Matching voters with political parties and candidates: an empirical test of four algorithms

Fernando Mendez

Zentrum fur Demokratie (ZDA),
University of Zurich,
21 Kuttigerstrasse Aarau 5000,
Switzerland
Email: Fernando.Mendez@zda.uzh.ch

Abstract: Voting Advice Applications (VAAs) have enjoyed a growing popularity across Europe in recent years and, increasingly, in other parts of the globe too. In response a growing literature is emerging around the promises and potential perils associated with VAAs. This paper contributes to these debates by addressing a core methodological aspect of VAA design: how voters' policy preferences are aggregated to produce measures of concordance with parties/candidates. To this end, the paper analyses the performance of four VAA models that are based on competing algorithms. The findings, which draw from the analysis of VAA-generated data from four experiments, raise a number of concerns for the future design of VAA systems.

Keywords: voting advice applications; political behaviour; election studies; political science.

Reference to this paper should be made as follows: Mendez, F. (2012) 'Matching voters with political parties and candidates: an empirical test of four algorithms', *Int. J. Electronic Governance*, Vol. 5, Nos. 3/4, pp.264–278.

Biographical notes: Fernando Mendez is Director of the e-Democracy centre (eDC) and a Lecturer in Political Science on the Masters Programme at the University of Zurich and the Swiss Federal Institute of Technology (ETH) since 2007. Apart from his work at eDC he is also a Senior Researcher at the Centre for Research on Direct Democracy. His academic interests include internet politics, direct democracy, comparative federalism, European integration, and comparative public policy.

1 Introduction

Voting Advice Applications (VAAs) have enjoyed a growing popularity across Europe in recent years. During some of the most successful deployments of VAA's in an electoral setting millions of citizens have been issued with so-called voting 'recommendations' (Marschall, 2008; Ladner and Pianzola, 2010). The proliferation of such tools is therefore of special interest to political scientists and researchers investigating the impact of new technologies on society and politics. Reflecting this interest, a growing literature has emerged around the promises and possible problems with VAAs (see Cedroni and Garzia, 2010 for a recent review). This paper contributes to these debates by addressing various methodological aspects related to the design of VAAs. Specifically, it focuses on the issue of how voters' policy preferences are aggregated to produce measures of concordance with parties/candidates. In other words, the paper deals with the core function of a VAA.

The following section begins by providing some background on VAAs and an account of the 'world view' that underpins their design. This provides the basis for discussing four VAA models based on competing algorithms. In Section 3 the datasets generated by four VAA experiments are described. It also discusses the methodology used for testing the various VAA algorithms. Section 4 presents the results of an empirical test of the four VAA models. The discussion in the concluding section then relates the analysis to questions of VAA further design and development.

2 VAAs: what they do and how they view the voter

The idea behind VAAs is to allow citizens to better define their own subjective, political preferences and to match these with the stated (or academically coded) preferences of candidates or political parties that are stored on the online application. Around 30 policy items are typically included in a VAA although in some cases such as the Swiss smarvote it can be up to around 60. The core output of most VAAs is usually a concordance/similarity score between the user and the parties/candidates across the 30-odd policy statements. Preference matching therefore lies at the core of a VAA even though many of the sites add extra layers of graphical visualisation. Common additional graphical features include mapping on an X,Y scatterplot the two dimensions that are widely held to define the issue space in most Western democracies – on the one hand social liberalism versus social conservativism [sometimes referred to as GAL/TAN, (see e.g. Hooghe et al., 2002)], and on the other hand economic left versus economic right. In some VAAs there is an additional mapping of multiple dimensions using a radar charts (sometimes referred to as Spider graphs).

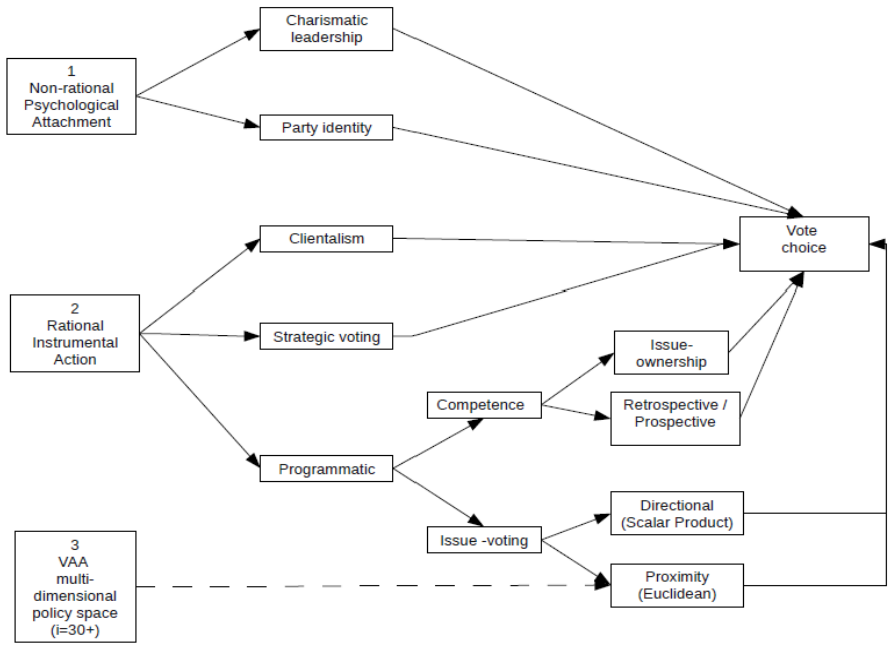
I leave the accuracy of multi-dimensional mapping to one side in this paper, whilst acknowledging that there is good reason to believe that this too raises a number of methodological concerns about accuracy (Louwerse and Rosema, 2011; Gemenis, forthcoming). Also design issues related to statement formulation (Walgrave et al., 2009) and party coding (Wagner and Ruusuvirta, 2009; Gemenis, forthcoming) are outside the scope of this paper. Instead I shall concentrate exclusively on what is usually the core output for most VAAs: the rank ordering of party/candidates in terms of concordance/similarity scores.

What, then, is the 'world view' of the citizen voter that informs a VAA and by extension its recommendation? The short answer is that VAA design is predicated on a rational model of how a voter decides that goes back to Downs' influential economic theory of democracy (Downs, 1957). The voter is assumed to compare his/her stance with that of the candidates on all of the issues. The voter then aggregates his/her preferences on the different issues and votes for the candidate which offers, on balance, the largest number of preferred policies. Over the years these ideas about 'issue voting' have been formalised into a spatial theory of electoral choice.

It is useful to note the underlying assumptions of the spatial models since they have a direct influence on the theoretical basis of a VAA. The key feature of these models of electoral choice is that each voter can be represented by a point on a hypothetical space which reflects the voter's ideal set of policies and that the policy position of each party/candidate can also be represented on that same space (Westholm, 1997; Tomz and Houweling, 2008). A voter chooses the party/candidate whose policy position is closest to theirs. Variants of this approach to electoral choice are called 'proximity models' and such models form the theoretical core of existing VAA models. In such a rationalistic model there is little scope for sociological/psychological factors such as party identification to determine voter choice or for valence factors such as the perceived competence of a candidate/party to deliver the desired goals – all of which have informed alternative theories of voting behaviour.

A crucial distinction can thus be made between a) so-called 'issue-voters' whose vote choice is based on the policy stance of a candidate/party on a given set of policy issues and b) non-issue voters whose vote choice is based on other factors. This distinction will be crucial to the rest of the paper and the analysis conducted in the empirical sections. This distinction is graphically represented in Figure 1. The dotted line shows the VAA world view – a view that is predicated on a particular variant of issue voting theory. VAAs are therefore ill-suited to psychological/sociological models of voting such as those based on factors like party identification or charismatic leadership. Indeed, VAA models are also not suited to other rational models of voting included in Figure 1.

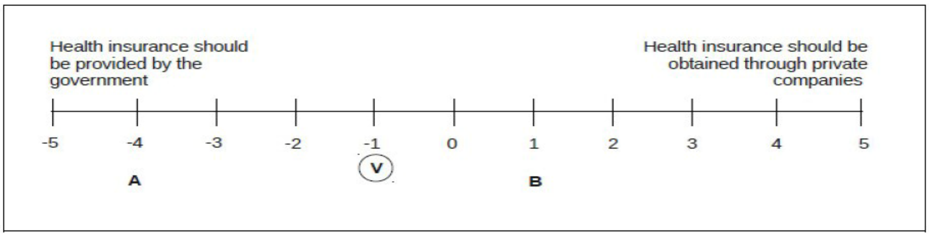
Figure 1 Theoretical linkages in mainstream models of voting



Even accepting the issue voting assumption behind VAAs, there are still two prominent theories of issue voting. The classic model of issue voting – the proximity model – has already been mentioned. The other theory of issue-voting is the directional model (originally formulated in Rabinowitz and Macdonald, 1989). The key difference between the two models relates to how a policy dimension is conceptualised (Westholm, 1997; Tomz and Houweling, 2008; Lacy and Paolino, 2010). In the proximity model what matters most is the distance between the policy alternatives whereas in the directional model what matters most is to be on the correct 'side' of the argument. The directional assumption is that voters do not discriminate between fine policy gradations but are more concerned with the direction or the side of the argument on a particular policy issue.

The differences between the two models of issue voting can be seen in Figure 2 as originally formulated by Rabinowitz and Macdonald (1989). In the health policy example in Figure 2, proximity models and directional models predict different behaviour from the voter, represented by the letter V. According to proximity theory the voter would prefer candidate B since it is closer to the voter's position. In the directional model, however, the voter would prefer candidate A, since both A and V are on the same side of the argument. The zero point represents a neutral point, once it is crossed one is on a different side of the argument. In sum, and as the figure suggests, the two theories generate competing predictions about vote choice (this is a topic to which I shall return in section 3).

Figure 2 Directional and proximity models of issue-voter



Note: A and B are candidate positions, V refers to the voter's position on the issue statement

Source: Rabinowitz and Macdonald (1989)

In addition to competing theories of behaviour there is a question -especially relevant to VAAs- of which metric to use for calculating distances. In terms of the metrics based on a proximity model, two different metrics, a Euclidean distance or a City Block distance, can be used. For the directional model of issue voting an alternative metric has been proposed based on the Scalar Product measure (Rabinowitz et al., 1989). These metrics provide for at least three different algorithms that can be used in a VAA. To the best of our knowledge most VAA models have used the City Block metric rather than the Euclidean distance. The difference between the two methods for calculating proximity is that the city block metric is based on a calculation of the sum of the absolute differences between voter and party on each issue, whilst the Euclidean metric is based on the square of the differences between voter and party on each issue.[1]

The differences can be illustrated rather neatly in matrix form. Consider the following similarity matrix based on City Block logic:

$$\begin{array}{cc} & \begin{matrix} CA & A & N & D & CD \end{matrix} \\ \begin{matrix} CA \\ A \\ N \\ D \\ CD \end{matrix} & \begin{bmatrix} 1 & 0.5 & 0 & -0.5 & -1 \\ 0.5 & 1 & 0.5 & 0 & -0.5 \\ 0 & 0.5 & 1 & 0.5 & 0 \\ -0.5 & 0 & 0.5 & 1 & 0.5 \\ -1 & -0.5 & 0 & 0.5 & 1 \end{bmatrix} \end{array}$$

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). The numbers in the cells of the matrix represent the points assigned for a 'hit' by the VAA algorithm when a voter (rows) and a candidate (columns) land in one of the possible cells for each policy statement. The scale ranges from –1 to +1. Excluding the 'no opinion' answer category from the matrix, this results in 25 possible combinations or 'matches' (i.e. 5^2). The distances are scaled so that the maximum distance, e.g a completely agree by a voter and a completely disagree by a candidate, equals –1 whereas a perfect match, e.g. completely agree by both voter and candidate, equals 1. In short, the similarity coefficient will vary from -1 for complete disagreement to +1 for complete agreement. The overall similarity coefficient is given by summing the total number of hits and dividing by the total number of items answered.

Having presented the City Block matrix we can now take a look at the logic of alternative algorithms for calculating similarity scores. The following matrices present the corresponding similarity scores assigned to each cell based on Euclidean (proximity logic) or Scalar Product (directional logic) and an additional algorithm inspired by directional theory which is referred to as an Hybrid model. A Euclidean[2] similarity matrix is given by the following:

$$\begin{array}{cc} & \begin{array}{ccccc} CA & A & N & D & CD \end{array} \\ \begin{array}{c} CA \\ A \\ N \\ D \\ CD \end{array} & \begin{bmatrix} 1 & 0.875 & 0.5 & -0.125 & -1 \\ 0.875 & 1 & 0.875 & 0.5 & -0.125 \\ 0.5 & 0.875 & 1 & 0.875 & 0.5 \\ -0.125 & 0.5 & 0.875 & 1 & 0.875 \\ -1 & -0.125 & 0.5 & 0.875 & 1 \end{bmatrix} \end{array}$$

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

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

Lastly, a fourth model is introduced inspired by directional logic and referred to as the Hybrid model

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

Essentially, the Hybrid model simply splits the difference between the City Block proximity model and the directional model based on the Scalar Product matrices above. A key difference is illustrated by looking at the respective scores attached to a match on 'neither agree nor disagree' by a voter (row) and a candidate (column). In the Hybrid model the score of 0.5 is half way between the Scalar Product score (0 points) and the proximity model (1 point). Indeed, the 0.5 score for a match in the 'neither agree nor disagree' cell captures the intuition that prospective voters/candidates can see both sides of the argument on a given issue statement. A further difference in the Hybrid model is that it aims to capture the intensity of a preference. This can be gleaned by comparing a match in the 'agree/agree' cell compared to 'completely agree/completely agree' cell in both tables. In short, the Hybrid model conceives the middle category differently whilst also taking intensity of preferences into account. The more intense the preference, the higher the match on a given policy question.

3 Description of the experiments, data and empirical test

In the period between October 2010 and May 2011 the research team deployed a VAA during four electoral contests. Two of the contests were conducted for Presidential elections that took place in Latin America, Brazil in October 2010 and Peru in April 2011, whilst the other two elections were for Parliamentary elections held in Europe, Scotland and Cyprus both in May 2011. Thus in the two Latin America Presidential races VAA users were matched with candidates whereas in the two European cases they were matched with political parties. In all cases the research team partnered with election studies experts from the countries in drawing up the 30 item questionnaire and in coding the candidates/parties.[3]

In section 2 we identified four VAA algorithms based on different theories and metrics. Two proximity models based on a Euclidean metric and a City Block metric and two directional inspired models based on the Scalar Product metric and a Hybrid metric. In terms of testing these VAA models and their performance across the four electoral settings there are three main steps to the analysis:

Step 1 The first step involved the selection of candidates/parties to be used. We opted for a common criterion: to only consider candidates/parties that crossed the five per cent vote share threshold. This threshold was chosen largely on pragmatic grounds. To begin with marginal parties were not always included in every VAA experiment given the difficulties involved in coding them on all the policy issues. Furthermore, using a different threshold, say 4% or 3%, would have resulted in only one extra party being included, the Scottish Green Party -a party which did not compete in the constituency vote for the Scottish elections. Because of these factors, as well as the typically low number of respondents supporting some of the very marginal parties (see additional criteria below), a common five per cent threshold was used across the five electoral settings.

Step 2: Involved identifying committed voters with a specific vote intention. This could be done quite easily since all the VAA experiments included a set of questions on vote intention that were asked *before* the results were produced thereby avoiding the need to rely on respondents filling in an additional opt-in survey (and the additional selection bias that this would entail). By filtering out ‘floating voters’ it was possible to create a dataset of respondents whose vote intention was for one of the candidates/parties that had crossed the 5 percent vote share threshold. Interestingly, in all four VAA settings a large proportion of respondents did not actually specify a vote intention for any of the candidates/parties. Another question that was asked before the VAA results were displayed was the reason why a respondent would vote for a particular candidate/party. In addition to issue voting, a range of answer options were included such as party identification, charismatic leadership, perceptions of competence and clientalism. It was thus straightforward to create a binary variable, issue-voter or non-issue voter, for each respondent in the dataset.

Step 3: involves comparing the performance of each algorithm in terms of matching respondents with their preferred candidate/party. All algorithms assign a coefficient of similarity between the respondent and every candidate/party. It is therefore possible to generate a rank order of candidates/parties for each respondent based on their vote intention. We assign a rank of 1 if the highest coefficient is assigned to the respondent’s chosen candidate/party, a rank of 2 if it is the second highest coefficient, and so on. Also, if two candidates/parties, including the respondent’s preferred one, rank equal first, we assign a rank of 1.5, if they rank equal second, we assign a rank of 2.5 and so on.

The empirical test is to see how well the four VAA models can ‘correctly’ rank respondents’ chosen candidate/party. Since respondents only reveal their first preference, this is what we use as a benchmark for evaluating the degree of convergence between candidate/party preference and voting advice. The closer to 1 the mean rank, the better the performance of the VAA algorithm since a score of 1 would mean that the VAA model ‘correctly’ ranked in first place the candidate/party of all respondents that had the same vote intention. Conversely, the lowest possible score will depend on the number of candidates/parties above the five per cent threshold in a given electoral contest. Where there are five candidates (as in Peru) the lowest possible score is 5, which would occur if the VAA model ranked a users’ candidate choice last, i.e. fifth in the Peru case.

Before turning to the results it is necessary to briefly mention the nature of the datasets generated by VAAs. The datasets typically produced can be large, and the four experiments were no exception. It is crucial to point out that not all data entries are relevant for our analysis. VAA sites can receive many visitors, for instance around 70,000 in the case of Peru or over 7,000 in the case of Cyprus. But not all database entries can be considered ‘valid’ for analysing the questions dealt with in this paper. Typically many users may enter repeated entries, or they may get ‘bored’ towards the end of the questionnaire, and many do not provide additional data (such as party affiliation or their vote intention). Various safeguards to screen out rogue entries such as item response timers or identifiers for multiple entries from the same web browser have been devised. We have preferred to err on the side of caution in terms of identifying potentially rogue entries and the end result is usually a much reduced dataset of ‘core’ users who are more likely to have provided valid entries. The dataset is nonetheless large with over 50,000 respondents with a vote intention for the 16 candidates/parties that crossed the five per cent threshold for each election (see Table 1).

Table 1 Number of users that expressed a vote intention for one of the candidates/parties that crossed the 5% vote share threshold

Country	*Party/candidate*	*Type of voter*	*Number of respondents*
Brazil	Dilma	Issue	1374
		Non-issue	2228
	Serra	Issue	1052
		Non-Issue	2858
	Marina	Issue	2285
		Non-Issue	2238
Cyprus	Akel	Issue	678
		Non-Issue	243
	Dissy	Issue	595
		Non-Issue	686
	Diko	Issue	173
		Non-Issue	144
	Edek	Issue	149
		Non-Issue	114
Peru	Toledo	Issue	1774
		Non-Issue	3817
	Fujimori	Issue	492
		Non-Issue	993
	Castañeda	Issue	347
		Non-Issue	1471
	Kuczynski	Issue	6029
		Non-Issue	13506
	Humala	Issue	1262
		Non-Issue	1105
Scotland	SNP	Issue	1234
		Non-Issue	2694
	Labour	Issue	964
		Non-Issue	1119
	Tory	Issue	481
		Non-Issue	391
	Libdem	Issue	560
		Non-Issue	343
	Total number of respondents		53,399

4 Analysis and results

As noted in the section above, our focus is on the actual rank order produced by a VAA algorithm rather than the precise coefficient (concordance score). What is likely to matter most for an issue driven VAA respondent with a clear vote intention is where their candidate/party is placed. The crucial VAA question is this: Which candidate/party is ranked first?

Let us begin by examining if the data reveal differences between issue voters and non-issue voters by comparing the rank scores of the two types of respondents when they have expressed the *same* candidate/party vote intention. Given that a core part of the argument being advanced rests on issue voters it is expected that the overall scores of non-issue voters, i.e. those respondents who base their vote choice on matters not necessarily connected to the policy positions of candidates/parties, to be less consistent than issue voters. Much as expected, the average rank generated by the four algorithms across the 16 candidates/parties were better on average for issue-voters than they were for non-issue voters. In Table 2, the closer to 1 the average rank for respondents whose vote intention was for a party/candidate in the rows, the better the score.

Table 2 Average rank performance of issue voters versus non-issue voters across algorithms

Country	*Party/candidate*	*Type of voter*	*Euclid*	*City Block*	*Scalar Product*	*Hybrid*
Brazil	Dilma	Issue	1.2	1.33	1.77	1.46
		Non-Issue	1.31	1.45	1.88	1.59
	Serra	Issue	1.61	1.69	1.69	1.66
		Non-Issue	1.66	1.72	1.74	1.68
	Marina	Issue	2.09	1.9	1.37	1.71
		Non-Issue	2.19	2	1.42	1.8
Cyprus	Akel	Issue	1.4	1.37	1.03	1.15
		Non-Issue	2.22	2.1	1.16	1.64
	Dissy	Issue	2.7	2.6	1.83	2.36
		Non-Issue	2.82	2.74	2.05	2.55
	Diko	Issue	1.56	1.7	2.42	1.84
		Non-Issue	1.58	1.68	2.54	1.89
	Edek	Issue	1.32	1.34	1.78	1.41
		Non-Issue	1.55	1.58	2.01	1.64
Peru	Toledo	Issue	1.73	1.93	2.62	2.01
		Non-Issue	1.84	2.01	2.76	2.1
	Fujimori	Issue	2.05	1.77	2.33	1.84
		Non-Issue	2.12	1.87	2.4	1.94
	Castañeda	Issue	2.62	3	3.55	3.31
		Non-Issue	2.62	3.01	3.5	3.27
	Kuczynski	Issue	4.14	3.88	1.67	3.14
		Non-Issue	4.38	4.15	1.82	3.48
	Humala	Issue	1.43	1.36	1.41	1.3
		Non-Issue	1.53	1.45	1.52	1.39
Scotland	SNP	Issue	1.66	1.6	1.4	1.48
		Non-Issue	1.99	1.93	1.67	1.8

Table 2 Average rank performance of issue voters versus non-issue voters across algorithms (continued)

Country	*Party/candidate*	*Type of voter*	*Euclid*	*City Block*	*Scalar Product*	*Hybrid*
Scotland	Labour	Issue	1.89	1.92	2.24	2.05
		Non-Issue	1.96	1.99	2.39	2.13
	Tory	Issue	1.77	1.73	1.15	1.39
		Non-Issue	2.19	2.12	1.29	1.66
	LibDem	Issue	1.23	1.31	1.76	1.45
		Non-Issue	1.29	1.43	1.98	1.59

It is important to stress that the results of Table 2 relate to how well the VAA models perform in terms of ranking respondents with the *same* candidate/party vote intention. Of the 128 cells in the results matrix (16 candidates/parties x 2 types of voters x 4 VAA models) there are only three cells where the average rank of non-issue voters is marginally better than for issue voters (in bold). This was the case with the Cypriot party, Diko, and with the Peruvian presidential candidate, Castaneda. Except for these outlier cases, issue voters were more consistent than non-issue voters in terms of the rankings produced by the various VAA models. We have reasonable support for the argument that issue voters using VAAs appear to be more consistent in terms of the congruence of their policy positions and their vote intention. For the remainder of the paper, therefore, I shall concentrate most attention on issue voters. How well do the various algorithms perform in terms of 'correctly' placing this group of core voters?

Table 3 Issue voters overall scores for each election

Country	*Euclid*	*City Block*	*Scalar*	*Hybrid*	*Best Performer*
Brazil	1.63	1.64	**1.61**	**1.61**	Scalar/Hybrid
Cyprus	1.74	1.75	1.76	**1.69**	Hybrid
Peru	2.39	2.39	**2.31**	**2.31**	Scalar/Hybrid
Scotland	1.64	1.64	1.63	**1.59**	Hybrid

Note: Best scores for each election in bold

Table 4 Non issue voters overall scores for each election

Country	*Euclid*	*City Block*	*Scalar*	*Hybrid*	*Best Performer*
Brazil	1.72	1.72	**1.68**	1.69	Scalar
Cyprus	2.04	2.03	1.94	**1.93**	Hybrid
Peru	2.50	2.5	**2.4**	2.44	Scalar
Scotland	1.86	1.87	1.83	**1.79**	Hybrid

Note: Best scores for each election in bold

Table 3 reports the average performance of the four models for each of the elections. As with the earlier table, the closer to 1 the final score, the better the predictive performance of the respective algorithm. The table shows that the directional inspired models (based on the Scalar Product or the Hybrid algorithm) outperformed the proximity models (based on Euclidean or City Block algorithms) in ranking issue voters' preferred candidate/party across the four electoral settings. The Hybrid model came first or equal first in all the electoral settings. In the case of Brazil and Peru the Scalar Product algorithm shared first place with the Hybrid model. However, the Scalar Product performed comparatively worse in the case of Cyprus, coming last. In Table 4 we report the average rank performance of the four algorithms for non issue voters. Not surprisingly, the overall scores are worse and in the hypothesised direction (i.e. higher) across all models

Table 5 Issue voters correctly ranked in first place

Country	*Euclid*	*City Block*	*Scalar*	*Hybrid*	*Best performer*
Brazil	53.1	50	49.6	**53.8**	Hybrid
Cyprus	50.6	49.6	50.8	**53.3**	Hybrid
Peru	37.2	37.1	35.8	**39.2**	Hybrid
Scotland	58.4	57.4	59.9	**62**	Hybrid

Note: Best scores for each election in bold

Table 6 Non issue voters correctly ranked in first place

Country	*Euclid*	*City Block*	*Scalar*	*Hybrid*	*Best performer*
Brazil	48.6	44.6	46.2	**49.1**	Hybrid
Cyprus	36.9	37	**44.5**	43	Scalar
Peru	34.5	33.9	33.9	**35.2**	Hybrid
Scotland	47.3	47.1	50.2	**52.1**	Hybrid

Note: Best scores for each election in bold

Tables 5 and 6 present the results of a slightly different test. The aim is to see whether the same results hold if the focus is only on the first ranked party/candidate. Note that we are here concentrating solely on the proportion of first placed parties/candidates and not assigning further points to other ranks. Thus, the higher the percentage, the better the predictive power of the algorithm. The table presents the average for each of the four models across the five settings. Again, we find that the Hybrid model performed best across all settings. In short, the directional inspired algorithms seem to perform better in terms of correctly assigning respondents to their preferred party/candidate. Crucially, this is the case for the important subset of respondents that are at the core of our analysis -issue voters. The results hold whether we focus on correctly first placed candidates/parties or use a more refined average rank order measure.

Across the four electoral settings the worse scores tend to be in the Cyprus and Peru cases. Part of the reason for the Peru case, in particular, is that there were 5 main candidates competing for the Presidential elections. But there is another, potentially more important, factor explaining the difficulty in these two cases which can be gleaned from Table 7. The correlation matrix provides an indication of the overlap between candidates/parties in a given electoral contest. It measures the bivariate correlation of the policy positions of the candidates on the 30 policy statements in their respective electoral contest. In other words, to what extent did candidates/parties take similar positions across the 30 policy statements.

Table 7 Correlation matrix (n = 30 policy statements)

Party/Candidate Coding		*PERU*					*SCOTLAND*				*CYPRUS*						
		Toled	*Castaneda*	*Fujimori*	*Kuczynski*	*Humala*	*SNP*	*Lab*	*Tory*	*Libdem*	*Akel*	*Disy*	*Diko*	*Edek*	*Dilma*	*Serra*	*Silva*
PERU	Toledo	1															
	Castaneda	**.761****	1														
	Fujimori	–0.036	–0.054	1													
	Kuczynski	**.583****	**.585****	–0.145	1												
	Humala	0.249	0.336	–0.311	0.066	1											
SCOTLAND	SNP						1										
	Lab						–0.28	1									
	Tory						–.682**	0.241	1								
	Libdem						0.03	–0.345	–0.039	1							
CYPRUS	Akel										1						
	Disy										–.475**	1					
	Diko										–0.241	**.425***	1				
	Edek										0.142	–0.295	**.387***	1			
BRAZIL	Dilma														1		
	Serra														0.318	1	
	Silva														0.151	–0.316	1

Notes: ** Pearson's R is significant at the 0.01 level (2-tailed).
* Pearson's R is significant at the 0.05 level (2-tailed).
§ Positive correlations in bold type.

We can see in the table that there are a number of statistically significant correlations. Significant positive correlations, which are in bold in the matrix, will make it more difficult for the VAA to distinguish amongst the various candidate/party options. On the other hand, negative correlations amongst candidates/parties are a welcome sign of policy differentiation. Not surprisingly, the two electoral cases where there are significant positive correlations, Cyprus and Peru, are also the cases where the algorithms tend to perform worse. Nonetheless, in these difficult cases, especially Peru, the toughest case, we also find that the directional inspired models performed better.

5 Discussion

The VAA phenomenon appears to be spreading beyond its former home in Western Europe to other countries across the globe. In some European countries the number of 'voting recommendations' produced have the potential to conceivably affect electoral outcomes. In response, political scientists have begun to critically evaluate methodological issues that are raised in the design of VAAs. In this paper we raised the issue of rival theoretical models upon which VAAs could be built and the various competing metrics that they give rise to: a proximity model based on a City Block metric (the most common method) or a Euclidean distance, and a directional model based on a Scalar Product metric. We also introduced a so-called Hybrid model that takes into account both the 'side of the argument' and the 'intensity' of preferences on a given policy statement. We then set up an empirical test to examine the performance of the four VAA models based on data generated by four real-world experiments. The basic claim is that the directional inspired models performed better than proximity based algorithms.

The results presented in this paper with regard to directional algorithms should not be viewed as a call for using the Scalar Product (the metric used in directional theory) as a model for a VAA design. Such a metric, I believe, is not particularly well suited to a VAA. Other things being equal the metric would, for example, tend to consistently match a moderate right leaning voter with a more extreme right party across a given set of policy items. Nonetheless, directional theory does point to two factors -being on the 'correct side of the argument' and the intensity of preferences- that could be important for a VAA design, especially when based on a five-point Likert scale that includes a middle category. These two elements were incorporated into the Hybrid model, which tended to outperform the other models in terms of being a better predictor of the vote intention of issue voters.

Evidently, these results are based on but one of a variety of tests that could be used to potentially adjudicate between competing algorithms. Many other tests will no doubt be performed on the datasets produced by VAAs. What is likely to emerge from the resulting research is that there is no single best way to aggregate policy preferences to produce a voting recommendation. This holds even if VAA designers base their algorithms on a proximity model. A Euclidean metric and City Block metric will generate different similarity coefficients as well as rank ordering of parties/candidates. This suggests that to the extent that a voting recommendation can be produced it must be treated with some degree of scepticism – much will depend on institutional factors, such as the type of electoral system, the structure of political competition, and the number of effective parties included in a VAA, the coding of parties, as well as design issues concerning the type of answer scales used and how middle categories and 'no opinions' are treated. Many of these issues have been beyond the scope of this paper but further research is likely to be conducted in this direction. Instead, I have argued that VAA designers' choice of algorithm -i.e. the one that goes into the 'black box'- is not necessarily neutral. Because of this, and the fact that the different metrics can produce alternative rank ordering of parties/candidates, VAA designers may be well advised to dispense with the concept of a voting 'recommendation' altogether. At a minimum, efforts should be made to 'qualify' the results produced by VAAs to the public using these websites.

References

Cedroni, L. and Garzia, D. (2010) *Voting Advice Applications in Europe: The state of the art*, Scriptaweb, Napoli.

Downs, A. (1957) 'An economic theory of political action in a democracy', *The Journal of Political Economy*, Vol. 65, No. 2, pp.135–150.

Gemenis, K. (forthcoming) '"Estimating parties" policy positions through voting advice applications: some methodological considerations', *Acta Politica*.

Hooghe, L., Marks, G. and Wilson, C.J. (2002) 'Does left/right structure party positions on European integration?', *Comparative Political Studies*, Vol. 35, No. 8, pp.965–989.

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

Ladner, A. and Pianzola, J. (2010) 'Do Voting Advice Applications Have an Effect on Electoral Participation and Voter Turnout? Evidence from the 2007 Swiss Federal Elections', *Electronic Participation*, pp.211–224.

Louwerse, T. and Rosema, M. (2011) 'The design effects of Voting Advice Applications: Comparing methods of calculating results', *Paper presented at the ECPR, Section on Voting Advice Applications under Scrutiny-Assessing their impact on elections*, August 25–27, Reykjavik, Iceland.

Marschall, S. (2008) 'The Online Making of Citizens: Wahl-O-Mat', in Georgi, V.B. (Ed.): *The Making of Citizens in Europe: New Perspectives on Citizenship Education*, Bonn: Bundeszentrale fur politische Bildung, pp.137–141.

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

Ramonaitė, A. (2010) 'Voting Advice Applications in Lithuania: Promoting Programmatic Competition or Breeding Populism?', *Policy and Internet*, Vol. 2, No. 1, pp.117–141.

Tomz, M. and Van Houweling, R.P. (2008) 'Candidate positioning and voter choice', *American Political Science Review*, Vol. 102, No. 3, pp.303–318.

Wagner, M. and Ruusuvirta, O. (2009) 'Faulty Recommendations? Party Positions in Online Voting Advice Applications', *Presented at the APSA Annual Meeting*, 3–6 September 2009, Toronto, Canada.

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

Westholm, A. (1997) 'Distance versus Direction: The Illusory Defeat of the Proximity Theory of Electoral Choice', *The American Political Science Review*, Vol. 91, No. 4, pp.865–883.

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

Notes

1 Technically, there is a difference between the Euclidean distance and a Euclidean squared distance. In the former case, distance is calculated by the square root of the sum of the squared differences for each issue. The additional square root transformation need not concern us for this paper since it does not affect the order in which candidates/parties are ranked by a VAA (see methodology discussion in section 3).

2 To see how the values in the matrix cells were derived consider a fivepoint Likert scale with the following values CA=1, A=2; N=3, D=4, and CD=5. The maximimum distance is 4 (i.e. the difference between CA and CD is 5–1). The next maximum distance is 3 (5–2) and so on. The square of the maximum distance is 16 (4x4), the second maximum distance is 9 (3x3) and so on for each cell. The matrix simply rescales the Euclidean distances so that they range from –1 to +1, rather than from 0–16.

3 In terms of coding the policy positions of the candidates/parties on the 30 issue statements included in the VAA a preliminary coding was carried out by the relevant research teams. Candidates/parties were then invited to check their assigned policy positions and were able to 'correct' codes where justified. Quite uniquely amongst all VAA experiments thus far, in the Peru case the five leading presidential candidates actually filled in the questionnaire themselves. In other cases, such as Scotland, the codes were also checked by party headquarters or by leading political journalists (Wheatley et al., forthcoming).