting the performance to predict each class equally; but not based on the overall accuracies as compared to the NIR.

For predictions of intentions to vote for left- versus right-from-the-center parties within the group of putative voters of one of the major German parties, mean prediction accuracies across repeated cross-validations ranged from 64.15% for models comprising Big Five personality domains to 68.85% for models comprising nuances. Again, however, all of these accuracies were below the NIR. But balanced accuracies ranged from 55.87% for predictions from Big Five personality facets, to 57.70% for prediction models based on nuances. Hence, the balanced accuracies were exceeding the balanced accuracy of 50.00% achieved by a model always predicting the majority class.

Regarding the aim of the present study to test whether Big Five personality facets and nuances exhibit higher predictive accuracies compared to domains, the overall answer would be: no and yes. For a more elaborate discussion on this, we compare the balanced accuracies across 10-times repeated 10-fold cross-validations of models derived from Big Five personality domains, facets, and nuances. Moreover, we focus on the prediction of voting intentions for specific parties and for left- versus right-from-the-center parties since only models to predict these variables exceeded a balanced accuracy of 50.00%.

PsychOpen GOLD

As can be seen in the results section, the mean balanced accuracies were highest for models based on Big Five personality nuances. However, the balanced accuracies for models based on Big Five personality facets were roughly the same, once even lower, compared to models based on Big Five personality domains. Thus, using nuances increased balanced accuracies of predictions in comparison to using domains and facets; but the increase was only around 1.8% at most. The finding that models based on Big Five personality facets did barely exhibit higher balanced accuracies of predictions over models based on domains might be due to the following fact: in the specific questionnaire applied in this study, not all items are included in facets; but domains and single-item analyses comprise all 44 items (see Supplementary Materials [D] for an overview on which items do and do not belong to a certain facet and a certain domain). Items not included in facets might play a role in the prediction.

Importantly, what is understood as satisfying prediction performance is to a certain degree subjective. Similarly, whether the overall accuracy or the balanced accuracy of a model is considered when evaluating prediction performance is a subjective decision. Therefore, we transparently present both accuracy measures. Nevertheless, it is clear that the predictions of intentions to not vote versus to vote based on the present models were not better than predictions of a baseline learner/model always predicting the majority class: accuracies consistently lied below the NIR and the balanced accuracies consistently lied below 50.00%. Regarding the prediction of voting intentions for a specific party and for left- versus right-from-the-center parties, the present results (regarding accuracies and balanced accuracies) indicate that additional variables need to be taken into account in order to increase the prediction performance (see further discussion below). The fact that accuracies did not exceed the NIR but balanced accuracies exceeded the 50.00% threshold indicates that the algorithms used in the present study might overall perform better in samples with more balanced class distributions.

To judge on the effect sizes found in the present study (for voting for a specific party or left- vs. right-from-the-center parties), we would like to take into account effect sizes reported in other related studies. As such, the correlation between Big Five personality domains and voting for center-right versus center-left parties in the study by Vecchione et al. (2011) ranged from |.12| to |.24| in the German sample. In a meta-analysis on associations between the Big Five personality domains and political ideology in German samples, the highest correlation was |.07| (Krieger et al., 2019). These effect sizes derived from previous studies together with the imbalanced class distributions might explain why accuracies of none of the present models did exceed the NIR (but please see interpretation of balanced accuracies). Moreover, these results indicate that personality does not seem to play a major role in predicting voting intentions. Putatively, one's voting intention is a complex psychological construct and many different variables and their interactions might contribute to it; each with a small singular effect (Götz et al., 2021). Thus, taking into account more variables to predict voting intentions, like

socio-demographics (Nier, 2017) or political attitudes such as attitudes towards refugees (Igarashi, 2021), might improve the prediction. However, for the federal elections in 2017 in Germany, nearly 62,000,000 individuals were eligible to vote (Der Bundeswahlleiter, 2021). On a population scale, therefore, even seemingly small effect sizes, such as those found in the present study, can be of importance.

The results have important implications for current debates. As such, it is often discussed in how far knowledge of individuals' personality can be used to influence their opinions, for example by microtargeting, in the political as well as economic field (Matz et al., 2017). A prominent example from the political field is the Cambridge Analytica scandal. Inference on, among others, the Big Five characteristics of citizens from their digital footprint data appears to have been used to predict and manipulate voting preferences, for example in the 2016 American presidential elections and the Brexit referendum (Wylie, 2019). Supporting evidence that personality traits can be predicted from digital footprint data to a certain degree comes from various studies (Kosinski et al., 2013; Marengo & Montag, 2020). However, with regard to political attitudes and according to the results of the present study, more variables aside from (self-reported) personality should be taken into account to improve prediction performance – at least in the German context. If voting intentions cannot be accurately predicted from personality traits, this makes it seem unlikely that purely personality-based campaigning effectively influences one's voting intentions. However, by taking into account more variables (see above) the prediction of voting intentions might be improved. In line with this, also Cambridge Analytica did putatively not use knowledge on the Big Five personality traits exclusively.

Finally, some limitations of the present study need to be mentioned. First, the sample was not representative of the general German population. The distribution of current voting intentions observed in the present sample did neither reflect the election outcome from the federal elections in 2017, nor the distributions found in general population samples in the time of data collection (Bundeszentrale für politische Bildung, 2018; Guttmann, 2020). This might limit generalizability of findings. Moreover, the imbalance in the class sizes of the criterion variables seems to have posed a problem. Next, the current sampling procedure was focused on individuals from Germany. However, we did not specifically ask whether participants were eligible to vote in the general German elections. Even if unlikely, it is possible that some of the study participants are German residents, but are not eligible to vote, for example, because they are not citizens. Moreover, the present study is limited to one self-report measure of the Big Five. We therefore call for future studies reinvestigating the present research questions using different Big Five measures to test replicability of results across different measurements. Similarly, longer questionnaires assessing more nuances and facets of the Big Five might give more detailed insights, since the BFI was not specifically designed to assess facets, never mind nuances, in-depth. In the same vein, also applying different algorithms with different tuning parameter settings might lead to (slightly) different results. Next, it is also impor-

tant to note that in general German elections, one has two votes – one to vote for a specific candidate from one's electoral district to become a member of the *Bundestag* and one for a specific party. In line with common surveys of professional research institutes on voting preferences (Willkow & Cantow, 2020), we focused on the second vote in the present study. Another limitation of the present study is that the generalizability of the findings is necessarily limited to the German context. Applicability of the findings to other countries must be tested in future studies.

Conclusion

This study sheds light on the predictability of voting intentions by Big Five personality domains, facets, and nuances in the German context. The differentiation between German individuals with the intention to not vote versus to vote was not possible via the random forest analysis approach applied in this work. The predictions of voting intentions for specific parties or for parties on the left versus right side of the spectrum was barely possible. The prediction accuracies might be improved by adding more variables to the models. Still, the accuracies of predictions from models based on Big Five personality nuances were slightly higher than predictions from models based on Big Five personality facets and domains.

PsychOpen GOLD

Supplementary Materials

For this article the following Supplementary Materials are available via PsychArchives (for access see Index of Supplementary Materials below):

> (A): File "datafile_sample_participants_rf_bf_voting.xlsx" contains data and codebook.
> (B): File "data_and_codebook.RData" contains data and codebook.
> (C): File "Rscript_rf_bf_voting.R" contains the R-code to reproduce the findings.
> (D): File "Supplementary_Material_1.pdf" contains additional information.
> (E): File "Supplementary_Material_2.xlsx" contains additional information.

Index of Supplementary Materials

Sindermann, C., Mõttus, R., Rozgonjuk, D., & Montag, C. (2021). *Supplementary materials to: Predicting current voting intentions by Big Five personality domains, facets, and nuances – A random forest analysis approach in a German sample* [Data, codebooks, R-code, additional information]. PsychOpen GOLD. https://doi.org/10.23668/psycharchives.5114

References

Berbuir, N., Lewandowsky, M., & Siri, J. (2015). The AfD and its sympathisers: Finally a right-wing populist movement in Germany? *German Politics, 24*(2), 154-178. https://doi.org/10.1080/09644008.2014.982546

Breiman, L., & Cutler, A. (n.d.). *Random forests—Classification description.* Retrieved March 30, 2021, from https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm#ooberr

Bundeszentrale für politische Bildung. (2017). *Parteien in Deutschland | Dossier Parteien | bpb.* bpb. https://www.bpb.de/politik/grundfragen/parteien-in-deutschland/

Bundeszentrale für politische Bildung. (2018). *Wählerstimmen | Zahlen und Fakten Bundestagswahl 2017.* bpb. https://www.bpb.de/nachschlagen/zahlen-und-fakten/bundestagswahlen/279740/waehlerstimmen

Buskirk, T. D., & Kolenikov, S. (2015). Finding respondents in the forest: A comparison of logistic regression and random forest models for response propensity weighting and stratification. *Survey Methods: Insights from the Field.* https://doi.org/10.13094/SMIF-2015-00003

Chen, C., Liaw, A., & Breiman, L. (2004). *Using random forest to learn imbalanced data* (Technical Report 666, p. 12). Department of Statistics, University of California.

Cooper, C. A., Golden, L., & Socha, A. (2013). The Big Five personality factors and mass politics. *Journal of Applied Social Psychology, 43*(1), 68-82. https://doi.org/10.1111/j.1559-1816.2012.00982.x

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual.* Psychological Assessment Resources.

Dawkins, R. (2017). Political participation, personality, and the conditional effect of campaign mobilization. *Electoral Studies, 45*, 100-109. https://doi.org/10.1016/j.electstud.2016.11.018

Der Bundeswahlleiter. . (2021). *Wahlberechtigte.* bundeswahlleiter. https://www.bundeswahlleiter.de

PsychOpen GOLD

/service/glossar/w/wahlberechtigte.html

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology, 93*(5), 880-896. https://doi.org/10.1037/0022-3514.93.5.880

Expatica. (2020). *The main political parties in Germany.* Expatica. https://www.expatica.com /de/living/gov-law-admin/political-parties-in-germany-107953/

Gerber, A. S., Huber, G. A., Doherty, D., & Dowling, C. M. (2011). The Big Five personality traits in the political arena. *Annual Review of Political Science, 14*(1), 265-287. https://doi.org/10.1146/annurev-polisci-051010-111659

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 59*(6), 1216-1229. https://doi.org/10.1037/0022-3514.59.6.1216

Götz, F., Gosling, S., & Rentfrow, J. (2021). *Small effects: The indispensable foundation for a cumulative psychological science.* PsyArXiv. https://psyarxiv.com/hzrxf/

Guttmann, P. (2020, September 8). *Bundestagswahl: Neueste Wahlumfragen im Wahltrend | Sonntagsfrage #btw.* Chronik aller Wahlumfragen zur Bundestagswahl. https://dawum.de /Bundestag/32

Hirsh, J. B., Deyoung, C. G., Xu, X., & Peterson, J. B. (2010). Compassionate liberals and polite conservatives: Associations of Agreeableness with political ideology and moral values. *Personality and Social Psychology Bulletin, 36*(5), 655-664. https://doi.org/10.1177/0146167210366854

Igarashi, A. (2021). Hate begets hate: Anti-refugee violence increases anti-refugee attitudes in Germany. *Ethnic and Racial Studies, 44*(11), 1914-1934. https://doi.org/10.1080/01419870.2020.1802499

Jang, K. L., McCrae, R. R., Angleitner, A., Riemann, R., & Livesley, W. J. (1998). Heritability of facet-level traits in a cross-cultural twin sample: Support for a hierarchical model of personality. *Journal of Personality and Social Psychology, 74*(6), 1556-1565. https://doi.org/10.1037/0022-3514.74.6.1556

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big-Five Inventory-version 4a and 54.* Berkeley Institute of Personality and Social Research, University of California.

Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives. *The Journal of Applied Psychology, 98*(6), 875-925. https://doi.org/10.1037/a0033901

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America, 110*(15), 5802-5805. https://doi.org/10.1073/pnas.1218772110

Krieger, F., Becker, N., Greiff, S., & Spinath, F. M. (2019). Big-Five personality and political orientation: Results from four panel studies with representative German samples. *Journal of Research in Personality, 80*, 78-83. https://doi.org/10.1016/j.jrp.2019.04.012

PsychOpen GOLD

Kristensen, J. B., Albrechtsen, T., Dahl-Nielsen, E., Jensen, M., Skovrind, M., & Bornakke, T. (2017). Parsimonious data: How a single Facebook like predicts voting behavior in multiparty systems. *PLOS ONE, 12*(9), Article e0184562. https://doi.org/10.1371/journal.pone.0184562

Lewandowsky, M. (2015). Eine rechtspopulistische Protestpartei? Die AfD in der öffentlichen und politikwissenschaftlichen Debatte. *Zeitschrift fur Politikwissenschaft, 25*(1), 119-134. https://doi.org/10.5771/1430-6387-2015-1-119

Lingjun, H., Levine, R., Fan, J., Beemer, J., & Stronach, J. (2019). Random Forest as a predictive analytics alternative to regression in institutional research. *Practical Assessment, Research & Evaluation, 23*, Article 1. https://doi.org/10.7275/1wpr-m024

Marengo, D., & Montag, C. (2020). Digital phenotyping of Big Five Personality via Facebook data mining: A meta-analysis. *Digital Psychology, 1*(1), 52-64. https://doi.org/10.24989/dp.v1i1.1823

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences of the United States of America, 114*(48), 12714-12719. https://doi.org/10.1073/pnas.1710966114

McCrae, R. R. (2015). A more nuanced view of reliability: Specificity in the trait hierarchy. *Personality and Social Psychology Review, 19*(2), 97-112. https://doi.org/10.1177/1088868314541857

McCrae, R. R., & Mõttus, R. (2019). What personality scales measure: A new psychometrics and its implications for theory and assessment. *Current Directions in Psychological Science, 28*(4), 415-420. https://doi.org/10.1177/0963721419849559

Mendez, G., Buskirk, T. D., Lohr, S., & Haag, S. (2008). Factors associated with persistence in science and engineering majors: An exploratory study using classification trees and random forests. *Journal of Engineering Education, 97*(1), 57-70. https://doi.org/10.1002/j.2168-9830.2008.tb00954.x

Mondak, J. J., & Halperin, K. D. (2008). A framework for the study of personality and political behaviour. *British Journal of Political Science, 38*(2), 335-362. https://doi.org/10.1017/S0007123408000173

Mondak, J. J., Hibbing, M. V., Canache, D., Seligson, M. A., & Anderson, M. R. (2010). Personality and civic engagement: An integrative framework for the study of trait effects on political behavior. *The American Political Science Review, 104*(1), 85-110. https://doi.org/10.1017/S0003055409990359

Mõttus, R., Bates, T., Condon, D. M., Mroczek, D., & Revelle, W. (2017). *Leveraging a more nuanced view of personality: Narrow characteristics predict and explain variance in life outcomes.* PsyArXiv. https://psyarxiv.com/4q9gv/

Mõttus, R., Kandler, C., Bleidorn, W., Riemann, R., & McCrae, R. R. (2017). Personality traits below facets: The consensual validity, longitudinal stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology, 112*(3), 474-490. https://doi.org/10.1037/pspp0000100

Mõttus, R., & Rozgonjuk, D. (2021). Development is in the details: Age differences in the Big Five domains, facets, and nuances. *Journal of Personality and Social Psychology, 120*(4), 1035-1048. https://doi.org/10.1037/pspp0000276

Mõttus, R., Wood, D., Condon, D. M., Back, M., Baumert, A., Costantini, G., Epskamp, S., Greiff, S., Johnson, W., Lukaszewski, A., Murray, A., Revelle, W., Wright, A., Yarkoni, T., Ziegler, M., & Zimmermann, J. (2020). Descriptive, predictive and explanatory personality research: Different goals, different approaches, but a shared need to move beyond the Big Few traits. *European Journal of Personality, 34*(6), 1175-1201. https://doi.org/10.1002/per.2311

Nier, H. (2017, August 30). *Infografik: Wer wählt wen?* Statista Infografiken. https://de.statista.com/infografik/10882/parteipraeferenz-und-soziostrukturellen-merkmale-der-wahlberechtigten-in-deutschland/

Paunonen, S. V., & Ashton, M. C. (2001). Big five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology, 81*(3), 524-539. https://doi.org/10.1037/0022-3514.81.3.524

Rammstedt, B., & Danner, D. (2017). Die Facettenstruktur des Big Five Inventory (BFI). *Diagnostica, 63*(1), 70-84. https://doi.org/10.1026/0012-1924/a000161

R Core Team. (2018). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

RStudio Team. (2020). *RStudio: Integrated development for R.* RStudio. http://www.rstudio.com/

Schleunes, K. A., Turner, H. A., Barkin, K., Bayley, C. C., Duggan, L. G., Berentsen, W. H., Kirby, G. H., Elkins, T. H., Geary, P. J., Sheehan, J. J., Heather, P. J., Leyser, K. J., Wallace-Hadrill, J. M., Hamerow, T. S., & Strauss, G. (2020). *Germany—Political parties.* Encyclopedia Britannica. https://www.britannica.com/place/Germany/Political-parties

Schoen, H., & Schumann, S. (2007). Personality traits, partisan attitudes, and voting behavior. Evidence from Germany. *Political Psychology, 28*(4), 471-498. https://doi.org/10.1111/j.1467-9221.2007.00582.x

Seeboth, A., & Mõttus, R. (2018). Successful explanations start with accurate descriptions: Questionnaire items as personality markers for more accurate predictions. *European Journal of Personality, 32*(3), 186-201. https://doi.org/10.1002/per.2147

Sibley, C. G., Osborne, D., & Duckitt, J. (2012). Personality and political orientation: Meta-analysis and test of a Threat-Constraint Model. *Journal of Research in Personality, 46*(6), 664-677. https://doi.org/10.1016/j.jrp.2012.08.002

Sindermann, C., Elhai, J. D., Moshagen, M., & Montag, C. (2020). Age, gender, personality, ideological attitudes and individual differences in a person's news spectrum: How many and who might be prone to "filter bubbles" and "echo chambers" online? *Heliyon, 6*(1), Article e03214. https://doi.org/10.1016/j.heliyon.2020.e03214

Stachl, C., Pargent, F., Hilbert, S., Harari, G. M., Schoedel, R., Vaid, S., Gosling, S. D., & Bühner, M. (2020). Personality research and assessment in the era of machine learning. *European Journal of Personality, 34*(5), 613-631. https://doi.org/10.1002/per.2257

PsychOpen GOLD

Tupes, E. C., & Christal, R. E. (1992). Recurrent personality factors based on trait ratings. *Journal of Personality, 60*(2), 225-251. https://doi.org/10.1111/j.1467-6494.1992.tb00973.x

Vecchione, M., Schoen, H., Castro, J. L. G., Cieciuch, J., Pavlopoulos, V., & Caprara, G. V. (2011). Personality correlates of party preference: The Big Five in five big European countries. *Personality and Individual Differences, 51*(6), 737-742. https://doi.org/10.1016/j.paid.2011.06.015

Volkens, A., Burst, T., Krause, W., Lehmann, P., Matthieß, T., Merz, N., Regel, S., Weßels, B., & Zehnter, L. (2020). *The manifesto data collection: Manifesto project* (MRG/CMP/MARPOR, Version 2020b). Berlin: Wissenschaftszentrum Berlin Für Sozialforschung.

Willkow, Z., & Cantow, M. (2020). *Wahlumfragen zur Bundestagswahl.* wahlrecht. https://www.wahlrecht.de/umfragen/index.htm

Wylie, C. (2019). *Mindf*ck: Cambridge Analytica and the Plot to Break America.* Random House Audio.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*(6), 1100-1122. https://doi.org/10.1177/1745691617693393