

A normative set of criteria to increase political competence through Voting Advice Applications

Voting Advice Applications (VAAs) have proliferated in the last years in many European countries, and their effects have been extensively discussed. One of their main objectives is to increase voter's political competence by notifying users of their closest political party according to their own preferences. In order to do this, VAAs compare and aggregate users' and political parties' preferences on a set of policy issues. The main goal of this paper is to argue that current VAAs do not fulfill their stated aims. First, we discuss the notion of political competence advanced by VAAs. Second, we define four normative criteria to evaluate whether their methods of recommendation are likely to increase voters' political competence: informativeness, respect for users' way of comparing and aggregating policy issues, reliability, and transparency. Third, we argue that current VAAs compare and aggregate users' and parties' policy preferences following a weak method. It fails to respect sufficiently the users' way of comparing and aggregating policy issues and is not reliable. To prove it, we analyze the methodology of current VAAs and use the outcomes from the EU-Vox 2014 in several countries. Fourth, we discuss the two possibilities by which VAAs could improve these problems: 1) by using ex-ante survey data to fill their gaps 2) by creating a learning algorithm to adapt the VAA to users' preferences. We found that some changes need to be made if VAAs aim to have an impact on users' political competence.

1. Introduction

This paper aims to critically evaluate the current models of Voting Advice Applications (VAAs) by proposing a set of normative criteria to evaluate the application. The discussion on the methodology behind VAAs and its normative implications is important because these applications have proliferated in recent years [Garzia and Marschall, 2012, Marschall and Garzia, 2014, Ladner and Pianzola, 2015, Ramos et al., 2019]. VAAs have become a part of the electoral process in many countries, reaching a considerable part of the electorate in most European elections, and their roles have diversified [Alvarez et al., 2014]. According to the literature, one of their main goals is to increase users’ political competence by allowing them to acknowledge the disagreement between their political preferences and “the policy standpoints of political parties” [Garzia and Marschall, 2014b, 233]. As VAAs receive a lot of publicity during the election campaign, it is important to evaluate the empirical validity and the normative implications of their methodology [Fossen and Anderson, 2014].

VAAs have been subject of research in different fields of political science [Alvarez et al., 2014, Anderson et al., 2014, Garzia and Marschall, 2014b]. There is a branch of literature focusing on the different effects of these applications. While the majority of research shows that VAAs have an effect on increasing turnout and political knowledge at the aggregate level [Garzia et al., 2017], those experiments discounting users’ self-selection bias found a scarce impact on increasing turnout at the individual level [Mahéo, 2017]. Another branch of literature studies the socio-demographic characteristics of VAA users. The usual conclusion is that the average user “is young, highly educated and keenly interested in politics” [Cedroni, 2010, 253]. This essay addresses a nascent area in VAA literature that mixes a methodological and a normative approach. Regarding the former, there has been various articles on the methods used by VAAs to provide recommendations [Mendez, 2017, Romero et al., 2020], the importance of the statement selection [Walgrave et al., 2009] and the treatment of the data [Djouvas et al., 2016]. Concerning the normative implications of the current methods of VAAs, the few philosophical papers on these applications have typically focused on the influence that they have on a citizen’s conception of democracy, but not on the different methods to provide a recommendation [Anderson et al., 2014]. Only in exceptional cases have some authors pointed out from a normative perspective the importance of VAAs methods for increasing voter’s political competence [Fossen and Anderson, 2014]. One of the most visible outcomes of the crystallization of the normative concerns of VAAs’ developers was the so-called “Lausanne Declaration on Voting Advice Applications.” It consists of a set of articles that establish a “certain standards and minimal requirements that should be respected by all the makers of VAAs” [Garzia and Marschall, 2014a]. The declaration addresses issues of organisation and management of VAAs, usability, access, selectiveness, usability and functioning [Garzia and Marschall, 2014a].

In the following, we present the problem of the voters’ lack of political competence and how VAAs pretend to alleviate this by assuming a narrow conception of democracy

inspired by social choice theory. Second, we analyse four criteria to evaluate VAAs' methods of recommendation: informativeness, respect for the users' way of comparing and aggregating policy issues, reliability, and transparency. Third, we explain the current VAA methods of recommendation and argue that they have structural failures regarding the established criteria, undermining their goal of increasing political competence. Fourth, we probe empirically our affirmations by replicating in several European countries the process of recommendation used by the EU-Vox 2014. Last, we suggest overcoming some of VAAs current failures by employing empirical data and discuss the implications of methods of doing so.

2. VAAs and the problem of citizen (in)competence

2.1. Democracy and voters' political competence

The question of the importance of voter's political competence for the correct functioning of democracy has been subject to unresolved controversies. As Dahl puts it, the standpoint is the notion that "if democracy is to work, it would seem to require a certain level of political competence on the part of its citizens" [Dahl, 1992, 46]. Concerns about voters' lack of political competence were present in the views of some striking philosophers such as Plato, Cicero, and Schumpeter and were explanatory of their concerns about democracy [Caplan, 2011]. For example, Mill, worried about possible undesirable consequences due to the extension of universal suffrage, argued that giving additional weight to politically competent voters would improve collective political intelligence and the outcomes of democratic elections [Stuart Mill, 1859]. He considered politically competent voters to be the well-educated or vocationally well-positioned citizens and believed that giving additional voting to them would be positive for everyone as they would have more opportunities to make better political decisions [Stuart Mill, 1859]. Partially following Mill's insights, we define a competent voter as someone with adequate capability and knowledge on policy issues to cast a vote that could advance her policy preferences [Carpini and Keeter, 1996].

This definition leads to the following question: What does "adequate" knowledge on policy issues mean? In this paper, we do not evaluate the adequacy of voter's knowledge as a function of any appreciation of the "goodness" of preferences nor on the rationality of the arguments displayed by the voter in defending her choice. Instead, VAAs take voters' preferences on policy issues for granted [Anderson et al., 2014]. They map the policy preferences of users and political parties on a set of issues and recommend the party that best fits with the users' policy preferences according to a criterion of comparison and aggregation of policy issues [Mendez, 2012]. Therefore, VAAs implicitly define having political competence as having 1) adequate knowledge about the levels of agreement between voters' preferences on policy issues and the positions of political parties, and 2) the ability to compare and aggregate these policy issues into a single agreement score that determines the party to vote for [Mendez, 2017].

In order to know the level of agreement between her preferences on any policy issue and the positions of the political party within it, a voter must know the existence of this policy issue and the position of political parties. It is expected that these knowledge conditions will be satisfied by VAAs: the very same questionnaire gives an approximated idea of the prominent policy issues at stake, and after receiving a recommendation for a party users can typically see the position of political parties in any policy issue and compare them with their positions [Garzia and Marschall, 2014b]. There are more complexities involved in the comparison and aggregation of these policy issues, which leads to VAAs’ main outcome: the agreement score that determines the recommended party [Louwerse and Rosema, 2014]. This score refers to one of the conditions stated as fundamental for voter’s political competence in VAAs: the aggregation of the levels of agreement with the political parties in the set of policy issues in order to decide which party to vote for [Mendez, 2017]. The agreement score aims to increase voters’ political competence by simplifying, in a meaningful way, the complex process of comparing and aggregating policy preferences [Marschall and Schmidt, 2010]. From this perspective, VAAs aim to be like “a computerized expert system” that “assist an architectural engineer in navigating complex decisions about the construction of a building” [Fossen and Anderson, 2014, 246].

The first question to determine is what conception of democracy is presupposed by VAAs with this understanding of voter’s political competence. The approach of VAAs fits tightly with the normative conception of democracy defended by social choice theory, in which the democratic process can be understood as a means of aggregating preferences on policy issues in order to come up with a political party that respects and considers these policy preferences fairly [Anderson et al., 2014]. According to this view, a competent voter “is well informed about the options on the electoral menu, and therefore competent to choose a political party that matches his or her preferences” [Fossen and Anderson, 2014, 247]. However, the notion assumed by VAAs on democracy and political competence is disputed. Different conceptions of democracy and citizenship cast the problem of citizen competence in a different light than social choice theory [Fossen and Anderson, 2014]. For example, advocates of deliberative democracy are more worried about how voters form their policy preferences than about their lack of knowledge about the disagreement between their policy positions and the positions of political parties [Caplan, 2011]. This discussion is a deeper one that goes beyond the objective of this paper; VAAs are unable to inform citizens regarding the genuine goodness of a specific position on any issue and any design of the application assumes that citizens have legitimate policy preferences on a set of policy issues that should be compared with those of the political parties in order to decide which party to vote for [Anderson et al., 2014]. VAAs are only concerned with matching citizens and political parties using a similarity criterion: the closer the party and the citizen on the policy issues, the higher the chance of being the adequate voting party for the citizen [Mendez, 2012]. Therefore, VAAs treat given policy-preferences on a set of issues as the right way to choose one’s voting party [Fossen and Anderson, 2014]. In the following, we assume the arguable and narrow conception of political competency that every VAA

so far has assumed [Anderson et al., 2014]. The goal of this paper is not to challenge it. Instead, we argue that the methods of current VAAs are insufficient to advance even this narrow conception of political competence.

2.2. Advancing VAAs’ vision of voters’ political competence

According to the literature, low levels of political knowledge are widely exhibited by most citizens [Carpini and Keeter, 1996]. For example, Ferejohn et al. stated that “nothing is certain (to) the student of public opinion and democracy more forcefully than the paucity of information most people possess about politics” [Ferejohn and Kuklinski, 1990, 3]. Certainly, it has been empirically shown that voters lack knowledge about policy issues, suffer predictable biases due to their lack of information regarding how they place political parties in these issues, and experience cognitive dissonance when explaining their reasons for voting [Caplan, 2011]. Moreover, citizens’ lack of knowledge about the policy issues at stake and the position of the political parties within them is not normally an impediment to going to the polls [Hardin, 2006]. For that reason, few researchers deny that there are “room for improvement” [Fossen and Anderson, 2014, 245] regarding the notion of voter’s political competence advanced by VAAs. If this is true, there is a gap between what might be desirable in order to meet the standards for being a competent voter according to VAAs and the current level of voters’ political competence. This political competence gap has been seen as an unavoidable reality to which political systems can respond in different ways, like enforcing politically independent authorities or increasing the role of experts [Caplan, 2011]. Another common alternative is increasing voter’s political competence in order to bridge the gap between what they should know to vote competently and what they actually know [Dahl, 1992]. This is what VAAs do, by offering the users knowledge on the policy issues at stake, the positions of the parties within it and an agreement score based on a comparison and aggregation of the preferences of the user and the positions of the political parties [Fossen and Anderson, 2014].

From our perspective, if VAAs can make citizens more competent, even in the narrow sense of political competence that VAAs advance, they can play a positive role for individuals and society. Voters’ lack of political competence poses a problem because politicians are responsible for a large part of the budget of a country, and they make fundamental decisions regarding the lives of their governed on many policy issues [Somin, 1998]. Therefore, it seems problematic to cast a vote based on mistaken beliefs about the policy issues or the political positions of the parties because this would mean that ballots may have no relationship with what voters care about [Garzia and Marschall, 2014b]. Brennan (2011) has argued that those people who vote without knowing the policy issues or the positions of the political parties in them are engaging in a form of recklessness, as they act without knowing whether their voting behaviour causes harm to others [Brennan, 2011]. Carpini and Keeter (1996) consider being a politically competent voter to be comparable with having access to “the currency of citizenship”. From their standpoint, a politically competent voter acquires a significant resource “for meeting the role of the politically active and involved citizen” [Carpini and Keeter, 1996, 8].

VAAAs are supposed to increase this notion of political competence transparently and respecting users' preferences [Marschall and Garzia, 2014]. They are expected to perform an important function for voters, especially for those considering multiple parties and not knowing exactly what the total disagreement between their policy preferences and the positions of these political parties is [Garzia et al., 2017].

3. Basic operation of VAAAs

Current VAAAs comprise a set of questions of size N . Answers to the questions are typically restricted to a Likert scale, expressing agreement in five degrees plus a *No opinion* answer: $\mathcal{L} = \{CA, A, N, D, CD, NO\}$ ¹. Each user gives an answer belonging to \mathcal{L} for each question. A user *profile*, \vec{u}_i , is the vector containing the answers of user i for all N questions. VAAAs give recommendations among K parties. Each party also has answers for every question, leading to K analogous profiles, \vec{p}_k . The questions relate to the set of policy issues that should determine the chosen party (Figure 1).

The screenshot shows the EUvox 2014 web interface. At the top, there is a logo with a checkmark and the text 'EUvox 2014', and links for 'About' and 'FAQ'. Below the header is a progress bar with a blue circle containing the number '5'. The main content area displays the question: 'The EU should redistribute resources from richer to poorer EU regions'. Below the question, there are five buttons for the Likert scale: 'Completely agree', 'Agree', 'Neither agree nor disagree', 'Disagree', and 'Completely disagree'. A sixth button, 'No opinion', is positioned below the other five.

Figure 1: A question in the EU-Vox 2014 [Agathokleous and Tsapatsoulis, 2016]

The recommended party is the one with the highest agreement score, which is calculated by fixed algorithms computing each question independently. First, answers are transformed into a vector with 1 on the chosen answer and 0 on the others. Second, each question is evaluated according to a *distance matrix* D emulating the comparison process. In this process, the score for each question, q_{ikj} , will be determined by users

¹Completely Agree, Agree, Neither Agree or Disagree, Disagree, Completely Disagree, No Opinion.

and parties' encoded answers (\vec{u}_{ij} and \vec{p}_{kj}) and the distance matrix (D). The way in which each policy issue is compared is chosen ex-ante according to proximity, directional or hybrid logic [Mendez, 2012 and 2017]². Once the preference in the policy issue of the user is compared with the positions of the political party, the outcome for that user in the question for each party, q_{ikj} , is calculated. Third, the total agreement score, s_{ik} , is computed by squashing the scores for the N questions according to some weights, \vec{w} , that represent the importance of each question. In many cases, VAAs weigh all questions equally, although it is common to allow users to coarsely tune the weights.

$$s_{ik} = \sum_{j=1}^N w_j q_{ijk} = \vec{w} \cdot \vec{q}_{ik} . \quad (1)$$

In every VAA, developers must unavoidably take some decisions that affect the agreement coefficient [Fossen and van den Brink, 2015]. First, they decide the set of policy issues considered to provide the agreement coefficient. Second, they place the political parties in these issues. Third, they elaborate a distance matrix for weighting the levels of agreement between users and parties' preferences in the issues. Lastly, they decide the weight of each of the policy issues for the final recommendation. Therefore, even the best imaginable developed VAA will not "simply reflect what is at stake in the election by neutrally passing along information" [Fossen and van den Brink, 2015, 341]. Rather, political information is structured by developers' decisions. These decisions of the developers should be evaluated to acknowledge whether the VAA is likely to increase voters' political competence.

In previous papers, it has been argued that VAAs' notion of political competence assumes an arguable conception of democracy [Fossen and Anderson, 2014] and that the decisions of the developers affect substantively to how the information is structured [Fossen and van den Brink, 2015]. However, the discussion on the methodology of VAAs have always focused on particular aspects of the application, such as whether the policy issues and the positions of the parties are correctly selected [Gemenis, 2012, Gemenis, 2013], the statement of the questions are neutral and understandable [Walgrave et al., 2009, Kamoen and Holleman, 2017] and the way of comparing preferences is adapted to a previous issue-voting theory [Mendez, 2012, Mendez, 2017]. Thanks to that, some of the decisions that the developers must take when designing a VAA have been improved. However, there is a lack of a comprehensive set of criteria to evaluate VAAs methods of recommendation. A normative set of criteria regarding whether the VAA is likely to increase voters' political competence can be helpful to evaluate whether the decisions of the developers are sound and justified. In the following, we establish a set of criteria to evaluate VAAs and argue that current methods of recommendation fail in some of them.

²These issue-voting theories are further explained in the next sections

4. Evaluating VAAs

VAA developers have previously discussed the criteria that should drive the development of these applications. So far, the main outcome of this ongoing discussion has been the *Lausanne Declaration on Voting Advice Applications*. The signers of this declaration do not try to establish what is the ideal kind of VAA, and rather “recommend certain standards and minimal requirements that should be respected by all the makers of VAAs” [Garzia and Marschall, 2014b, 227]. This has been the most serious intent of establishing a set of criteria for future VAAs. Our normative criteria are built on this declaration and have the objective of evaluating current and future VAAs’ capability of increasing voters’ political competence.

4.1. Informativeness

To help voters in their goal of voting competently, VAAs must be informative on those aspects of the application that can increase their notion of voter’s political competence. First, a VAA must inform users on the scope of the application and its limits to increase political competence. Therefore, a VAA must clearly state what are their assumptions regarding issue-voting theories as a valid proxy to vote competently. This is stated in the Lausanne Declaration in article 5.1: “VAAs are based on the assumption that users’ proximity to parties and candidates can be measured by their degree of accordance on political issue positions. Ideally, VAAs make this presumption visible” [Garzia and Marschall, 2014b, 228]. At this point, we have to distinguish between the ideal and the real world for VAAs. In the former, voters are issue-voters as predicted by Downs’ theory of 1957. In this world, a citizen is considered to be competent if she is able to place herself and political parties on a set of policy issues and aggregate distances in policy issues relying on the importance of each issue [Mendez, 2017]. In this case, VAAs appear as an optimal tool for increasing voter’s political competence as it provides an expertise placement of the policy issues and the positions of political parties in them, and it offers a supposedly scientific agreement method to offer an agreement score [van der Linden and Dufresne, 2017]. However, in the real world, we cannot assume entirely issue voters. Certainly, many voters use different strategies to vote for a party beyond their position in the policy issues: it has been empirically tested that voters use party identification, charismatic leadership or social cues to decide which party to vote [Mendez, 2012]. If there are voters not worried about their agreement with the political parties in the policy issues, it is certainly difficult to think that VAAs can increase their political competence in the sense we previously define it [Anderson et al., 2014]. As these voters do not fit in the conception of political competence that VAAs advance they are beyond the direct scope of the application. Therefore, VAAs should inform users that they are only able to provide a recommendation based on issue-voting and are unable to offer any other kind of advice.

Moreover, VAAs must include those features increasing their conception of political competence. We have previously argued that VAAs consider a competent voter such as

one that 1) compare their policy preferences with the positions of the political parties in a set of prominent policy issues, and 2) aggregate these disagreements in the policy issues in a meaningful way to decide which party to vote for. Therefore, VAAs should provide information on the policy issues at stake, the parties’ positions on these policy issues and a recommended party, which is the result of an aggregation of disagreement in the policy issues between users and political parties [Fossen and Anderson, 2014]. In the Lausanne Declaration, it is argued that the selected policy issues must be relevant and reveal the dimensions of political competition [Garzia and Marschall, 2014b]. Furthermore, it is stated that the positions of the political parties must be decided upon a series of methods such as “expert opinions”, “party manifestos”, and self-placement. As previous researchers have studied, the important point is that this has correspondence with the real positions of the political parties in the policy issues [Gemenis, 2013]. If VAAs provide that information to the user, it will facilitate the process of comparing and aggregating their levels of agreement with the parties in the policy issues. Therefore, we can set three sub-criteria to assess whether a VAA is informative to increase political competence.

1. The VAA informs the users of the scope and limitations of the application.
2. The VAA informs the user of the relevant policy issues at stake.
3. The VAA informs the user of the position of the political parties running for the election in these policy issues.

4.2. Respect for users’ way of comparing and aggregating policy preferences

Ideal VAAs are supposed to offer users “an unobstructed view of the political landscape, and their place within it” [Dinas et al., 2014, 291]. In this sense, unobstructed means that the recommendation depends on users’ preferences on policy issues. This idea can be considered in different ways. For example, making the whole process of recommendation without changing where the user has placed herself is uncontroverted as every VAA actually completes this operation because it takes policy preferences as given [Anderson et al., 2014]. However, this understanding of respect for the users’ preferences does not fully inform how users truly form their voting decisions by comparing and aggregating their policy preferences. For example, a VAA might not recommend the closest political party according to users’ own preferences if it does not give enough weight to a decisive policy issue according to users. If a VAA does not correctly represent the saliency of the policy issues, VAAs would evaluate how users should aggregate policy issues to cast their votes. Therefore, if there is a systematic bias on VAAs’ way of comparing and aggregating policy preferences, the goal of increasing voters’ political competence will be critically undermined. In that scenario, the recommendation would complicate instead of facilitating the already complex process of voting competently. In that case, users would have excellent reasons for not considering the recommendation.

Unfortunately, it is impossible to expect a VAA offering an entirely unobstructed view of users’ disagreement with political parties as the unavoidable decisions will affect how

the application represents the user and provides a recommendation [Fossen and van den Brink, 2015]. However, there can be levels of respect for users’ way of comparing and aggregating policy preferences relying on the method used. The design of the application will determine whether the developers of the application are considering how users aggregate and compare policy issues. The aggregation of policy issues is directly related to the saliency of the issue: more important policies for the users should have more weight than unimportant ones. Another alternative would be to put additional questions on that topic, but it seems unpractical due to the limitations of questions that any VAA confronts. There are at least two ways to take users into account regarding the saliency of the policy issues. First, VAAs can give users the opportunity to weigh the questions relying on their importance in choosing the voting party (i.e. set up a button in order to declare that a question is of special relevance). It is not feasible to allow users to entirely decide the weighting of the questions, as it would demand an impressive amount of time and knowledge on policy issues. Therefore, offering the option of weighting the questions will not entirely solve the problem of aggregating policy preference by respecting users’ way of doing it. Second, the weights of the question can be adapted to users following empirical methods and research. This can be mainly done in two different ways. First, using previous empirical data analysis that focuses on the importance of the policy issues (e.g. voter surveys). Second, using an algorithm that continuously learns to compare and aggregate policy issues to emulate better users’ way of doing it [Romero et al., 2020].

VAAs not only aggregate the policy issues, but they also have to compare the preferences of the users and the political parties. For that, they necessarily have to create some function of distances between users and parties’ positions [Mendez, 2017]. In a nutshell, they have to determine the value of the disagreement between parties and users: what happens if the user completely agrees with a topic and the party disagrees? This is a non-obvious question because it can be expected that different issues and circumstances create various ways of comparing [Dinas et al., 2016]. There are some issue-voting theories that can illuminate how users compare their policy preferences with the positions of the political parties, such as proximity and directionality [Mendez, 2012]. Depending on whether the distance functions are adapted to the users’ way of comparing policy issues, we can consider whether the VAA is accomplishing this criterion. A VAA using distance matrices that do not correspond to users’ way of comparing their preferences is unlikely to increase voters’ political competence, as it is not respectful for the users’ way of comparing them. According to this criterion, VAAs’ notion of political competence is neutral regarding how users compare policy issues and the saliency of the questions. Therefore, if VAAs aim to increase voters’ political competence, they should try to offer users a recommendation based on a user-based saliency and a users-way comparison of policy preferences.

1. The VAA establishes the saliency of the policy issues respecting users’ way of establishing it.
2. The VAA respects users’ way of comparing preferences in policy issues.

4.3. Reliability

To increase political competence, it has been claimed that VAAs must try to reflect the political reality of the world and be able to offer a (simplified) picture of the real world [Dinas et al., 2014]. If VAAs do not capture a reliable disagreement between users and political parties, they are unable to offer a recommendation worth consideration. This recommendation would be incapable of increasing voter’s political competence in the real world, and thus it could not be applied to any practical decision; it would only be valid for the imaginary-world of VAA developers. It has often been claimed that VAAs must be a mirror of the political landscape in which the user can get a complete look at her place within it [Dinas et al., 2014]. This claim is an oversimplification of VAAs that can lead to misunderstandings because the application has a very narrow conception of democracy and voting informed by social choice theory, spatial politics, and issue voting, which cannot claim to be an entirely realistic version of politics. Because many important voting factors are not considered by the application, VAAs will always fail to be entirely pure representations of political reality. For that reason, some authors have argued that, instead of mirrors of electoral reality, VAAs should be understood as dioramas [Fossen and van den Brink, 2015]. From our point of view, understanding VAAs as dioramas depict more accurately the scope of VAAs; this application cannot aspire to capture all the complex aspects of reality and must necessarily depict a not entirely realistic political landscape that is based on issue voting. Having said that, if VAAs aim to increase political competence they must aspire to be as reliable as possible.

This criterion has a relationship with the previous one: a VAA considering users’ way of comparing and aggregating preferences with a biased method will hardly be reliable. The main difference is that this criterion takes into consideration the political landscape that VAAs try to emulate. It accounts for whether VAAs offer a reliable and unbiased matching to issue voter users, and it does not misrepresent any political party. This criterion follows article 4.2 of the Lausanne Declaration: “VAA makers ought to carefully watch that the design does not favour a party/candidate in a systematic manner” [Garzia and Marschall, 2014b, 228]. Therefore, according to this criterion, a VAA will fail if it is mostly recommending a set of political parties that have no relationship with users’ expectations, and also if it over or under-represents some political parties substantively without sound justification. There are different ways to measure this condition of reliability that has been used in previous VAAs [Tsapatsoulis et al., 2015]. First, accuracy measures the percentage of predictions in accordance with users’ expectations: the VAA recommends the same party that the users expected to vote for before filling in the application. Second, the mean rank evaluates how high the VAA placed the preferred political party of the users. It offers complementary information to the accuracy measure, as it gives an idea of the position in which the previously preferred party is placed in the ranking of recommended political parties according to the agreement coefficient. Third, the confusion matrix assesses whether some political parties are over or under-represented by the VAA. A confusion matrix is a specific table layout that allows visualization of whether the recommended political parties correspond to those

expected by users. Fourth, *f-score* summarizes two key performance measures: *precision* (how many of the recommendations of each party were given to the right people) and *recall* (how many of the declared voters of each party were given it as a recommendation).

The reliability criterion faces two main objections. First, judging the VAA by how it meets an unfiltered set of users’ expectations about preferred voting parties may be wrong as the voting parties are determined by many factors that VAAs are not meant to help a voter consider. Certainly, in the real world, issue voters cannot be assumed entirely. Many voters use different strategies to vote for a party beyond their position on policy issues; it has been empirically shown that voters use party identification, charismatic leadership or social cues to decide which party to vote for [Mendez, 2012]. If there are voters who are not worried about their agreement with political parties on policy issues, VAAs cannot be expected to fulfill the expectations of these users by recommending a party that is in line with their policy preferences [Anderson et al., 2014]. Therefore, in these cases, it is problematic to negatively evaluate the fact that the VAA does not recommend the users’ expected party, as this situation may be due to factors beyond the control of the application. There are at least two ways to improve this situation. First, we can filter those users considering themselves “issue voters” (i.e. those declaring that they cast their vote “solely on policy positions of parties”) and evaluate the reliability of the VAA according only to these users. The problem with this approach is that citizens may have intertwined reasons for voting for a party, and even if they consider themselves “issue voters” there will probably be many factors that are not considered by VAAs. Another alternative would be to ask users to reply to the following question: “which party would you vote for insofar as your project is to base your vote solely on parties’ policy positions?” In that case, it can be evaluated whether the VAA is giving the users the parties that they would have chosen in the case of entirely being issue voters.

The second objection is that this criterion can undermine the informativeness of VAAs, as it can limit the amount of learning a VAA can provide to users. Indeed, a VAA always offering the same party that users expect would be useless in increasing political competence, as it would not give any valuable information to cast a competent vote. This VAA would not be informative on the disagreement between users and political parties as it would merely adjust to users’ wishes. This is not what this criterion aspires to. Instead, the reliability criterion informs on VAAs’ capability to draw a reliable political landscape based on a single premise: users’ political knowledge is imperfect but not total. Studies on VAA users have shown that they tend to be more interested in politics and have more knowledge about policy issues than the average citizen [Cedroni, 2010]. Assuming these knowledge conditions and “issue voters”, this criterion informs that a VAA will hardly be reliable if it mostly recommends marginal parties to a representative sample of users not previously interested in these parties. There are three considerations to be made regarding the previous statement. First, the reliability criterion must be balanced with the other criteria in order to evaluate VAAs (especially informativeness). Therefore, a VAA designer could decide to lose a little bit of reliability in order to increase the informativeness of the application or vice versa. Second, the realism criterion not

only considers whether VAAs recommend the expected party to the user (accuracy), but also the position of the expected party in the set of possible recommended parties (mean rank) and two other measures to know whether the VAA is over or under-valuating some political parties (confusion matrix and F-score). Therefore, it is a more comprehensive measure than a single comparison between users' expected and recommended party. Third, a reliable VAA does not compulsory need to have an excellent percentage of accuracy, F-score or mean rank because the role of this criterion is to preclude those VAAs having a percentage close to random matching or leading to a misrepresentation of particular parties without sound justifications. From this perspective, the reliability criterion can be used rather to exclude than to recommend any VAA method.

4.4. Transparency

According to the Lausanne Declaration, article 5.4, "following the principle of transparency, the algorithm matching users to parties and candidates should be documented and clearly explained to users" [Garzia and Marschall, 2014b, 238]. From our point of view, the criterion of transparency can be divided into two different subcriteria: openness and accountability. First, it is essential that VAAs are open in the sense of allowing users and researchers to access to their methods. This criterion means that the decisions taken by VAA developers should be accessible to both users and researchers. This is important in order to help the users understand why they have been recommended to a particular party. A closed method will make difficult any process of accountability by the developers of the application, who might have hidden interests or just rely on non-accurate methods. Not having open methods may prevent any replication and improvement of the methods. In that case, the reason behind a VAA recommendation may remain hidden and useless to any future research on these applications and political science.

Openness is a necessary condition, but it is not enough to advance accountability. We believe that a VAA method will be accountable if it allows researchers to understand the reasons behind a recommendation. Moreover, the recommendation process must be interpretable in light of previous theories and methods of political science and replicable by other researchers. Therefore, in an accountable VAA, all the decisions taken by the designers lead to a process of recommendation in which every step is sound and it is possible to understand what should be changed to have a different recommendation. For example, the weights of the questions and the distance functions determining how the disagreement in the policy preferences are compared must be determined from an understandable and sound process. A VAA making an aggregation in which the different policy issues cannot be disentangled will hardly be accountable. This is the case of the Support Vector Machine and the Neural Network Machine used by [Katakis et al., 2014]. In these proposed methods of recommendation, it is not possible to clearly evaluate the relationship between the recommendation and the policy issues. Other very sophisticated methods for recommending a party may suffer a problem of transparency by making impossible to really understand why a recommendation is given. This can be

a problem as the process of recommendation could be taking non-logical decisions and assuming unsound decisions that cannot be justified from the perspective of political science literature.

From the aforementioned, it should be clear that VAA methods should be understandable by researchers if sound conclusions from their usage are to be made. However, it is arguable whether users should understand the method of recommendation to consider that the VAA fulfills the transparency. On the one hand, a non-understandable method may undermine the goal of increasing users' political competence by providing a meaningful aggregation of their disagreement with the political parties. Users could feel that the application is useless if they do not understand the process and the accountability of VAA developers can be questioned. However, on the other hand, it is not clear whether VAAs can aspire to be understandable by non-politically sophisticated users. Even the simplest imaginable VAA assumes an imaginary world of conceptual spaces in which some spatial voting theory and maths must be used. This is hardly understood by anyone who is not a geek into political science. Therefore, it can be expected that any VAA will never be fully understood by the users, and consider that this is a genuine limitation of the application that cannot be transformed in a normative criterion to evaluate them. In that sense, it can be argued that VAAs are like cars: users do not need to understand their internal mechanism to be helped by them³. However, there are some important differences between cars and VAAs. We can expect that citizens treat VAAs differently than cars because these applications provide some political information that helps them to form an important decision in their life. In that sense, cars are different from VAAs as voting is not the same than transporting. It can be acceptable to use a car to move from one place to another without knowing anything about the internal mechanism of the car. However, users may have very good reasons to expect to know why an application recommends them to vote for a particular party. If users do not understand anything on the process of recommendation, two options may be expected: 1- A user refuses to accept the recommendation and political competence is not increased. 2- A user accepts the advice and votes the recommended party without understanding the reasons behind their decision. In that case, although it can be considered that the outcome of the application has been achieved, political competence would have not been increased.

Following the above, we argue that it is important for users to understand the method of recommendation and the logic employed by VAAs. However, we are conscious of the limitation of the application, and cannot expect users to fully understand VAAs. Therefore, we conclude by saying that this part of the criterion of transparency should not be part of the normative criterion to evaluate whether VAAs are likely to increase voters' political competence. It should rather be considered as a limitation that any VAA must confront when dealing with the problem of increasing voters' political competence. Therefore, we believe that the only two subcriteria to evaluate the transparency of a VAA are our previously defined conception of openness and accountability.

³We would like to thank Kostas Geminis to provide us with this idea.

1. VAA methods are open to the public and researchers.
2. VAAs methods are accountable, meaning that the reasons behind a recommendation can be interpreted and replicated by researchers.

5. Status Quo: Current VAAs

5.1. VAAs selection

For our analysis of the decisions that the developers take in the design of current VAAs, we have used previous papers that overview the state of research in VAAs. First, [Wagner and Ruusuvirta, 2012] analyses a series of thirteen VAAs launched in Europe from 2002 to 2009. Second, Garzia and Marschall [2014, chapter 3] analyzed VAAs launched until 2014. To be sure that we covered current VAAs, we have studied other VAAs launched since 2014, reaching to the conclusion that they all follow a similar logic than those studied by Garzia and Marschall. We exclude from our analysis the Social VAAs, which rely on the positions of the users who declare to vote for a party to make the recommendation [Katakis et al., 2014], and the very recent Learning VAA [Romero et al., 2020]. We made this because they are still in a very early stage, the methods have not yet been established, there are doubts regarding their likelihood to increase voters' political competence and their usage has been exceptionally limited [Katakis et al., 2014]. However, our criteria could also be used to evaluate whether these VAAs are likely to increase voter's political competence.

Moreover, to prove that there is a problem of reliability in current VAAs, we have replicated the method of recommendation of the VAA launched for the Parliamentary European Union Election in 2014, the EU-Vox 2014. This VAA has been designed by the same developers that most of the previous applications and their design are very similar regarding the critical points that we want to probe. Its data has been used for many other studies [Mendez, 2017, Ramos et al., 2017]. We have analyzed the results in eight countries: Spain, the Netherlands, France, Germany, United Kingdom, Poland, Sweden, Portugal, and Denmark. In the annex, there is a descriptive analysis of the users who filled the EU-Vox 2014 in each of the countries.

5.2. Explaining developers' decisions

In current VAAs, most important decisions (i.e. establishing the policy issues and positions of the parties, weighting policy issues and creating the distance function for measuring disagreement) are taken ex-ante by the developers [Louwerse and Rosema, 2014]. Parties' policy positions are generally agreed upon after a process of searching for information about the parties and a process of validation [Wagner and Ruusuvirta, 2012]. Moreover, all the questions are equally weighted by the developers for the aggregation whose outcome is the agreement score determining the recommended party [Mendez, 2012]. At most, many VAAs give users the opportunity to declare that a question is unimportant or very important for them [Louwerse and Rosema, 2014]. The effect of pointing

out that a question is of particular importance is also chosen ex-ante: weights are usually halved or doubled depending on the user’s declaration [Wagner and Ruusuvirta, 2012]. For example, if the user declares that a question regarding taxes is very important for her, the weight of that question will be multiplied by two. Therefore, current VAAs assume that users’ way of aggregating policy issues is the simplest imaginable: all policy issues are as important in deciding which party to vote for except if the user declares that a question is more or less important. In that case, a single multiplication/division is made without any evaluation of whether it is exactly that what the users mean when declaring that a question is more/less important. As far as our knowledge goes, there has been no single empirical study on VAAs justifying the decision to weigh all policy issues equally.

Regarding the comparison between users and parties’ policy positions, VAA developers choose ex-ante a matrix of distances between users and political parties and use the same matrix throughout the questionnaire [Mendez, 2012]. VAA developers have discussed whether to use proximity, directional or a hybrid distance function [Mendez, 2017]. This discussion echoes one of the fundamental disagreements about theories on issue voting that confront proximity and directional logic [Merrill III et al., 1999]. A proximity logic would mean that recommendations are based on voter-party distances, usually measured on a continuous scale [Downs, 1957]. In contrast, directional logic is indicated by three features. First, each policy issue has two different ‘sides’, for and against. Second, it is possible to express the intensity of preference on each of the sides of the spectrum (i.e. “agree” and “completely agree”). Third, parties are not punished for holding more intense positions than voters [Rabinowitz and Macdonald, 1989]. Therefore, defenders of the directional theory argue that voters mainly care about the fact that the political party is on their side of the ideological spectrum. For example, according to directional theory, a user declaring (“agree”) regarding euthanasia would only care that the political party is in those positions advancing euthanasia (“agree” and “completely agree”). In contrast, a proximity voter would give the same output to a party with the position “completely agree” and “neither agree or disagree”, as they both are at a distance of one from “agree.” Currently, distance matrices between users and parties’ positions following either proximity or directional logic are created to weight the disagreement for each policy issue [Mendez, 2012].

To evaluate whether VAAs are recommending a reliable set of parties, most VAA researchers have trusted in accuracy and mean rank between users’ expected and recommended party [Tsapatsoulis et al., 2015]. Those were the criteria used by [Mendez, 2012, Mendez, 2017] to acknowledge whether a proximity, a directional or a hybrid model was better fitting to VAA users. Some VAAs have tried to distinguish between issue voters and non-issue voters based on self-reporting [Mendez, 2017]. To our knowledge, when VAAs ask for which party the user prefers, it has never been explicitly emphasized that they should answer exclusively considering the policy positions of the parties. This addition would make the performance scores reliable, which is not currently the case.

5.3. Assessing the criteria

From our perspective, current VAAs use a correct method to choose the policy issues at stake and the positions of the political parties within them. Assuming VAA developers are policy experts acting in good faith, they can duly provide expertise on the issues at stake and the positions of the parties within them [Garzia and Marschall, 2014b]. Experts are better positioned than users to decide the policy issues and the positions of the parties because they have more knowledge of the political realm. Therefore, we believe that most current VAAs are informative in the sense that they inform the user of the policy issues, the positions of the political parties within them, and offer an agreement based on these differences, which simplifies the voting decision. However, they generally lack one of the subcriteria of informativeness as they do not inform on the limitations of the application and their assumptions relying on issue-voting and spatial politics. This has been addressed by [Fossen and Anderson, 2014] so it will not be further discussed here. Furthermore, we consider that most current VAAs are transparent as they have open methods that can be roughly understood by experts. Accountability in these circumstances can be easily achieved as it is straightforward to see the reasons behind the recommendation process. It is true that some VAAs do not publicise their distance matrix, but this is a failure that can be easily solved.

Our main concern regarding current VAAs is that they do not respect users' way of comparing and aggregating policy preferences. First, let us focus on the most controversial decision that VAA developers are taking: giving the same weight to every question. From our standpoint, this is far from respectful to users' way of aggregating policy preferences in order to come up with a single vote. It relies on an imaginary world in which voters give the same importance to every policy issue when casting a vote. This assumption is empirically flawed in issue voting as many studies have shown that voters consistently give more importance to some issues [Wlezien, 2005]. Beyond any empirical study, it is common sense to think that a VAA that weights all the issues equally will have problems in respecting users' own preferences. Indeed, it can be taken as given that voters care more about some specific policy issues when casting their votes, and this is not something that VAAs' vision of political competence should challenge. There is no way in which voter's political competence should be understood as the equalization of every policy issue to find which political party is closer to the user preferences "if aggregated in such a way." Therefore, current VAAs, instead of declaring that they recommend political parties to the users according to their own preferences, should claim that "given a world in which everyone cares exactly the same about a policy issue regarding euthanasia, one on taxes and another on immigration", VAAs can increase political competence. This is not solved by giving users the opportunity to add more/less weight to a question, as the multiplication/division by two is still arbitrary and can only marginally improve the respect toward users' own preferences. Is there any specific reason to multiply/divide by two and not another number?

A similar problem occurs with matrix distances. The method used to establish the

same matrix for every question according to proximity or directional theory is criticisable. First, it assumes that these theories are the only sources of inspiration for the construction of the matrix. This is an important decision that should be clarified by VAA developers, as there is an ongoing discussion in political science on whether voters compare their positions on the policy issues with those of the political parties following any proximity or directional theory [Dinas et al., 2016]. Second, even assuming that these theories are the most appropriate in order to build the matrix, it is not really clear that the same matrix can be used for every policy issue. For example, [Kropko and Banda, 2018] found that each policy issue follows its own logic, as they potentially respond to different patterns of behaviour that are difficult to disentangle ex-ante. Beyond empirical research, it is quite imaginable that users may be driven differently in their disagreement about visceral policy issues (such as some regarding social rights. e.g. homosexual rights) in which directional logic is more sound than in others regarding taxes, in which a proximity logic may be expected. Therefore, it is difficult to know ex-ante which theory of voting is better fitted to the users of the application. This can be seen in those studies that have tried to evaluate the distance matrix, such as those of Mendez [2012 and 2017]. Their main conclusion is that a hybrid matrix, not following any clear voting logic, fits better with users' policy preferences [Mendez, 2017]. Once accepted that an ex-ante established hybrid matrix better fits users' way of comparing their policy preferences with those of political parties, the central question that remains unanswered is why other distance matrices have not been tested.

Moreover, current VAAs do not perform properly at the criterion of reliability. First, VAA developers have only tried to improve accuracy and mean rank between users' expected and recommended party without differentiating between issue and non-issue voters [Louwerse and Rosema, 2014]. This is a poor way of improving the reliability of the application, as it does not consider all the other possible factors that can influence non-issue voters and does not differentiate between kinds of users. Second, in the scarce studies in which issue voters have been differentiated, accuracy and mean rank of VAAs have been below expectations [Mendez, 2017]. We have replicated the process of recommendation of the EU-Vox 2014 in eight countries to test whether it accomplishes the criterion of reliability. For example, in Spain, after differentiating between issue and non-issue voters by their responses to the supporting questions, we found that the EU-Vox 2014 recommended to issue voter users their previously preferred party in an astonishingly low percentage (approximately 25% accuracy and 3 in mean rank, meaning that the preferred party appeared on average in the third position in the set of recommended parties). To illustrate the importance of this point, we hypothesize a dumb algorithm that recommends everyone the party most prevalent in the overall voting intentions, then the second one and so on, regardless of users' and parties' answers to the questions. In Spain, such an algorithm obtained similar outcomes regarding accuracy (28%) and mean rank (3.44). This is not an exclusive problem of the EU-Vox 2014, because as we have argued this methodology is representative of other VAAs. Regarding the overall matching with political parties, we emulated the recommendation process for the Spanish issue voter users in the EU-Vox 2014 and found that some parties were strikingly

overestimated or underestimated by the application (figure 2 and figure 3). The results for the rest of the countries, which are similar, can be seen in the annex.

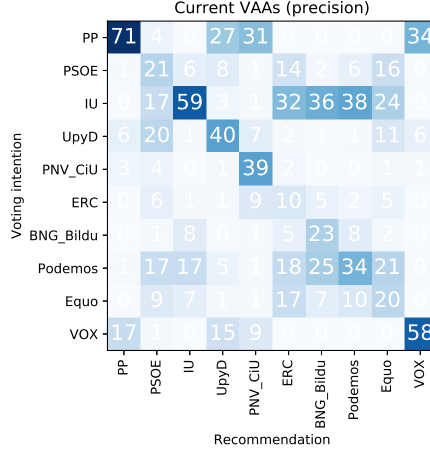


Figure 2: Confusion matrix between recommended and preferred parties. Darker colours represent a higher percentage of coincidence while lighter colours mean lower. The squares forming the diagonal line correspond to the cases in which the recommended and the preferred party were the same. The values are normalized by columns so the values of each column sum up to 100.

Figure 2 reveals two problems of current VAAs. First, low accuracy is reflected in the fact that the squares forming the diagonal line are not always darker than other squares in the same column. This means that some parties are mainly being recommended to users who intended to vote for another party. Second, the fact that the diagonal line is sharply heterogeneous means that parties are receiving different treatment by the application. This information can be complemented by the fact that the mean rank is also heterogeneous among the political parties, with some of them appearing in the last positions for their declared voters. For example, while most users receiving a recommendation to vote for the Partido Popular (PP) had declared that they would vote for that party, the opposite is true for their major competitor in 2014, the Partido Socialista Obrero Español (PSOE), whose recommendations match users' voting intentions very oddly. This situation is similar in the other studied countries as can be seen in the annex. This is problematic as VAAs should offer a reliable landscape of the political parties to all the users, and this can hardly be happening if the application significantly overvalues and undervalues some political parties without sound justification. As can be seen in figure 3, some parties are outstandingly over-represented by the application (for example Esquerra Republicana de Catalunya (ERC) and Podemos) while others such as Izquierda Unida (IU) are ignored. The remaining question is the following: Is there any possible justification for benefiting some parties over others in such a remarkable way? We believe that this is not the case, at least if we take as given that they are issue voters and the knowledge on politics of VAA users is more or less evenly distributed

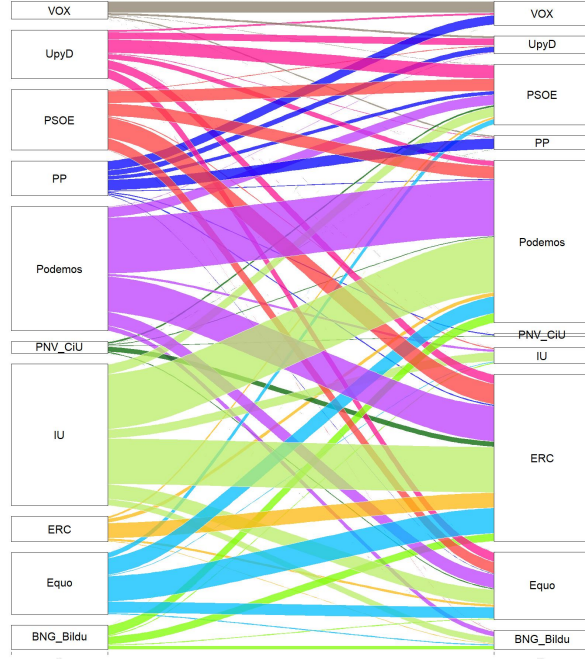


Figure 3: Alluvial Chart showing the preferred parties of users on the left side and the recommended party provided by the VAA on the right side for Spain in the EU-Vox 2014.

between voters of different parties. Therefore, matching among users declaring different initial voting parties should not totally differ when a large and representative sample of voters of the party is taken, as was the case with PP and PSOE declared voters in the EU-Vox 2014. Some differences between the parties may be acceptable (e.g. most users could mistakenly conceive the position of their preferred party on some policy issues and therefore lead to a misrepresentation of that party), but a confusion matrix with such a heterogeneous diagonal line is problematic and suggests a consistent bias towards some political parties due to the VAA method.

In short, this confusion matrix shows the flaws of current VAA methods. From our perspective, the main problem in the EU-Vox 2014 was in ignoring the user saliency and the way of comparing their preferences. However, it is worth exploring whether the political parties are well-positioned, although there is already promising work on the subject [Gemenis, 2013]. Again, this lack of reliability is not an exclusive problem from the EU-Vox 2014 as this methodology is representative of most VAAs. This problem of misrepresentation of political parties is acknowledged by previous VAA researchers. For example, worried about how parties are positioned in the application, [Gemenis, 2013, 281] reaches to the conclusion that “some parties may find themselves in the beneficial position of receiving more voting recommendations because of artefacts in the VAA construction.”

To sum up, all the symptoms suggest a lack of reliability; accuracy and mean rank are close to random matching, and there is a misrepresentation of the political landscape clearly benefiting some parties. Our results suggest a problematic situation in current VAAs in which the respect for the users’ preferences and the reliability of the application can be questioned. Considering the above, we need to find out why, despite these critical problems in increasing political competence, VAA developers have decided to give a priori the same saliency to all the policy issues, use a predetermined matrix distance, and not differentiate between issue and non-issue voters. A possible justification is that they did not find any other way to do it better. Every possible decision set ex-ante without empirical validation will be liable to the same criticism made before and, in that case, it seems logical to equally weight all the questions to give an “appearance of neutrality” to the method. Therefore, giving the same weight to the questions may be seen as preferable to any other ex-ante non-empirical weighting decision, and following this argument, the problem is intrinsic to the ex-ante non-empirical logic employed by VAAs to weight the questions.

6. Conclusion

In this paper, we have argued that current VAAs are not adequate to increase voters’ political competence. First, we have established and discussed four criteria to evaluate whether a VAA is likely to increase political competence: informativeness, respect for users’ way of comparing and aggregating policy preferences, reliability, and transparency. For that, we have used as a starting point the Lausanne Declaration on VAAs, which was made by the own developers of the application with the objective of establishing the correct methods for designing the application. Second, we have argued and empirically proved that current VAAs fail fundamentally in two of them: respect for users’ way of comparing and aggregating policy issues and reliability. Therefore, this research tackles the problem of previous VAAs by pointing out the necessity of looking for different methods of recommendation that can improve their performance.

The remaining question is whether other methods that empirically establish the weights of the policy issues and the matrix distances can solve the current VAA problems to increase voters’ political competence. One option is to find an empirical way to establish ex-ante the weight of the questions (i.e. to adopt the result of other surveys to establish the weights of the policy issues). This would be an interesting idea worth discussion by VAA developers, but the difficulties with current survey data are enormous, as there is no available information on the exact value that users give to the thirty policy issues used by VAAs. There is another alternative that needs to be developed further in the future: a learning algorithm that adapts the weights of the questions and the distance functions to users’ way of comparing and aggregating policy issues. There are two reasons why we believe that a learning method is superior. First, VAA users suffer from a self-selection bias and they are not representative of other surveys’ samples [Cedroni, 2010]. There-

fore, in order to get a representative sample of VAA users to adapt the questions and the distance functions, we should use data from VAA surveys, which in their current state are totally insufficient to carry out such a sophisticated analysis. Second, other non-learning methods would have problems if the users of the application are not as previously expected. For example, this can happen if some citizens who previously were not using VAAs start filling in the application because of an external factor. In that case, a learning method could be adapted to these new users, but a pre-established method could not. In any case, our research suggests that some changes need to be made in future VAAs for accomplishing their goal of increasing voters' political competence.

In this sense, Social VAAs using learning methods might be a solution to the problems aforementioned. They have already been proposed by previous researchers, but their results are still provisional and their usage limited [Katakis et al., 2014, Tsapatsoulis et al., 2015, Agathokleous and Tsapatsoulis, 2016]. However, it should be studied whether these methods, which are various and have many nuances, would be able to accomplish the normative criteria established in this article. Moreover, other alternative methods of VAA appear as promising alternatives to current methods of recommendation according to our normative set of criteria. For example, a learning method of VAA adapting the weights of the questions and the distance functions to compare the levels of disagreement seems like a promising option for new VAAs [Mendez, 2017, Romero et al., 2020]. Given the failures of current VAAs to improve voters' political competence, new methodologies should be tested to further advance on the possibilities of these applications to increase voters' political competence.

References

- [Agathokleous and Tsapatsoulis, 2016] Agathokleous, M. and Tsapatsoulis, N. (2016). Applying hidden markov models to voting advice applications. *EPJ Data Science*, 5(1):34.
- [Alvarez et al., 2014] Alvarez, R. M., Levin, I., Mair, P., and Trechsel, A. (2014). Party preferences in the digital age: The impact of voting advice applications. *Party Politics*, 20(2):227–236.
- [Anderson et al., 2014] Anderson, J., Andreadis, I., Dumont, P., De Angelis, A., Fivas, J., Fossen, T., Gemenis, K., Van Ham, C., Kies, R., Krouwel, A., et al. (2014). *Matching Voters with Parties and Candidates: Voting Advice Applications in Comparative Perspective*. Ecpr Press.
- [Brennan, 2011] Brennan, J. (2011). The right to a competent electorate. *The philosophical quarterly*, 61(245):700–724.
- [Caplan, 2011] Caplan, B. (2011). *The myth of the rational voter: Why democracies choose bad policies*. Princeton University Press.
- [Carpini and Keeter, 1996] Carpini, M. X. D. and Keeter, S. (1996). *What Americans know about politics and why it matters*. Yale University Press.
- [Cedroni, 2010] Cedroni, L. (2010). *Voting Advice Applications in Europe: The state of the art*. Scriptaweb.
- [Dahl, 1992] Dahl, R. A. (1992). The problem of civic competence. *Journal of Democracy*, 3(4):45–59.
- [Dinas et al., 2016] Dinas, E., Hartman, E., and van Spanje, J. (2016). Dead man walking: The affective roots of issue proximity between voters and parties. *Political Behavior*, 38(3):659–687.
- [Dinas et al., 2014] Dinas, E., Trechsel, A. H., and Vassil, K. (2014). A look into the mirror: preferences, representation and electoral participation. *Electoral studies*, 36:290–297.
- [Djouvas et al., 2016] Djouvas, C., Mendez, F., and Tsapatsoulis, N. (2016). Mining online political opinion surveys for suspect entries: An interdisciplinary comparison. *Journal of Innovation in Digital Ecosystems*.
- [Downs, 1957] Downs, A. (1957). An economic theory of political action in a democracy. *Journal of political economy*, 65(2):135–150.
- [Ferejohn and Kuklinski, 1990] Ferejohn, J. A. and Kuklinski, J. H. (1990). *Information and democratic processes*. Univ of Illinois Pr.

- [Fossen and Anderson, 2014] Fossen, T. and Anderson, J. (2014). What’s the point of voting advice applications? competing perspectives on democracy and citizenship. *Electoral Studies*, 36:244–251.
- [Fossen and van den Brink, 2015] Fossen, T. and van den Brink, B. (2015). Electoral dioramas: On the problem of representation in voting advice applications. *Representation*, 51(3):341–358.
- [Garzia and Marschall, 2012] Garzia, D. and Marschall, S. (2012). Voting advice applications under review: the state of research. *International Journal of Electronic Governance*, 5(3-4):203–222.
- [Garzia and Marschall, 2014a] Garzia, D. and Marschall, S. (2014a). The lausanne declaration on voting advice applications. *W: D. Garzia, S. Marschall (red.). Matching voters with parties and candidates. Voting advice applications in comparative perspective*, pages 227–228.
- [Garzia and Marschall, 2014b] Garzia, D. and Marschall, S. (2014b). *Matching Voters with Parties and Candidates: Voting Advice Applications in a Comparative Perspective*. Number January.
- [Garzia et al., 2017] Garzia, D., Trechsel, A. H., and De Angelis, A. (2017). Voting advice applications and electoral participation: a multi-method study. *Political Communication*, 34(3):424–443.
- [Gemenis, 2012] Gemenis, K. (2012). A new approach for estimating parties’ positions in voting advice applications. In *Workshop, Cyprus University of Technology*, volume 23, page 24.
- [Gemenis, 2013] Gemenis, K. (2013). Estimating parties’ policy positions through voting advice applications: Some methodological considerations. *Acta politica*, 48(3):268–295.
- [Hardin, 2006] Hardin, R. (2006). Ignorant democracy. *Critical Review*, 18(1-3):179–195.
- [Kamoen and Holleman, 2017] Kamoen, N. and Holleman, B. (2017). I don’t get it: Response difficulties in answering political attitude statements in voting advice applications. In *7th International European Survey Research Association Conference*, volume 11, pages 125–140. European Survey Research Association.
- [Katakis et al., 2014] Katakis, I., Tsapatsoulis, N., Mendez, F., Triga, V., and Djouvas, C. (2014). Social voting advice applications-definitions, challenges, datasets and evaluation. *IEEE Transactions on Cybernetics*.
- [Kropko and Banda, 2018] Kropko, J. and Banda, K. K. (2018). Issue scales, information cues, and the proximity and directional models of voter choice. *Political Research Quarterly*, page 1065912918760729.

- [Ladner and Pianzola, 2015] Ladner, A. and Pianzola, J. (2015). Voting advice applications. In *Encyclopedia of Information Science and Technology, Third Edition*, pages 6427–6436. IGI Global.
- [Louwerse and Rosema, 2014] Louwerse, T. and Rosema, M. (2014). The design effects of voting advice applications: Comparing methods of calculating matches. *Acta politica*, 49(3):286–312.
- [Mahéo, 2017] Mahéo, V.-A. (2017). Information campaigns and (under) privileged citizens: An experiment on the differential effects of a voting advice application. *Political Communication*, 34(4):511–529.
- [Marschall and Garzia, 2014] Marschall, S. and Garzia, D. (2014). Voting advice applications in a comparative perspective: an introduction. *Matching Voters with Parties and Candidates*. ECPR Press, Colchester, pages 1–10.
- [Marschall and Schmidt, 2010] Marschall, S. and Schmidt, C. K. (2010). The impact of voting indicators: the case of the german wahl-o-mat. *Voting advice applications in europe. The state of the art*, pages 61–86.
- [Mendez, 2012] Mendez, F. (2012). Matching voters with political parties and candidates: an empirical test of four algorithms. *International Journal of Electronic Governance*.
- [Mendez, 2017] Mendez, F. (2017). Modeling proximity and directional decisional logic: What can we learn from applying statistical learning techniques to VAA-generated data? *Journal of Elections, Public Opinion and Parties*.
- [Merrill III et al., 1999] Merrill III, S., Merrill, S., and Grofman, B. (1999). *A unified theory of voting: Directional and proximity spatial models*. Cambridge University Press.
- [Rabinowitz and Macdonald, 1989] Rabinowitz, G. and Macdonald, S. E. (1989). A directional theory of issue voting. *American Political Science Review*, 83(1):93–121.
- [Ramos et al., 2017] Ramos, J., Padilla, J., and Chueca, E. (2017). Icts and democratic deficit in the eu: how vaa can help improve the democratization of eu institutions and practices: the case of eu-vox 2014, european elections in spain. In *Proceedings of the International Conference on Electronic Governance and Open Society: Challenges in Eurasia*, pages 70–77. ACM.
- [Ramos et al., 2019] Ramos, J., Padilla, J., and Chueca, E. (2019). Abstentionism, voting advice applications and voting activation. *Statistics, Politics and Policy*, 10(1):55–85.
- [Romero et al., 2020] Romero, G., Padilla, J., and Chueca, E. (2020). Learning vaas: A new method for matching users to parties in voting advice applications. *Journal of Elections, Public Opinion and Parties*, Forthcoming.

- [Somin, 1998] Somin, I. (1998). Voter ignorance and the democratic ideal. *Critical Review*, 12(4):413–458.
- [Stuart Mill, 1859] Stuart Mill, J. (1859). On liberty. *Collected works of John Stuart Mill*, pages 259–340.
- [Tsapatsoulis et al., 2015] Tsapatsoulis, N., Agathokleous, M., Djouvas, C., and Mendez, F. (2015). On the design of social voting recommendation applications. *International Journal on Artificial Intelligence Tools*, 24(03):1550009.
- [van der Linden and Dufresne, 2017] van der Linden, C. and Dufresne, Y. (2017). The curse of dimensionality in Voting Advice Applications: reliability and validity in algorithm design. *Journal of Elections, Public Opinion and Parties*, 27(1):9–30.
- [Wagner and Ruusuvirta, 2012] Wagner, M. and Ruusuvirta, O. (2012). Matching voters to parties: Voting advice applications and models of party choice. *Acta politica*, 47(4):400–422.
- [Walgrave et al., 2009] Walgrave, S., Nuytemans, M., and Pepermans, K. (2009). Voting aid applications and the effect of statement selection. *West European Politics*, 32(6):1161–1180.
- [Wlezien, 2005] Wlezien, C. (2005). On the salience of political issues: The problem with ‘most important problem’. *Electoral Studies*, 24(4):555–579.

A. Dataset & Cleaning

The data was retrieved from the repository of the EU-Vox 2014 before it stopped being available. From that original data, a subset of the users was selected. In order to obtain a clearer picture, only those users that had answered all the 30 items were considered disregarding those that stopped the survey before reaching the end. This condition is critical to ensure that the evaluated method of party matching is operated over observations with complete responses to the party items. Following Andreadis (2012, 2014), further filters were imposed upon the subset of observations. This removal was done with the intention of diminishing the potential impact of rogue users that did not fulfill the survey in a timely manner or that shown patterns of disregard when answering the questions in the survey. To this end, the following filters were imposed on the original retrieval of the VAA data.

1. Users whose time to complete the items section took under 120 seconds were removed.
2. Users with response time for an item under 1 second were removed.
3. Users who had responded 3 or more item questions under 2 seconds were also removed.
4. Users whose age were stated to had taken place prior to 1920 were removed.
5. Users whose ages were stated to be below 17 were removed.
6. Users who completed the survey using either the Android or the iPhone platform were removed.

Filters 1 to 3 intend to ensure that enough care has been put into the survey to prove that a level of attention and care sufficient to guarantee well-thought trough response was being taken. Filter 4 intends to ensure that users were stating the right age when filling the survey. While filter 5 intends to remove users that could not have been able to vote in the EU 2014 elections to enter the sample. Lastly, filter 6 is devised to correct the failure to present the questions appropriately in the mobile device version of the survey.

The resulting composition of the survey variable is as follows in the table below:

Country	Freq.	Percent
Denmark	60,975	26.18
England	36,557	15.7
France	3,414	1.47
Netherlands	684	0.29
Northern_Ireland	857	0.37
Poland	31,388	13.48
Portugal	19,681	8.45
Scotland	3,685	1.58
Spain	71,918	30.88
Sweden	1,585	0.68
Wales	2,121	0.91
Total	232,865	100

Figure 4: Number of Users per country.

Country	Age Groups									Total
	18-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-94	
Denmark	2,519	12,842	9,213	9,149	7,426	4,947	1,598	158	4	47,856
England	5.26%	26.83%	19.25%	19.12%	15.52%	10.34%	3.34%	0.33%	0.01%	100%
	1,853	10,790	6,762	5,339	5,081	4,515	1,807	224	4	36,375
France	5.09%	29.66%	18.59%	14.68%	13.97%	12.41%	4.97%	0.62%	0.01%	100%
	376	1,259	538	407	404	323	76	12	0	3,395
Netherlands	11.08%	37.08%	15.85%	11.99%	11.9%	9.51%	2.24%	0.35%	0%	100%
	26	296	112	85	91	56	12	4	0	682
Northern_Ireland	3.81%	43.4%	16.42%	12.46%	13.34%	8.21%	1.76%	0.59%	0%	100%
	63	341	159	132	90	57	10	3	0	855
Poland	7.37%	39.88%	18.6%	15.44%	10.53%	6.67%	1.17%	0.35%	0%	100%
	3,040	15,394	4,736	2,356	2,588	2,404	494	72	1	31,085
Portugal	9.78%	49.52%	15.24%	7.58%	8.33%	7.73%	1.59%	0.23%	0%	100%
	863	6,809	5,360	2,904	1,980	1,320	349	47	1	19,633
Scotland	4.4%	34.68%	27.3%	14.79%	10.09%	6.72%	1.78%	0.24%	0.01%	100%
	131	1,217	786	580	488	354	107	12	0	3,675
Spain	3.56%	33.12%	21.39%	15.78%	13.28%	9.63%	2.91%	0.33%	0%	100%
	4,091	28,234	17,497	9,734	8,003	3,394	601	79	6	71,639
Sweden	5.71%	39.41%	24.42%	13.59%	11.17%	4.74%	0.84%	0.11%	0.01%	100%
	164	714	276	197	151	63	14	1	0	1,580
Wales	10.38%	45.19%	17.47%	12.47%	9.56%	3.99%	0.89%	0.06%	0%	100%
	98	585	373	321	295	317	117	7	2	2,115
Total	4.63%	27.66%	17.64%	15.18%	13.95%	14.99%	5.53%	0.33%	0.09%	100%
	13,224	78,481	45,812	31,204	26,597	17,750	5,185	619	18	218,890
	6.04%	35.85%	20.93%	14.26%	12.15%	8.11%	2.37%	0.28%	0.01%	100%

Figure 5: Age Composition per country.

Gender			
Country	0	1	Total
Denmark	16,918	32,050	48,968
	34.55%	65.45%	100%
England	10,872	25,459	36,331
	29.92%	70.08%	100%
France	875	2,518	3,393
	25.79%	74.21%	100%
Netherlands	157	526	683
	22.99%	77.01%	100%
Northern_Ireland	219	636	855
	25.61%	74.39%	100%
Poland	5,994	25,249	31,243
	19.19%	80.81%	100%
Portugal	6,554	13,016	19,570
	33.49%	66.51%	100%
Scotland	1,133	2,529	3,662
	30.94%	69.06%	100%
Spain	19,262	52,196	71,458
	26.96%	73.04%	100%
Sweden	429	1,135	1,564
	27.43%	72.57%	100%
Wales	647	1,454	2,101
	30.79%	69.21%	100%
Total	63,060	156,768	219,828
	28.69%	71.31%	100%

Figure 6: Gender Composition per country.

B. Extended results

B.1. Confusion Matrix every country

	FolkeModEU	Folkeparti	Konservative	LiberalAlliance	RadikaleVenstre	Socialdemok	Socialistisk	Venstre
FolkeModEU	33	7	2	2	1	2	3	2
Folkeparti	31	77	18	19	3	10	6	18
Konservative	0	5	16	12	2	1	0	17
LiberalAlliance	2	5	10	36	2	1	0	10
RadikaleVenstre	2	0	13	6	45	19	17	11
Socialdemok	4	3	13	4	25	39	21	13
Socialistisk	28	2	5	4	15	25	51	5
Venstre	0	2	24	17	7	3	1	23

Figure 7: Confusion Matrix with the originally chosen parties in the Y axis and the VAA recommended parties in the X axis for Denmark.

	Conservatives	Greens	Labour	LibDems	UKIP
Conservatives	36	1	2	4	8
Greens	17	68	52	33	5
Labour	8	15	23	25	0
LibDems	7	14	20	36	0
UKIP	31	1	3	2	86

Figure 8: Confusion Matrix with the originally chosen parties in the Y axis and the VAA recommended parties in the X axis for England.

	Debout	EELV	FG	FN	FV	Modem	NPA	PS	RC	UMP
Debout	28	2	2	53	22	1	2	1	6	7
EELV	6	77	51	1	0	30	54	68	19	8
FG	0	2	24	1	0	1	15	5	0	1
FN	35	0	0	21	17	1	0	0	6	3
FV	13	1	0	13	22	3	2	1	0	9
Modem	6	5	2	4	11	38	4	7	19	30
NPA	2	5	17	1	0	0	23	1	0	0
PS	3	7	4	1	0	9	0	15	19	5
RC	0	1	0	0	0	1	0	0	6	0
UMP	7	1	0	4	28	16	0	2	25	37

Figure 9: Confusion Matrix with the originally chosen parties in the Y axis and the VAA recommended parties in the X axis for France.

	50PLUS	CDA	CU.SGP	D66	GL	PvdA	PvdD	PVV	SP	VVD
50PLUS	0	6	4	2	2	0	10	0	10	4
CDA	0	13	8	5	0	2	0	0	0	12
CU.SGP	0	10	25	1	1	2	0	2	0	1
D66	0	10	4	47	7	15	0	2	0	38
GL	0	16	4	24	78	47	40	0	24	0
PvdA	0	3	0	9	7	18	10	2	4	3
PvdD	0	13	17	8	2	9	20	10	24	4
PVV	33	10	21	0	0	2	10	67	4	5
SP	67	0	8	1	2	2	0	13	29	3
VVD	0	19	8	4	0	4	10	4	4	31

Figure 10: Confusion Matrix with the originally chosen parties in the Y axis and the VAA recommended parties in the X axis for The Netherlands.

	Europa.	NP	PiS	PO	PR	PSL	RN	SLD.UP	SP
Europa.	82	6	4	45	10	23	2	54	7
NP	2	73	8	3	23	6	25	2	7
PiS	1	3	40	8	25	22	7	5	35
PO	3	1	5	28	9	19	1	11	9
PR	1	5	11	3	17	4	4	2	11
PSL	1	1	5	4	5	13	1	3	6
RN	0	10	20	1	8	5	58	1	18
SLD.UP	10	0	1	9	1	6	0	22	4
SP	0	0	5	1	3	2	1	1	3

Figure 11: Confusion Matrix with the originally chosen parties in the Y axis and the VAA recommended parties in the X axis for Poland.

	Conservatives	Greens	Labour	LibDems	SNP	UKIP
Conservatives	38	1	3	5	3	9
Greens	8	80	44	23	67	9
Labour	8	9	26	28	7	1
LibDems	7	9	23	42	6	1
SNP	0	1	2	1	12	1
UKIP	38	0	3	2	5	79

Figure 12: Confusion Matrix with the originally chosen parties in the Y axis and the VAA recommended parties in the X axis for Scotland.

	Center	Feminist	Folk	Kristdemok	Miljo	Moder	Pirat	Socialdemok	Sverigedemok	Vanster
Center	50	2	17	24	1	32	15	1	3	1
Feminist	2	33	7	3	28	6	13	29	1	10
Folk	5	1	24	3	0	13	5	5	1	0
Kristdemok	4	0	4	24	0	9	2	2	1	0
Miljo	10	10	4	3	25	3	7	7	1	3
Moder	7	0	16	14	2	22	5	2	1	0
Pirat	6	1	2	3	0	2	13	0	2	1
Socialdemok	9	15	22	10	22	8	3	25	1	5
Sverigedemok	2	1	3	12	1	6	4	6	83	1
Vanster	6	37	2	2	19	2	33	21	6	79

Figure 13: Confusion Matrix with the originally chosen parties in the Y axis and the VAA recommended parties in the X axis for Sweden.

	Alianca	BE	CDU	Livre	MPT	PAN	PCTP	PS
Alianca	53	0	1	2	11	4	3	2
BE	4	39	31	28	17	24	24	18
CDU	3	10	23	7	14	11	18	9
Livre	6	25	17	33	14	23	12	29
MPT	4	1	2	2	5	3	1	3
PAN	0	0	0	0	0	0	0	0
PCTP	1	7	16	2	6	7	33	2
PS	29	16	10	26	32	27	9	36

Figure 14: Confusion Matrix with the originally chosen parties in the Y axis and the VAA recommended parties in the X axis for Portugal.

B.2. Allubial every country

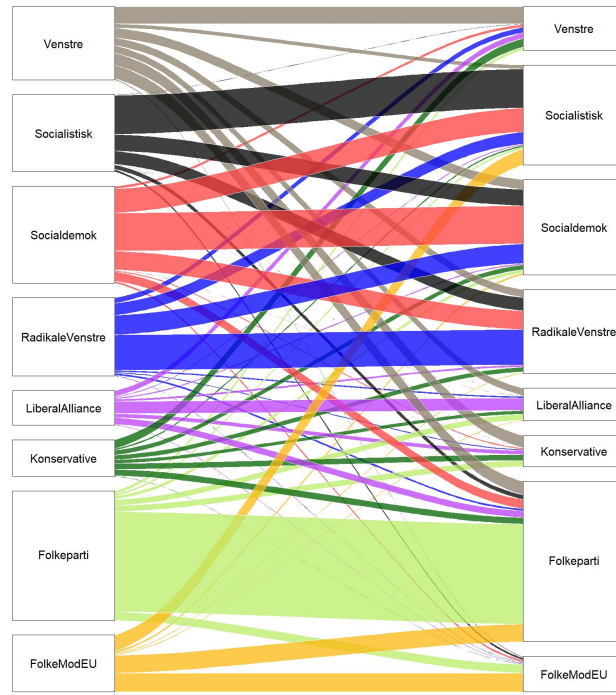


Figure 15: Alluvial Chart showing the preferred parties for the Danish users of the EU-Vox 2014 on the left side and the recommended party on the right side for Denmark.

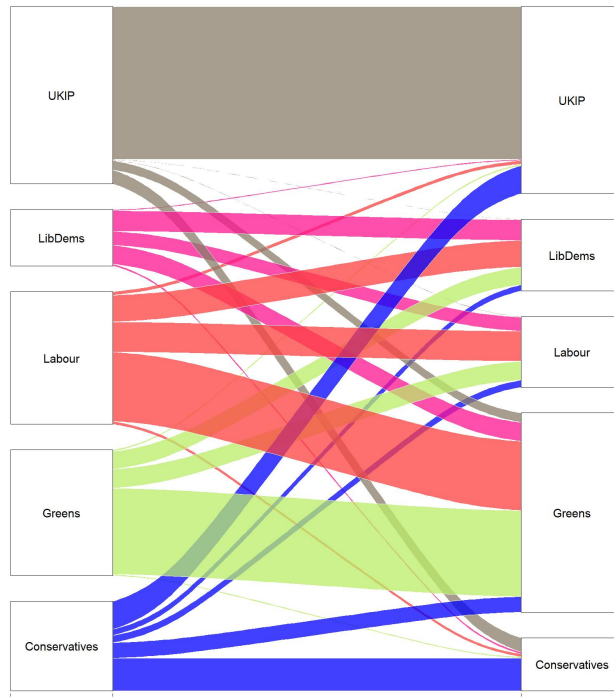


Figure 16: Alluvial Chart showing the preferred parties for the English users of the EU-Vox 2014 on the left side and the recommended party on the right side for England.

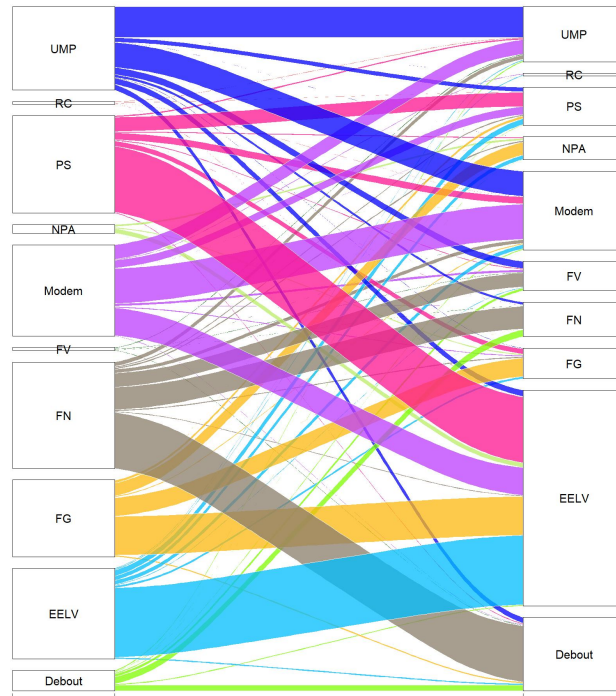


Figure 17: Alluvial Chart showing the preferred parties for the French users of the EU-Vox 2014 on the left side and the recommended party on the right side for France.

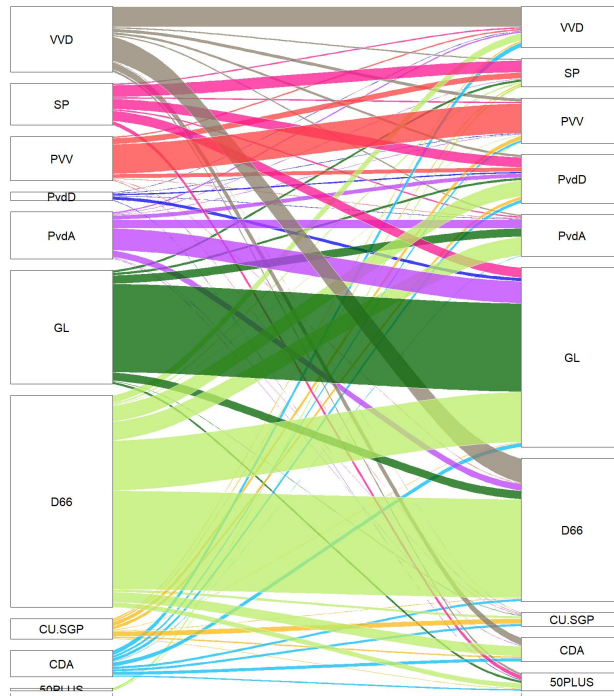


Figure 18: Alluvial Chart showing the preferred parties for the Dutch users of the EU-Vox 2014 on the left side and the recommended party on the right side for Netherlands.

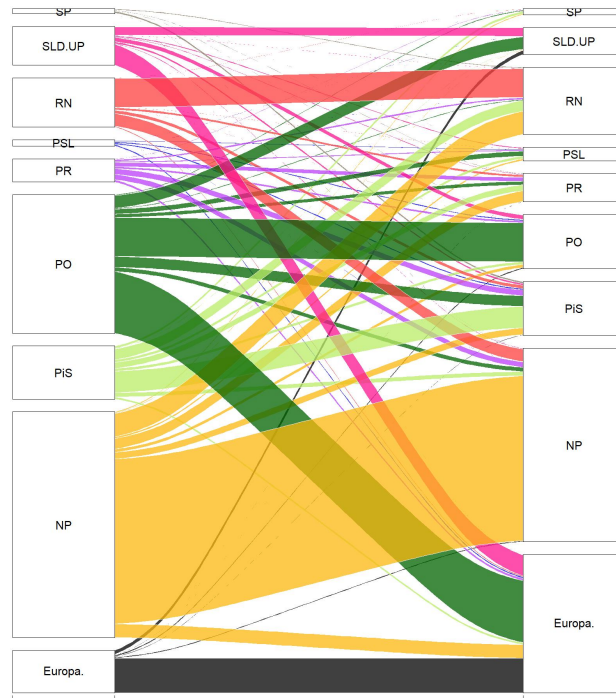


Figure 19: Alluvial Chart showing the preferred parties for the Polish users of the EU-Vox 2014 on the left side and the recommended party on the right side for Poland.

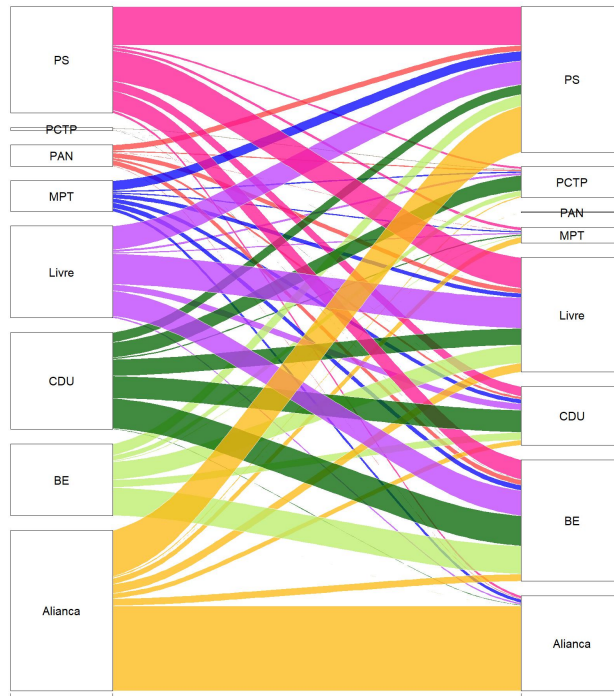


Figure 20: Alluvial Chart showing the preferred parties for the Portuguese users of the EU-Vox 2014 on the left side and the recommended party on the right side for Portugal.

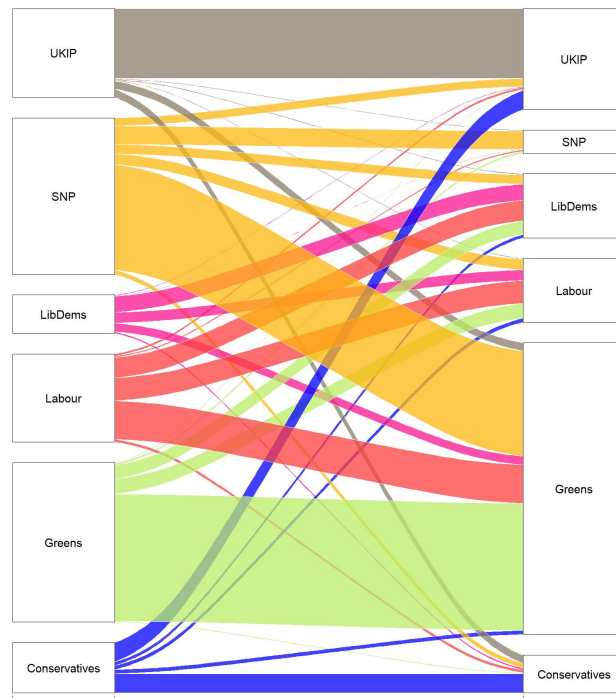


Figure 21: Alluvial Chart showing the preferred parties for the Scottish users of the EU-Vox 2014 on the left side and the recommended party on the right side for Scotland.

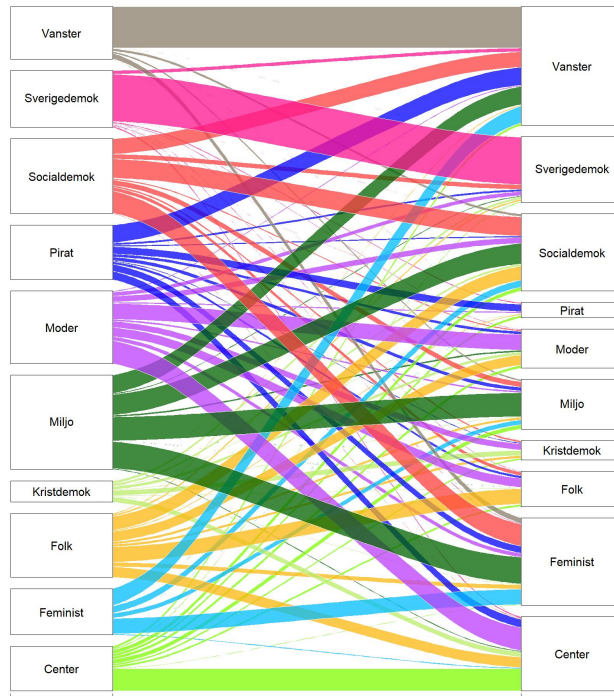


Figure 22: Alluvial Chart showing the preferred parties for the Swedish users of the EU-Vox 2014 on the left side and the recommended party on the right side for Sweden.