Modelling proximity and directional logic in VAAs

Fernando Mendez*
Centre for Democracy Studies Aarau
University of Zurich

Paper presented at ECPR
Glasgow, 5-7 September, 2014

Abstract

Matching algorithms based on competing theoretical models of issue voting (proximity and directional theory) are presented with a view to evaluating their predictive performance in diverse cross-national settings. Using methods from computer science and data mining, both high dimensional models as well as low dimensional models are tested. A core concern is whether voters' decisional logic is conditioned by contextual factors such as centripetal versus centrifugal party competition. The paper builds on recent literature that argues proximity models perform best in less polarised setting whereas directional models work better in a more polarised context, and shows how VAA–generated data can be used for evaluating the performance of competing theories of issue voting. Furthermore, it argues that engaging in such analyses can also lead to theoretically grounded improvements in the algorithms deployed by VAA designers.

Keywords: Voting advice applications — Political behaviour — Directional and Proximity theory — Predictive modeling

1 Introduction

This paper evaluates the predictive performance of matching algorithms based on competing theoretical models of issue voting. The theoretical inspiration comes from a set of practical concerns about how best to aggregate political preferences across a range of policy issues in order to match voters with political parties based on the similarity of their policy preferences. Framed in such terms, the research problem would appear to be mainly directed at the narrow community of scholars working on methodological aspects of VAA design. Yet, this would be mistaken. To the extent that any

*Email: fernando.mendez@zda.uzh.ch

matching of citizens to political parties takes places, the algorithms used will be based on an established body of theory in political science. Although this may not always be explicitly acknowledged VAA by designers, it is the case at least implicitly. Thus, as VAAs further proliferate and a growing body of evidence begins to accumulate suggesting that they have direct effects on citizens' behaviour this linkage between narrower VAA design issues and the broader field of political science will become ever more important.

One of the broader objectives of this paper is to show how research motivated by a concern surrounding methodological issues in VAA design is relevant to the broader field of political science. This comes in a number of guises, two of which will inform this paper: (1) first, is the contention that datasets generated by VAAs can be an extremely fruitful data source for examining non–VAA centric concerns; (2) second, as VAAs become more popular, VAA research is unlikely to remain restricted to political behaviour scholars. Instead it is likely to become more interdisciplinary as appears to already be the case. Such cross–fertilization from other disciplines can also feed into the wider political science literature. The next section of this paper addresses point (1) and describes the theoretical and empirical motivation behind this paper, namely to evaluate two competing theories of issue voting drawing on VAA–generated data. Section three focusses on the methodology and evaluation criteria underpinning the statistical analyses. Here I outline some of the techniques borrowed from the fields of computer science and data mining, in particular, the application of machine learning algorithms to model voter–party choice. This connects with the aim to further the interdisciplinary linkages mentioned in point (2) above. Section four presents the datasets used and the criteria for case and party selection as well as data cleaning methods. In section 5 the results of the empirical analysis are presented in terms of high and low dimensional predictive modeling. The discussion in the concluding section then relates the findings back to the wider literature as well as a VAA design issues.

2 Conceptulisation

As VAAs have proliferated and become a more permanent feature of the political landscape, so too has the literature on VAAs (see Triga et al. 2012; and the most recent review Garzia and Marschall 2014). This literature increasingly touches on many dimensions of the VAA phenomenon from the normative aspects of such a potentially significant intervention into the electoral process (Anderson and Fossen 2014), to the effects of its usage on aspects of individual political behaviour (Garzia et al. 2014, Vassil 2012). There is also a growing literature on methodological issues related to how VAAs are designed. In particular, a number of methodological concerns have emerged among political scientists such as (1) statement formu-

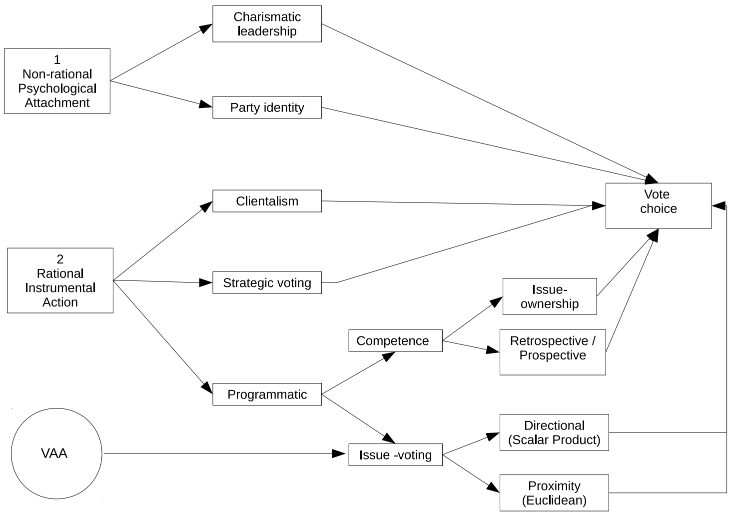

Figure 1: Theories of voter choice

lation(Gemenis 2013, Van Camp et al. 2014); (2) the effects of statement selection (Walgrave et al 2009), (3) the reliability of party coding (Gemenis 2013, Gemenis and Van Ham 2014), (4) the effects of different spatial frameworks (Louwerse and Otjes 2012; Louwerse and Rosema 2014, Germann et al 2014, Wheatley 2012; Wheatley et al. 2013; Mendez and Wheatley 2014); (5) the configuration of answer categories (Baka et al 2012); and (6) the choice of matching algorithms (Mendez 2012; 2014). This paper is mostly connected to this last strand. Given the central importance of an algorithm for the core function of a VAA, which is to match respondents to parties, there has been surprisingly little attention devoted to the issue from a political science perspective. On the other hand, 'interdisciplinary' approaches to VAA algorithms and matching functions are becoming more prevalent. From a computer science perspective there have been a number of papers, for instance, on the effect of missing values in vote recommendation (Agathokleous 2013) as well as alternative ways to match voters to parties using the logic of recommendation systems (Katakis et al. 2012, Teran et al. 2012) and in particular the development of a 'social vote recommendation' system (Katakis et al 2013). This paper builds on this strand of research in the emerging literature with a view to further explore the intersection between political science and computer science. This aspect is more fully addressed in the next section. For now, we turn to some of the political science theories that can be used to model voter–party ideological congruence.

The first step is to identify the potential repertoire of models a VAA could draw on from the various theories of voting behaviour within the broader field. Figure 1 provides this schematic overview to situate the VAA within a specific cluster of voting theories. Given the very function of a VAA is to match voters with the closest party in terms of the policy issue space, it is clear that a VAA falls within box 2 in figure 1, the rationalistic model, rather than say the more sociological—emotional theories of box 1 that are based on charismatic leadership or party identification. Furthermore, within the rationalistic models emanating from box 2, the VAA fits within the issue–voting box. This is why VAAs include a battery of policy items for the matching function rather than, say, perceptions of party leaders' competence to govern. The dashed line in the figure conveys the fact that almost VAAs draw their matching function from the issue voting model, and within this model mostly from proximity theory. By 'matching function' I refer to the metric used to measure voter–party congruence. Although it may be possible to incorporate other elements beyond policy issues into a matching function, I am not aware of this having been implemented. On the other hand, there have been VAAs from the PreferenceMatcher family that have drawn on the competing issue voting model, directional theory, to provide more than one matching metric to respondents. This was the case for the EUvox VAA, which has generated the datasets used in this paper.

Although the two issue voting models in figure 1 are based on a similar world–view there has been a long–standing controversy around these competing models of vote choice. In the one corner is a tradition that traces its lineage back to Downs' (1957) economic theory of democracy and the spatial models it has engendered. This approach, the proximity model, assumes a centripetal dynamic whereby voters prefer a party that is closest to them. On the other side, is a more recent theory that is based on a centrifugal logic that leads voters to choose a party according to the direction and the intensity of their issue preference (Macdonald et al. 1991, 1995). Not surprisingly, there are those who find mixed evidence for both explanations (Lewis and King 1999; Cho and Endersby 2003) while others have tried to combine elements of both (Merrill and Grofman 1999). In a way, this latter approach is partly adopted in this paper by also testing a model that draws on elements of both theories –although it is closest to the directional model. This draws on some earlier work with VAA algorithms (Mendez 2012, 2014) which presents the algorithms used for matching voters to parties according to proximity, directional and a directionally–inspired hybrid model (the corresponding matrices are reproduced in the annex 1). This type of analysis is based on high dimensional matching, which simply means that all VAA policy items are used to perform the matching –the precise dimensionality of the policy space being determined by the number of items in a VAA policy items (for a more detailed discussion of the distinction between the two

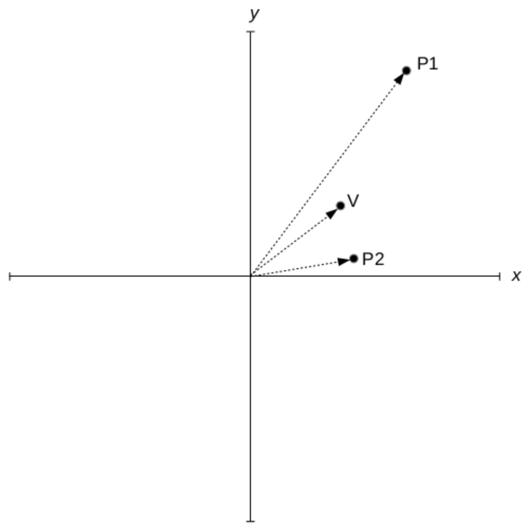

Figure 2: Voter choice in a two–dimensional space

types see Mendez 2014).[1]

In low dimensional matching, subsets of VAA policy items are grouped to form scales -typically a two dimensional scale such as a left/right vs. conservative/liberal– that is then used to map users and parties. Since the logic of high dimensional matching has been described in Mendez (2012; 2014) and the matrices for constructing the algorithms are included in a technical note in annex 1, I will focus my attention on low dimensional matching to illustrate the differences between proximity and directional models. The key distinction is between an voter's 'ideal position' and 'ideal direction'. Figure 2 tries to capture this distinction at its most basic level (this depiction draws on the basic intuitions in Linhart & Shikano's (2009) generalization of a directional model). The model assumes there is an 'origin' or 'status quo' position (the zero point in the coordinate system). Two competing predictions about voter V's preferred party, $P1$ or $P2$, are made by the theories. The proximity model seeks to minimise the distance between V and $P1$ or $P2$ in figure 2 –hence its proximity name and use of a Euclidean metric. On the other hand, the directional model aims to capture a voter's ideal policy direction. Here the intuition is that actors are less concerned with specific ideal points but want to shift policy in particular direction. In the graph the direction is depicted by the dashed arrow lines, which correspond

[1]The high dimensional match refers to the aggregation based on all policy items, which is reduced to a single coefficient. Paradoxically, then, a high dimensional match is reduce to a single coefficient whereas as low dimensional match may actually consist of two or more dimensions (see next section).

to the vectors from the origin of V, $P1$ and $P2$. Under directional logic the aim is to minimise not the distance between V and $P1$ or $P2$ but rather to minimise the angle between the corresponding vectors.

More formally to assign a preferred party let the voter's position be defined by V (x_v,y_v) and let $\vec{v}$ represent the vector defining his/her position with respect to the origin. Similarly let $P1$ (x_1,y_1) and $P2$ (x_2,y_2) define the positions of parties $P1$ and $P2$ respectively and let $\vec{p_1}$ and $\vec{p_2}$ define the corresponding vectors (see Figure 2). According to the proximity model, the distances d_1 and d_2 between the voter V, on the one hand, and $P1$ and $P2$, on the other, is defined by:

$$d_1 = \sqrt{(x_1 - x_v)^2 + (y_1 - y_v)^2} \tag{1}$$

$$d_2 = \sqrt{(x_2 - x_v)^2 + (y_2 - y_v)^2} \tag{2}$$

Looking at Figure 2, it is very clear that according to the proximity model voter V is closer to $P2$ than $P1$ (i.e., $d_2 < d_1$).

However, for the directional model, we focus not on the distance between V and $P1$ or V and $P2$, but instead on the angles θ_1 and θ_2 that the vectors $\vec{p_1}$ and $\vec{p_2}$ make with $\vec{v}$. This is defined by an inverse cosine function (arccos) in the equations below:

$$\theta_1 = \arccos\left(\frac{\vec{v}\cdot\vec{p_1}}{\sqrt{x_v^2 + y_v^2}\cdot\sqrt{x_1^2 + y_1^2}}\right) \tag{3}$$

$$\theta_2 = \arccos\left(\frac{\vec{v}\cdot\vec{p_2}}{\sqrt{x_v^2 + y_v^2}\cdot\sqrt{x_2^2 + y_2^2}}\right) \tag{4}$$

Turning once again to Figure 2, it is clear that the angle between the dashed lines representing V and $P1$ is smaller than between V and $P2$ (i.e., $\theta_1 < \theta_2$) and that voter V is therefore more congruent with $P1$ under the directional model.

Although our focus was on a simple two dimensional model for illustration purposes, the model can be easily extended into an n dimensional issue space (Linhart and Shikano 2009). This is important since we will be testing the performance of the two issue voting models in predicting vote choice in a three dimensional space as used by the EUvox VAA. The details surrounding the conceptualisation of political dimensionality and its operationalization are discussed in section 3.

Thus far we have only touched upon the underlying logic among the competing issue voting models. The more recent literature has added intervening or contextual factors that may impinge upon vote choice. The basic intuition is that rather than attempt to assign victory to one model over the other a more fruitful path is to investigate systemic features of the institutional context that may lead one model to perform better than the other. Kedar (2006), for instance, has stressed party–specific characteristics and institutional context in her study of issue–voting. More recently, Pardos-Prado and Dinas (2010) –hereafter PPD– have re–introduced Sartori's ideas on political polarization to bring institutional heterogeneity to the fore in the proximity–directional debate. Crucially, they distinguish between between two sides of polarization: a demand side (citizens) and the supply side (parties). It leads PPD to hypothesise the following: in systems where there is a major level of polarisation (where the party system is polarised and there is ideological polarisation among the electorate) the directional model will tend to have higher predictive weight; conversely, in those systems where party polarization is weaker and the majority of voters cluster around the centre, the proximity model will perform better in accounting for party choice (Pardos–Prado and Dinas 2010). PPD tested their theoretical model on data from European Parliament elections and derived a four–fold typology that is partly reproduced in figure 3. I draw on the PPD typology to generate some expectations about the institutional context in which proximity logic may outperform directional modeling and vice versa in terms of accounting for vote choice. The details informing case selection are discussed in greater detail in section 4. For now, figure 3 shows where the five countries used in this paper are placed in the PPD model. Accordingly, there are three cases where we could expect proximity models to work best, England, Ireland and Spain and one case where directional model should perform better Greece. Since Poland is in the 'high' ideological polarisation cluster we could also expect directional theory to perform better despite it being more of a 'mixed' case.

3 Methodology for modeling party choice

The aim of this paper is to build models based on proximity and directional voting logic with a view to evaluating their predictive performance. In terms of the predictions, we are here concerned with a theoretical model's ability to correctly predict a respondent's vote intention. Furthermore, we shall be testing the models at two levels of dimensionality: a high dimensional policy space based on all thirty policy items included in the typical VAA and a low dimensional policy space based on three scales derived the EUvox VAA.

Developing a discrimination index

We begin by presenting a discrimination index that can be applied to high dimensional models of party choice based on the entire range of policy ques-

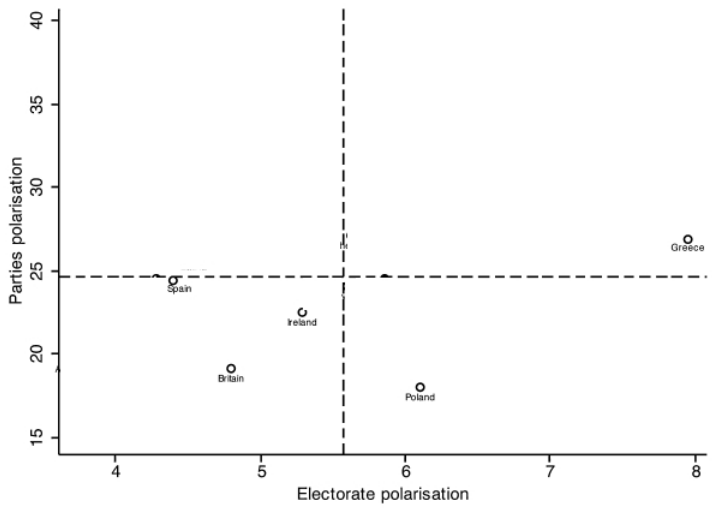

Figure 3: Country placements on polarisation scales based on Pardos–Prado and Dinas and Pardo (2010: 779)

tions included in the EUvox questionnaire.[2]. For the high dimensional modeling we can test the three algorithms that draw on proximity/directional theoretical models are described in annex 1. They are based on a city block metric (sometimes called the Manhattan distance) for modeling proximity theory and a scalar product metric for the directional model. A hybrid model is also introduced, which splits the difference between the two but mostly draws on a directional logic by giving more weight to the intensity of preferences. The proximity and hybrid algorithms were used in the live version of the EUvox (for a more detailed description of the logic see Mendez (2012; 2014)). However, before discussing some of the specific modeling techniques it is necessary to mention some of the data pre–processing that is carried out. In particular, the party matching coefficients that are derived from the three issue voting models need to be rescaled for performing the analyses. The matrices in the appendix produce coefficients that range from -1 to 1, which are rescaled to a -100 to 100 range for the online VAA. The idea is to provide respondents with 'negative' matches as well as 'positive' ones. However, even though the rank ordering of parties may be similar across the algorithms the actual coefficients will differ. Because of the way it is calculated the proximity model produces on average higher coefficients whereas the directionally–inspired models will tend to generate lower average coefficients (see matrices in appendix). Thus, in order to be able to compare the models it is necessary to normalize the coefficients. In this case, they are normalized to range from 0 to 100 for each algorithm so that the

[2]The index could also be applied to lower dimensional models, but here we are mostly concerned with presenting its logic

high–dimensional models can be directly compared.

Once each respondent's party match is calculated, according to the three different issue voting models, a number of established performance metrics can be used to evaluate the competing models. Our first performance metric attempts to measure the discriminating power of an algorithm. In broad terms, the discriminating power of an issue voting model as I define it in this paper concerns the extent to which its algorithm discriminates between, or separates, respondents with a vote intention for say, party k, from other parties by ranking party k higher than all other party alternatives. This information can be condensed into a single coefficient for each respondent with a stated vote intention. Let P_{c1} be the highest party coefficient for respondent j, P_{c2} the second highest coefficient, and so on through to...P_{cn}. Furthermore, let P_{ck} represent the coefficient of respondent j's party choice as revealed by their stated vote intention. The index is calculated as follows, where D_{ij} refers to the discrimination index for respondent j :

$$if(P_{c1} = P_{ck}), thenD_{ij} = P_{c1} - P_{c2}, elseD_{ij} = P_{ck} - P_{c1} \tag{5}$$

The higher the score, the better the performance of an algorithm.[3] Positive coefficient values are generated when the *if* condition in the formula above holds and indicate that an algorithm is able to discriminate between a respondent's preferred party and all other party alternatives. Negative coefficients are produced where the conditional statement does not hold; it indicates that an alternative party to a respondent's party choice is highest ranked. Thus, the lower the negative values the less the discrimination power of the algorithm. It is straightforward to then calculate a mean party discrimination index that can be averaged across all parties for a given country –this is especially important since the distribution of respondents' choice of parties is likely to be unbalanced.

Using supervised machine learning techniques to model voter-party ideological congruence

In a second step I use rather different methodology that draws on techniques commonly used in computer science and data mining, which fall under the rubric of *supervised machine learning.* These techniques will be applied to both high dimensional models and low dimensional models of voter-party choice. The aim is to build a predictive model using the party matching coefficients (i.e., those derived from the competing issue voting models) as predictors of party choice. However, unlike more traditional approaches in

[3]The index has the additional property that it can easily deal with tied observations. For example a tied observation with two party in first place with one being the respondent's party choice would lead to a '0' value after both scores are subtracted.

political science where the entire dataset is used to estimate model parameters, in machine learning the dataset is typically partitioned into a training set and a test set (see Hastie et al. 2009 for theoretical description of predictive modeling and Kuhn and Johnson 2013 for practical applications). After partitioning the dataset the model is built on the training set, which is used to estimate model parameters. The efficacy of the model built on the training set is then independently evaluated on the 'unseen' test set (aka the validation set). Crucially, the validation is performed only *once* on the test set. The idea is to prevent over–fitting where the model picks up on trends or quirks from a dataset that do not generalise to new 'unseen' samples. Hence the partitioning of the dataset into an 'unseen' test set from the training set. Furthermore, it is important that the validation is performed only *once* on the test set because repeated 'peeks' at the test set will also lead to over–fitting.

The logic of validation is of course not in any way exclusive to machine learning and there are good examples from political science. I use as an illustration a couple of examples with some relevance to the focus of this paper. Take for instance the seminal work by Laver et al. (2003) on wordscores for estimating policy dimensions. Their aim is to estimate latent policy positions based on a computational text analysis of mostly manifesto text (but it can be extended to any political text). Although they do not use the terminology of supervised machine learning, which can be thought of as a confirmatory method, they are engaged in the same exercise. Specfically, they train a 'reference text' (the training set) to classify what they call 'virgin' text (i.e., the test set) with a view to estimating classification performance. Similarly, Diermeier et al. (2012) use congressional speeches to estimate ideological positions of Senators using supervised machine learning algorithms -namely support vector machines- which they argue compare favourably to Poole and Rosenthal's (1985) *Nominate* approach based on voting behaviour. What all of these examples illustrate is the application of techniques from computer science and artifical intelligence to topics of interest for political scientists, which can best be thought of as 'computational political science'. This is trend is likely to increase but should also be distinguished from enhanced statistical analysis such as statistical modeling and regression analysis that is of course ubiquitous across the social sciences.

The 'computational political science' aspect to this paper is to build competing issue voting models, which are called the 'classifiers' in machine learning terms, and evaluate their classification performance. Furthermore, given that we have more than two parties for each country we are confronted with a multiclass classification problem that involves classifying instances into more than two classes. The statistical analysis is carried out in the R environment using the *caret* package for running the machine learning algorithms (Kuhn 2008). Each of the five country datasets is partitioned using a 70:30 split, a ratio that is quite common for predictive modeling. 70 per cent of the dataset is then used for training the model and the remaining 30 per cent is used as the test set to asses model performance. The samples for the training/test sets are randomly selected from the entire dataset so that the distribution of classes (parties) are preserved. I use *caret's* default parameters to tune the model in the training set, which is done by resampling the training set with 25 bootstrap iterations. Given the nature of the classification problem, a multinomial logistic regression is used. Although it would be possible to use alternative and commonly used machine learning algorithms such as tree classifiers, naive bayes or support vector machines, I rely here on a multinomial model that is more familiar to political scientists. I use the multinom function from the *nnet* package to estimate a multinomial logistic regression model based on neural networks.

We now turn to the issue of how to evaluate the performance of our predictive modeling. From a machine learning perspective common metrics for evaluating the performance of classifiers, which in this case refers to the theoretical models of issue voting, are: recall (R), precision (P) and accuracy (Acc). To see how these performance metrics are calculated let TP = number of true positives, FP = number of false positives, TN = the number of true negatives and FN= number of false negatives in the observed versus predicted classification table below. The formulae below the matrix, define how the performance metrics are calculated.

	Observed	
Predicted	TP	FP
	FN	TN

$$R = \frac{TP}{TP + FN} \tag{6}$$

$$P = \frac{TP}{TP + FP} \tag{7}$$

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \tag{8}$$

Recall is simply the 'hit rate' where a high recall means that an algorithm matched most respondents with their stated party choice, whereas a high precision means that an algorithm returned substantially more relevant party results than irrelevant ones. Finally, accuracy can be used on an overall basis for the entire country dataset to measure the proportion of correct party matches and null matches. Although the semantics across the disciplines differ, accuracy can be thought of as 'validity' and precision is synonymous with 'reliability' as used more commonly in political science or psychometrics.

$$K = \frac{O - E}{1 - E} \tag{9}$$

The three metrics are typically reported to evaluate the performance of classifiers. In particular, accuracy is usually reported as an overall measure for comparing multi–class classification models. However, one problem with overall accuracy occurs when the distribution of classes (in our case parties) is very unbalanced. This especially applies to the case of VAA datasets where party distributions based on respondents' party choice will inevitably be highly skewed. To control for unbalanced classes, the kappa statistic is used as a measure of performance. It is a metric that compares the observed accuracy with what could be achieved by random chance, which generally means that it is less misleading than simply using accuracy as a metric.

The kappa statistic, K in equation 9, takes into account the expected error rate where O is the observed accuracy and E is the expected accuracy under chance agreement. The kappa statistic is more familiar in political science since it is often used as a measure of inter-coder reliability. However, unlike comparing how two or more humans coded some (easily) observable phenomena as might be more usual in political science, in machine learning we are measuring how closely the instances classified by the machine learning classifier matched the unseen dataset, controlling for classifications that could have been produced by random chance. Thus, not only can this kappa statistic shed light into how the classifier itself performed, but it can also be directly compared with other models used for the same classification task. In terms of interpreting the kappa statistic, there is not a standardized interpretation although values <0 signal no agreement. Landis and Koch (1977) consider 0–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1 as almost perfect. Whereas according to Fleiss's (1981) kappa's >0.75 can be considered as excellent, 0.40 to 0.75 as fair to good, and <0.40 as poor. We report all the evaluation metrics in the results section, but focus most attention on the kappa statistic.

Modeling the policy space using 'objective' and 'subjective' evaluations

We have already covered most aspects of the high dimensional modeling in presenting the computation of the discrimination index so our attention here will be mainly focussed on further describing methodology for low dimensional modeling. For high dimensional modeling using supervised machine learning the same predictors as used for measuring the discrimination index, namely the party matching coefficients derived from the three competing issue voting models, will also be used for the predictive modeling. However, for the low dimensional analysis since there is no hybrid model we only compare a proximity model with its 'pure' directional counterpart as depicted in figure 2. For low dimensional modeling the supervised machine learning analysis is performed on two levels. In a first step we take the three overarching dimensions of political competition according to which the EUvox was structured as the main feature elements. The EUvox used three ex–ante defined dimensions to place respondents on the typical spatial maps included in VAAs: (1) economic (left vs. right), (2) cultural (socially conservative versus socially liberal) and (3) European (more vs less European integration). The thirty policy statements were structured so as to tap these three dimensions, with roughly ten items per dimension.[4] Evidently, the question of whether the political space is actually structured around these three or, more plausibly, less than three dimensions, is an empirical question. There is large literature on political dimensionality (Hooghe et al. 2002; Kriesi et al. 2006; Benoit and Laver 2012) and within it a strand that argues in favour of a three dimensional understanding of the European political space (Bakker et al. 2012; Costello et al. 2012).

The EUvox used some of the scaling techniques applied to VAA datasets as described in recent literature (see Wheatley 2012; Wheatley et al. 2013; Mendez and Wheatley 2014). In particular, a Mokken based 'dynamic' scale validation of the ex–ante defined dimensions was conducted on the live EUvox and where scales were found deficient the latter were modified (the method is described in Germann et al., 2014). In practice, this meant some minor adjustments to the three ex–ante defined scales, which usually involving the omission of an item or two that failed to load on a prescribed scale. For all five cases, scales for each of the three dimensions could be identified that satisfied the Mokken scaling benchmarks as described in the literature (e.g. Germann et al. 2014). For the low dimensional modeling I take the validated scales (rather than the ex–ante defined scales) in order to calculate respondents positions on the three dimensions. The parties are also mapped on the same three dimensional space. It is then straightforward to calculate the distance between respondents and parties using a Euclidean metric as defined in section 2 for a three-dimensional space. An analogous logic is used for modeling the competing directional theory, but here we use an angle minimising function to match a respondent to their closest party as described in equations 3 and 4. The resulting coefficients are then used as predictors of party choice. Thus, in this first step we are actually mimicking the VAA procedure, which is based on the match between the academically coded position of a party and a respondent on the n items used to construct the respective scales. Let us refer to this exercise, for want of a better label, as an evaluation of 'objective' matching. It is important to note that

[4] The precise number of items per scale varied across the VAA cases.

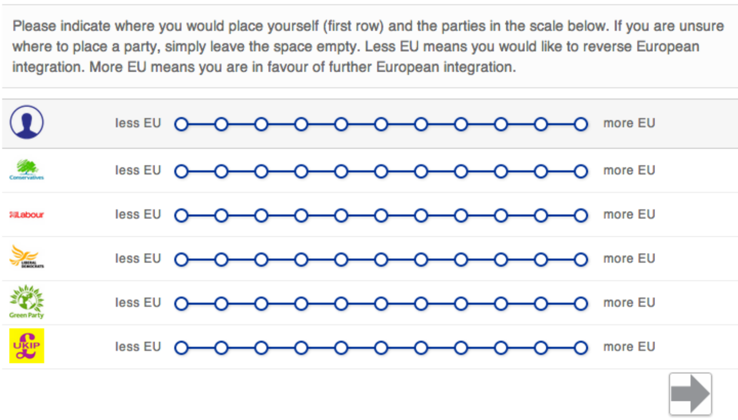

Figure 4: An example of a self–placement and party placement scale

according to this distinction the high dimensional modeling also constitutes an evaluation of 'objective' ideological matching.

For the second level in the analysis, it is possible to contrast the so–called 'objective' matching with a more 'subjective' model based on the personal evaluations of respondents. To this end, I rely on the EUvox set of self–placement and party placement questions, which are based on an 11–point scale for the three dimensions. Figure 3 provides an example of the England self–placement question for the European scale (the two other questions were similarly framed).[5] Using these data as the input values it is possible to repeat the low dimensional modeling process described above –with the crucial difference that we are using the respondent's subjective evaluations of their position and the party positions on the three scales. This approach, based on the subjective evaluation of respondents, is of course closer to how competing proximity and directional models are tested in the literature –though not to my knowledge using machine learning algorithms. Since the matching function can only be performed if the respondent has placed themselves and the five parties on the three scales, for the 'subjective' evaluations approach listwise deletion is applied. Lastly, in order to model directional theory it is necessary to have a '0' point. Accordingly, the three 11–point scales are coded on a –5 to +5 scale.

4 Datasets, selection criteria and pre–processing

The datasets used for the empirical analysis were generated through the Preference Matcher VAA system. Five datasets from the 2014 EUvox were used. The datasets include England, Ireland, Greece, Poland and Spain.

[5] In the actual VAA a mouse–over would alert the respondent to the name of each party when he/she hovered over the scale

Apart from trying to include a range of cases that vary geographically across Europe, from north to south and east to west, as well cases that vary according to the PPD typology, my selection was also premised on party coding criteria. The EUvox party coding system is based on an intensive, iterative and anonymised coding process (see Gemenis and Van Ham 2014 for a description). Given the importance of the party coding estimates for an algorithm's matching function, I have selected country cases where the party coding involved at least five coders per party and more than one round of anonymous coding. Not all EUvox country cases satisfied this party coding criteria, although a majority of the cases did so. Having selected the five country cases, the number of parties to be included for the empirical analysis was set by England. The English VAA was limited to five parties, which then set the maximum number of parties to be included across the remaining VAA datasets. Striving to ensure that all country cases included the same number of parties was done in order to enhance comparability across the performance metrics –whether high or low dimensional. Furthermore, using the five party threshold selection criterion also ensures that party position estimates across the policy items are likely to be more accurate since marginal parties, which may be harder to code or have many 'no opinions', were excluded. As can be seen from the table in annex 1 the distribution of vote share using this threshold is fairly similar across VAA datasets. A further criterion for party inclusion was the number of respondents per party, which was set at >200 across the datasets. With the exception of the fifth largest Polish party by vote share, all of the remaining five largest parties more than satisfied this criterion. The Polish party distributions for respondents with a vote intention for the PSL party were especially skewed with fewer than 200 respondents after cleaning compared to approximately 10 times this number for the next lowest party (see below).[6] This party was therefore excluded from the analysis.

Before any analysis could be carried out it was necessary to clean the raw datasets (on cleaning techniques see Andreadis 2014; and Tziouvas et al. 2014). Here I rely on a simple cleaning method. Since all Preference Matcher datasets have individual item response timers it is fairly straightforward to filter out 'rogue' entries such as those arising from visitors testing the VAA, skimming through the VAA or completing the questionnaire in a 'satisficing' manner (Krosnik 1999).[7] Applying robust cleaning criteria should ensure

[6]Part of the reason may be that the KNP –the other populist party attracted many respondents with well over 10,000 respondents after cleaning.

[7]The following criteria were used for filtering potential rogues during the cleaning process: (1) <120 seconds to complete the questionnaire; (2) $\leq$ 1 second for any of the 30 policy items; (3) $\leq$ 2 seconds for more than 4 policy items; (4) >20 No opinions ; (5) >15 same consecutive answers; (6) <20 seconds to fill out the webpage with political and socio–demographic variables; (7) respondents who filled in the questionnaire with the mobile phone version of the VAA (this had a slightly different answer category configuration); (8) respondents from a country other than the respective VAA country.

that the datasets used for analysis would have removed most 'rogue' entries even if this comes at the expense of removing a small number of potentially valid entries.

5 Results

Discriminating between algorithms

We begin by evaluating the performance of the models in terms of being able to discriminate between respondents' vote intention and other party alternatives as defined in equation 5. This is based on using as predictors the coefficients derived from the high dimensional modeling. The higher the mean discrimination index score –the average across all parties– the greater the discriminatory power of the model for a given country, with positive coefficients indicating on balance better discrimination while negative coefficients suggest greater problems in separating respondent's party of choice from alternatives.

Table 1 presents the results, where the first set of columns include the cases where proximity models could be expected to function better and the second set of repeated columns contain the directional cases. The cells with the best performing model per country are in bold. Looking at the overall classification we find some initial evidence to support the PPD model described in the theory section. That is, the directional model appears to be better able to separate respondents' party choice from other party alternatives in the countries implied by their model –Greece and Poland. Conversely, modeling using a proximity logic appears to generate better results in those countries which conform to the pure proximity model as depicted in the PPD model in figure 2 (i.e., in the lower left quadrant). Some notable variation exists across the cases, especially in Spain and Ireland where the best performing model returns negative coefficients suggesting that there is a possible clustering of parties in the ideological space that makes it difficult to separate parties.

Table 1: Discrimination power of three models

Case	*Model*	*Discrim.*	*Case*	*Model*	*Discrim.*
	Proximity	**0.701**		Proximity	0.530
England	Directional	-0.662	Greece	Directional	**1.422**
	Hybrid	-0.578		Hybrid	0.626
	Proximity	-1.454		Proximity	3.736
Spain	Directional	**-0.183**	Poland	Directional	**5.700**
	Hybrid	-1.501		Hybrid	4.242
	Proximity	**-1.613**			
Ireland	Directional	-2.825			
	Hybrid	-1.657			

Table 2: Overall summary statistics

Case	Statistic	Model 1			Model 2		
		Proximity	Directional	Hybrid	Proximity	Directional	Hybrid
England	Recall	0.516	0.535	0.524	0.537	0.554	0.546
	Precision	0.560	0.569	0.570	0.571	0.583	0.580
	Accuracy	0.617	0.633	0.623	0.631	0.645	0.638
	Kappa	0.490	**0.512**	0.499	0.510	**0.530**	0.520
Ireland	Recall	0.447	0.490	0.470	0.451	0.517	0.489
	Precision	0.453	0.504	0.485	0.446	0.531	0.496
	Accuracy	0.480	0.525	0.503	0.485	0.552	0.521
	Kappa	0.335	**0.392**	0.365	0.342	**0.428**	0.389
Spain	Recall	0.508	0.519	0.516	0.517	0.532	0.530
	Precision	0.516	0.527	0.524	0.526	0.536	0.536
	Accuracy	0.494	0.506	0.502	0.502	0.513	0.512
	Kappa	0.311	**0.327**	0.323	0.321	**0.338**	0.337
Greece	Recall	0.533	0.531	0.533	0.548	0.538	0.547
	Precision	0.510	0.507	0.508	0.612	0.593	0.612
	Accuracy	0.663	0.660	0.661	0.669	0.661	0.669
	Kappa	**0.526**	0.523	0.524	**0.537**	0.525	**0.537**
Poland	Recall	0.630	0.642	0.633	0.649	0.660	0.657
	Precision	0.729	0.737	0.725	0.738	0.745	0.742
	Accuracy	0.793	0.800	0.795	0.803	0.810	0.808
	Kappa	0.666	**0.678**	0.670	0.684	**0.694**	0.692

Predictive modeling using machine learning techniques

We now turn to the predictive modeling analysis based on machine learning algorithms to evaluate performance on 'unseen' samples. A number of findings can be noted from table 2, which is based on high dimensional models. Although I report all the commonly used evaluation metrics, we shall mostly rely on the kappa coefficient as an evaluation measure of predictive performance. As stated in section 3, the kappa statistic corrects for random chance classifications and is especially well–suited for dealing with unbalanced samples (which is the case for the EUvox datasets where the distributions of parties per country are skewed). Nonetheless, the reader may also want to check the overall recall measure which can simply be interpreted as the averaged proportion of correct predictions.

Let us start with the most basic model, which we can refer to as model 1. In this model only the party matching coefficients derived from the 3 issue voting models are used as predictors to train the model for classifying the 'unseen' respondents' vote choice. As described in section 3 the training model is estimated using a boostrapped multinomial logistic regression. In terms of the predictive performance of the classifiers, a fairly consistent picture emerges with regard to the basic model. Taking the kappa measure as our main evaluation measure, we find that models built using a pure directional model yield better results for all cases, with one counter–intuitive exception, the case of Greece. Here we find that the kappa coefficient is higher for the proximity model. However, on closer inspection it appears that for Greece the coefficients are virtually indistinguishable across the three issue voting models (similarly too for the other metrics including the recall measure).

In a second step, I develop a more fully specified model, model 2, which includes in addition to the party matching coefficients, four extra predictors: age, gender, education[8] or whether the respondent is an 'issue' voter.[9] What is of particular interest in this exploratory analysis is that adding extra predictors has only a marginal effect on performance. Looking at the result for model 2, we find that the pure directional model performs best across all country cases –with the exception that in Greece the first rank is shared between the proximity and hybrid models. Again, we find that in Greece the performance is virtually identical among the models.

Overall, then, when training models to evaluate subsequent predictive performance on 'unseen' samples, the pure directional model performs better with the exception of Greece where there is little difference among the models. There is some notable variance across the cases however. The re-

[8]This is coded as a binary variable where 1= degree and 0= no degree.

[9]A binary variable is created where 1 means the respondent's choice of party was based on the party's policy programme and 0 means the remaining options, which were charisma, competence to govern, clientalism, and other).

sults dovetail the problems which could be detected in the discrimination index (table 1) where Ireland and Spain also had comparatively lower scores. In particular, the kappa scores for Spain would be considered 'poor' using the Fleiss scales (<40) or 'fair' using the Landis and Koch evaluation (0.21–0.40). An ex–post explanation for this differential may be connected to the 'Podemos' phenomenon –a brand new party that did unexpectedly well and occupies a similar ideological space to the more traditional United Left party. For the remaining cases, the kappa coefficient would be considered moderate to good using the two scales with scores approaching .7 for Poland.[10]

We now move onto examine the results of the low dimensional models predictive modeling. Here, we are only comparing proximity and directional models since there is no hybrid model. In a first step we compare the so–called 'objective' evaluations, which are based on party and respondents' placements derived from the Mokken–validated scales for the economic, cultural and European dimensions. The results are presented in the columns under the 'Objective' heading in table 3.

Overall, it is quite clear that the proximity model performs best in a three dimensional space. There is also some notable variation across the cases. In particular, the results for Ireland and Spain comparatively poor compared to the other cases across all metrics. This mirrors what we found in the high dimensional modeling. As expected, Poland registers the highest scores given that the classification task is easier with four parties.

The second set of columns in table 3 provide the results for the low dimensional modeling based on the 'subjective' evaluations of the respondent. A somewhat similar picture emerges to the 'objective' evaluations with the proximity models easily outperforming the directional variant; and similar variance across the cases with machine learning algorithms experiencing comparatively greater classification difficulties in Ireland and Spain than with the remaining cases. Not surprisingly, the predictive performance is also superior when using the 'subjective' evaluations as predictors as can be seen by the considerably higher kappa values across all cases. In fact, according to the Fleiss scale the kappa scores could be considered excellent in Poland where the classification task is easier; but they are also relatively high for the case of Greece.

Comparing models

In the absence a benchmark it may be rather difficult to get a handle on the predictive performance of the supervised machine learning models. What we can therefore do is to estimate a model using a set of predictors that would be more commonly used in political science. For instance, self–placement

[10] Indeed, using Landis and Koch's scales the results for Poland would be considered as 'substantial'. The difference between Poland and the rest is of course due to the classification problem involving four parties instead of five.

Table 3: Summary statistics for low dimensional modeling

Case	Statistic	Objective		Subjective	
		Proximity	Directional	Proximity	Directional
England	Recall	0.514	0.459	0.644	0.605
	Precision	0.544	0.502	0.667	0.636
	Accuracy	0.612	0.573	0.686	0.651
	Kappa	**0.483**	0.423	**0.590**	0.545
Ireland	Recall	0.448	0.400	0.595	0.528
	Precision	0.421	0.469	0.606	0.529
	Accuracy	0.487	0.449	0.597	0.537
	Kappa	**0.342**	0.287	**0.490**	0.414
Spain	Recall	0.485	0.442	0.712	0.683
	Precision	0.496	0.463	0.702	0.682
	Accuracy	0.481	0.455	0.680	0.652
	Kappa	**0.288**	0.237	**0.553**	0.512
Greece	Recall	0.542	0.480	0.711	0.667
	Precision	0.512	0.540	0.747	0.697
	Accuracy	0.665	0.606	0.778	0.739
	Kappa	**0.532**	0.450	**0.694**	0.637
Poland	Recall	0.627	0.560	0.729	0.736
	Precision	0.704	0.653	0.816	0.801
	Accuracy	0.775	0.708	0.857	0.856
	Kappa	**0.644**	0.533	**0.753**	0.750

on a left–right scale is a key variable that is used in the political behaviour literature, with left–right orientations being strongly related to party choice (e.g. Van de Eijk et al. 2005). We can therefore build a predictive model based on the left right self–placement variable. Table 4 provides an overview of the predictive performance of the various models using the kappa criterion. To aid the performance of the left/right model we add the socio-demographic variables such as age, gender and education in order to train the model (as were used for model 2 in table 2). The kappa scores are noticeably poor across all cases with the exception of Poland with its relatively easier four party classification task. Indeed, for Spain and Ireland they drop below the <20 threshold considered as slight by Landis and Koch (1977).

For the remaining models only voter–party matching coefficients are used. For the high dimensional (HD) model I use the directional models' kappa scores from table 2 and for low dimensional (LD) models the proximity scores from table 3. Taking the two 'objective' models, it appears that the high dimensional (HD) model performs better, as can be seen by the higher mean K value. What is noteworthy here is that if scale validation had not been conducted for the LD 'objective' model, the kappa scores would have been considerably lower. This has obvious implications for VAA design, for instance there are VAA models that eschew high dimensional matches and only provide non–validated low dimensional matching. More generally what these analysis imply is that using so–called 'objective' evaluations for modeling ideological congruence among voters and parties could be a fruitful methodology for addressing broader questions and an alternative or complement to other methods such as relying on 'propensity to vote' scores. Nonetheless, it is also clear that when it comes to overall performance the low dimensional subjective evaluations has superior predictive performance.

Table 4: Kappa statistic across models using a Left/Right benchmark

Case	Left/Right model	HD 'objective'	LD 'objective'	LD 'subjective'
England	0.319	0.512	0.483	0.590
Ireland	0.177	0.392	0.342	0.490
Spain	0.197	0.327	0.288	0.553
Greece	0.282	0.526	0.532	0.694
Poland	0.431	0.678	0.644	0.753
Mean K	0.281	0.487	0.458	0.616

6 Discussion

The aim of the paper was to model voter–party choice based on a proximity and directional logic and to test the predictive performance of the ensuing models across a range of country cases. The datasets where generated by the EUvox VAA for the European Parliament elections of 2014. Furthermore, institutional context was used as a criterion for determining the selection of cases where proximity logic could be be expected to perform better than directional models and vice versa. In terms of the predictions, we were concerned with a model's performance in correctly assigning respondents to their party of choice. The analysis was performed in high dimensional space (using all VAA policy statements) as well as low dimensional space (using EUvox's three dimensional issue space). Furthermore, a distinction was made between so–called 'objective' and 'subjective' evaluations where the former referred to modeling voter–party matches based on the respondents' answers to a battery of policy items and the academically coded positions of parties on those same policy items. In testing 'subjective' evaluations the voter–party models were built using as inputs the respondent's self placement and his/her placement of the parties on three issue scales. Lastly, in addition to developing new metrics for evaluating algorithm performance, techniques borrowed from computer science and data mining were introduced. Specifically, a supervised machine learning methodology was implemented whereby the predictive performance of the competing models was evaluated in terms of their classification performance on 'unseen' samples.

A number of evaluation metrics where then used to compare predictive performance across the models. Echoing the more recent literature, in terms of the proximity versus directional controversy the results are somewhat mixed. Based on the empirical analysis conducted in this paper, much depends on whether the modeling is conducted in a high dimensional or low dimensional space, and whether objective or subjective evaluations are used as predictors for model building. Yet there are some clear patterns. When assessing the discriminatory power of the three issue voting algorithms in high dimensional matching (i.e., using as predictors the coefficients typically outputted by a VAA), the results seemed in line the Prados-Pardo and Dinas (2010) typology of polarization. The proximity model on average tended to better separate respondents' party choice from alternative parties configurations in the three country cases where this was expected. Conversely, the directional model performed better in the two cases where polarization was greater. However, when more sophisticated models were developed using supervised machine learning techniques to test the predictive performance of the competing theories a rather different picture emerged. With the exception of Greece, where the all three theoretical models performed equally well, directional models yielded better results. Another finding in this respect, is that adding additional predictors to the party matching coefficients –such as socio-demographics– have little impact on overall predictive performance.

The analysis was repeated in low dimensional space. Here a very clear picture emerged in which proximity modeling performed substantially better when using both 'objective' and 'subjective' evaluation as feature inputs for the machine learning task. Not surprisingly, performance was greater when respondents' subjective evaluations are used as input features. One noteworthy result in comparing high and low dimensional models (only the 'objective' evaluations models can be compared in this respect) is that the predictive performance based on high dimensional modeling tends to be superior. Although further analyses would be required to test this more thoroughly, a preliminary interpretation could be that the dimensionality reduction implied through scaling analyses comes at the expense of potential information loss *vis–a–vis* matching based on the entire range of policy items.

Overall this suggests that using 'objective' evaluations (i.e., VAA–based coefficients whether derived from a high or low dimensional space) can be used to examine ideological congruence among voters and parties. Such a measure clearly outperformed the type of approaches that are more common in the literature as demonstrated when comparing party matching coefficients with a more fully specified model using left/right ideological orientations. This leave a number of areas where this line of research could be taken. Obviously, the comparative scope could be *enlarged* by examining more cases. Or, alternatively, a *deeper* analysis of the variation in patterns of voter–party ideological congruence within countries could be conducted. This was beyond the scope of this paper, but it would certainly be an area where more could be done, especially where the models performed comparatively worse as was the case for Spain and Ireland. It would be also be possible to deepen by using more input features for the supervised machine learning, for example more voter or party characteristics. This was eschewed for the purposes of this paper where the aim was to illustrate the potential of the techniques by using parsimonious models –though I would argue that doing so also avoids model over–fitting. Indeed, in terms of the methods deployed only one of many supervised machine learning classifiers was presented, alternative commonly used classifiers which could have been tested include Bayesian models and Support Vector Machines models. Further dimensionality analysis could also be carried out, this paper only used the validated scales on an early subset of respondents. Such an analysis would no doubt reveal that, depending on the institutional context, different solutions to the three dimensional ex–ante imposed scales work best. For the broader literature, these various different research angles may shed some further light on aspects of the proximity vs. directional controversy –though mixed results seems to be the most likely outcome. More generally, using the VAA–generated data for mapping ideological congruence among voters and parties beyond some of the more limited measures currently used seems a promising track while at the same time feeding input into the very design of VAAs.

Annex 1: High dimensional matrices

The voter-party matching coefficients are based on the matrices below. The headings in the columns and rows are based a five–point Likert scale with the following answer categories: Completely Agree (CA); Agree (A); Neither Agree nor Disagree (N); Disagree (D); Completely Disagree (CD). Excluding the 'no opinion' answer category from the matrix, this results in 25 possible 'matches' between a respondent and party. The numbers in the cells of the matrix represent the points assigned for a 'hit' by the VAA algorithm. The scores are summed and divided by the total number of questions answered by the respondent. This will generate a coefficient that ranges from -1 to +1 (although for an online VAA it is rescaled to a range of -100 to 100).

A typical matching matrix for a VAA based on a City Block (proximity) logic is as follows:

$$\begin{array}{cc} & \begin{array}{ccccc} CA & A & N & D & CD \end{array} \\ \begin{array}{c} CA \\ A \\ N \\ D \\ CD \end{array} & \begin{pmatrix} 1 & 0.5 & 0 & -0.5 & -1 \\ 0.5 & 1 & 0.5 & 0 & -0.5 \\ 0 & 0.5 & 1 & 0.5 & 0 \\ -0.5 & 0 & 0.5 & 1 & 0.5 \\ -1 & -0.5 & 0 & 0.5 & 1 \end{pmatrix} \end{array}$$

A directional model based on a Scalar Product metric is given by the following matrix:

$$\begin{array}{cc} & \begin{array}{ccccc} CA & A & N & D & CD \end{array} \\ \begin{array}{c} CA \\ A \\ N \\ D \\ CD \end{array} & \begin{pmatrix} 1 & 0.5 & 0 & -0.5 & -1 \\ 0.5 & 0.25 & 0 & -0.25 & -0.5 \\ 0 & 0 & 0 & 0 & 0 \\ -0.5 & -0.25 & 0 & 0.25 & 0.5 \\ -1 & -0.5 & 0 & 0.5 & 1 \end{pmatrix} \end{array}$$

The third model, a hybrid model, draws mostly on directional logic splits the difference between the two:

$$\begin{array}{cc} & \begin{array}{ccccc} CA & A & N & D & CD \end{array} \\ \begin{array}{c} CA \\ A \\ N \\ D \\ CD \end{array} & \begin{pmatrix} 1 & 0.5 & 0 & 0.5 & 1 \\ 0.5 & 0.625 & 0.25 & -0.125 & -0.5 \\ 0 & 0.25 & 0.5 & 0.25 & 0 \\ -0.5 & -0.125 & 0.25 & 0.625 & 0.5 \\ -1 & -0.5 & 0 & 0.5 & 1 \end{pmatrix} \end{array}$$

Annex 2: Party selection

Table 5 provides includes all the parties used for the empirical analyses in terms of their ranking for the 2014 European Parliament elections and their voter share in per cent. Five parties were used for modeling for all cases with the exception of Poland, where the sample of respondents for the PSL was considered too unbalanced for predictive modeling.

Table 5: Parties selected for the empirical analysis by vote share

Case	*Party 1*	*Party 2*	*Party 3*	*Party 4*	*Party 5*
England	UKIP 26.60	Labour 24.43	Conservative 23.05	Green Party 6.91	LibDem 6.61
Greece	Syriza 26.57	New Democracy 22.72	Golden Dawn 9.39	Elia 8.02	Potami 6.60
Ireland	Fine Gael 22.30	Fianna Fáil 22.30	Sinn Féin 19.50	Labour 5.30	Green Party 4.90
Spain	Partido Popular 26.09	PSOE 23.01	United Left 10.03	Podemos 7.98	UPyD 6.51
Poland	Civic Platform 32.13	PiS 31.78	SLD-UP 9.44	KNP 7.15	

References

Agathokleous, M., N. Tsapatsoulis and I. Katakis, On the quantification of missing value impact on Voting Advice Applications, in L. Iliadis, H. Papadopoulos and C. Jayne (eds.): Proceedings of EANN 2013, Part I, CCIS 383 (Springer Berlin Heidelberg, 2013), pp. 496–505.

Anderson J., and T. Fossen, Voting Advice Applications and Political Theory: Citizenship, Participation and Representation, in D. Garzia and S. Marshall (eds.): Matching voters with parties and candidates (ECPR Press, Essex, 2014).

Andreadis, I., Data quality and data cleaning, in D. Garzia and S. Marshall (eds.): Matching voters with parties and candidates (ECPR Press, Essex, 2014).

Baka, A. L. Figgou and V. Triga, 'Neither agree, nor disagree': a critical analysis of the middle answer category in Voting Advice Applications, International Journal of Electronic Governance, Vol. 5, No. 3/4 (2012), pp. 244–263.

Bakker, Ryan, Seth Jolly, and Jonathan Polk. Complexity in the European Party Space: Exploring Dimensionality with Experts. European Union Politics 13, no. 2 (2012): 219–45.

Benoit, K., and M. Laver, (2012). The dimensionality of political space: Epistemological and methodological considerations. European Union Politics, 13(2):194–218.

choice. In: J. Thomassen, Editor, The European Voter. A Comparative Study of Modern

Costello, R, J Thomassen, and M Rosema. European Parliament Elections and Political Representation: Policy Congruence between Voters and Parties. West European Politics 35, no. 6 (2012): 1226–48.

Democracies, Oxford University Press, Oxford (2005), pp. 167–191.

Diermeier, Daniel, Jean–François Godbout, Bei Yu, and Stefan Kaufmann. Language and Ideology in Congress. British Journal of Political Science 42, no. 01 (2012)

Downs, A. An Economic Theory of Political Action in a Democracy. The Journal of Political Economy 65, no. 2 (1957): 135–50.

Garzia D. and S. Marshall (eds.), Matching voters with parties and candidates (ECPR Press, Essex, 2014).

Garzia D., A. De Angelis and J. The impact of Voting Advice Applications on electoral participation. in

Gemenis, K., and C. Van Ham, Comparing Methods for Estimating Parties Positions in Voting Advice Applications, in D. Garzia and S. Marshall (eds.): Matching voters with parties and candidates (ECPR Press, Essex, 2014)

Gemenis, K., Estimating parties' policy positions through voting advice applications: Some methodological considerations, Acta Politica 48 (2013), pp. 268–295.

Germann, M., F. Mendez, J. Wheatley and U. Serdult, Spatial maps in Voting Advice Applications: The case for dynamic scale validation, Acta Politica.

Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., and Tibshirani, R. (2009). The elements of statistical learning (Vol. 2, No. 1). New York: Springer.

Hooghe, L., G. Marks, and C. J. Wilson, C. J. (2002). Does left/right structure party positions on european integration? Comparative Political Studies, 35(8): 965–989.

Katakis, I., N. Tsapatsoulis, F. Mendez, V. Triga and C. Djouvas, Social Voting Advice Applications – Definitions, Challenges, Datasets and Evaluation, in IEEE Transactions on Cybernetics 2013.

Katakis, I., N. Tsapatsoulis, V. Triga, C. Tziouvas and F. Mendez, Clustering Online Poll Data: Towards a Voting Assistance System, in Proceedings of the 7th Semantic and Social Media Adaptation and Personalization workshop (SMAP'12), (IEEE Press, 2012), pp. 54–59.

Kedar O. How voters work around institutions: Policy balancing in staggered elections. Electoral Studies. 25(3):509–527; 2006.

Kriesi, H., E. Grande, R. Lachat, M. Dolezal, S, Bornschier, and T. Frey, (2006). Globalization and the transformation of the national political space: Six european countries compared. European Journal of Political Research, 45(6):921–956.

Krosnick, J A. Survey research. Annual review of psychology 50.1 (1999): 537–567.

Kuhn, M., and Johnson, K. Applied predictive modeling. New York: Springer. 2013

Lacy, D. and Paolino, P. (2010). Testing proximity versus directional voting using experiments. Electoral Studies, 29:460–471.

Landis, R., and G. Koch. The Measurement of Observer Agreement for Categorical Data. Biometrics, 1977, 159–74.

Laver, Michael, Kenneth Benoit, and John Garry. Extracting Policy Positions from Political Texts Using Words as Data. American Political Science Review 97, no. 02 (2003): 311–31.

Lewis J. B. and G. King, No evidence on directional vs. proximity voting, Political Analysis, vol. 8, no. 1, p. 21, 1999.

Linhart, E., S. Shikano. A Basic Tool Set for a Generalized Directional Model. Public Choice 140, no. 1–2 (2009): 85–104.

Louwerse T. and M. Rosema, The design effects of voting advice applications: Comparing methods of calculating matches, Acta Politica, doi: 10.1057/ap.2013.30 (2013).

Louwerse, T., and S. Otjes, S. (2012). Design challenges in cross–national VAAs: the case of the EU profiler. International Journal of Electronic Governance, 5(3/4):279–297.

Mendez, F. and J. Wheatley, UsingVAA–generated data for mapping partisan supporters in the ideological space in Matching voters to parties and candidates, in S. Marshall and D. Garzia (eds.): Matching voters to parties and candidates (ECPR Press, Essex, 2014).

Mendez, F. What's behind a matching algorithm: A critical assessment of how VAAs produce voting recommendations, in D. Garzia and S. Marshall (eds.): Matching voters with parties and candidates (ECPR Press, Essex, 2014).

Mendez, F., Matching voters with political parties and candidates:An empirical test of four algorithms, International Journal of Electronic Governance, Vol. 5, No. 3/4 (2012)

Merrill S, Grofman B. A Unified Theory of Voting: Directional and Proximity Spatial Models. Cambridge Mass.: Cambridge University Press; 1999.

Pardos-Prado, S., and D. Elias Dinas. Systemic Polarisation and Spatial Voting. European Journal of Political Research 49, no. 6 (2010): 759–86.

Rabinowitz G., and S.E. Macdonald ,A Directional Theory of Issue Voting,The American Political Science Review, Vol.83, No.1 (1989), pp.93–121.

Teran, L., A. Ladner, J. Fivaz and S. Gerber, Using a Fuzzy–Based Cluster Algorithm for Recommending Candidates in E–Elections, in A. Meier and L. Donze(eds.): Fuzzy Methods for Customer Relationship Management and Marketing: Applications and Classifications (IGI Global, 2012).

Triga, V., Social Representations of Voting Advice Applications:A Comparative Analysis, in D. Garzia and S. Marshall (eds.): Matching voters with parties and candidates (ECPR Press, Essex, 2014).

Triga, V., U. Serdult and T.Chadjipadelis, Special issue on Voting Advice Applications and State of the Art, International Journal of Electronic Governance, Vol.5, No.3/4 (Inderscience, 2012)

Van Camp, K., J. Lefevere and S. Walgrave, The Content and Formulation of Statements in Voting Advice Applications: A Comparative Analysis of 26 VAAs, in D. Garzia and S. Marshall (eds.): Matching voters with parties and candidates (ECPR Press, Essex, 2014).

van der Eijk et al., 2005 C. van der Eijk, H. Schmitt and T. Binder, Left–right orientations and party

Vassil, K., Voting Smarter? The Impact of Voting Advice Applications on Political Behavior, PhD Thesis, Dept. of Political and Social Sciences, European University Institute, 2012.

Walgrave, S., M. Nuytemans and K. Pepermans, Voting Aid Applications and the Effect of Statement Selection, West European Politics, Vol. 32, No. 6 (2009), pp. 1161–1180.

Wheatley, J., C. Carman, F. Mendez and J. Mitchell, The dimensionality of the Scottish political space: Results from an experiment on the 2011 Holyrood elections, Party Politics, (2013).

Wheatley, J., Using VAAs to explore the dimensionality of the policy space: experiments from Brazil, Peru, Scotland and Cyprus, International Journal of Electronic Governance, Vol. 5, No. 3/4 (2012), pp. 318–348.