Chapter X: Comparing methods for estimating parties' positions in Voting Advice Applications

Kostas Gemenis and Carolien van Ham[1]

1 Introduction

Despite the proliferation of Voting Advice Applications (VAAs), there has been limited interest in researching how the positions of parties (and/or candidates) are estimated in VAAs. This is quite surprising, considering that there are many competing methods for doing so (Laver 2001; Volkens 2007) and that the quest for identifying the most appropriate one in terms of validity and reliability has sparked extensive debates in political science. With some notable exceptions (see Gemenis 2013; Krouwel and van Elfrinkhof 2013), this debate has not permeated the VAA research community. This is unfortunate for two reasons. First of all, VAAs have potential consequences for voting behaviour (Ladner et al. 2012; Walgrave et al. 2008; Wall et al. 2012), and hence considering the reliability and validity of party position estimates used in VAAs is of vital importance to evaluate their quality as voter information tools. Secondly, VAAs generate a wealth of party position estimates on a variety of policy issues that could be potentially useful in answering questions of interest to political science outside the domain of VAAs (Gemenis 2013; Hansen and Rasmussen 2013; Krouwel 2012; Wagner and Ruusuvirta 2012; Wheatley et al. 2012).

With this paper we aim to contribute to the emerging debate by performing a direct comparison of four popular methods used to estimate party positions in VAAs. Specifically, we compare party self-placement, the conventional expert survey, the iterative method between party self-placement and expert coding proposed by *Kieskompas*, and the Delphi method as applied to the use of experts. Using data from the 2012 Dutch parliamentary election, we compare the competing methods in terms of their ease of use, the degree to which they provide estimates that have face validity, and where appropriate, in terms of inter-coder agreement. Our conclusions have implications for both VAA designers and third-party users of VAA party position data.

[1] Department of Public Administration, University of Twente. E-mail: k.gemenis@utwente.nl and c.t.vanham@utwente.nl. We would like to thank Martin Rosema for his invaluable help in designing the empirical study, André Krouwel for providing the party self-placement data, as well as all the colleagues who participated in the expert survey and the Delphi estimation. For the latter, we acknowledge the financial support of the Institute for Innovation and Governance Studies, University of Twente. The usual disclaimer applies.

Published (Chapter 3), in Matching Voters with Parties and Candidates: Voting Advice Applications in a Comparative Perspective, edited by Diego Garzia and Stefan Marschall

(Colchester, England: ECPR Press), pp. 33-47.

2 An overview of the methods compared

2.1 Party self-placement

Without a doubt, the most obvious thing to do if one wants to estimate the position of a particular party is to ask the party itself. This is what *StemWijzer* (The Netherlands), *VoteMatch* (UK), and *Wahl-O-Mat* (Germany) do, along with many other candidate-based VAAs such as *Smartvote* (Switzerland), *Vaalikone* (Finland), *Manobalsas* (Lithuania) and the VAAs designed by Danish newspapers and internet media. In its simplest form, this method consists of sending a questionnaire to each party (or its candidates) asking them to place themselves on a number of statements and provide a brief justification of its placement. Despite its intuitive appeal, this method has been proven difficult to replicate in many electoral settings. This is because, while political parties are often willing to reveal their positions on issues that they 'own', they are less likely to reveal their positions on controversial issues which they consider to be non-salient or electorally damaging. Parties have long been resisting attempts by political scientists to survey the attitudes of their cadres and MPs, and continue to do so when confronted with questionnaires sent by VAA designers. It is therefore telling that only 103 out of the 274 (37.6%) parties in the EU Profiler agreed to provide their placements on the 30 issues statements provided by the VAA designers (Trechsel and Mair 2011: 15).

Since VAA designers who adopt this method do not necessarily verify, let alone challenge, the positions and justifications provided by parties, we should also consider the possibility that parties may provide strategic responses intended to manipulate the direction of advice given to VAA users. Indeed there are several instances where parties attempted to manipulate the process of self-placement in order to place themselves in positions that are perceived to be more popular among voters (see Krouwel et al. 2012: 233; Ramonaitė 2010: 134-137; Wagner and Ruusuvirta 2012: 406). Most often, centrist positions are more popular as VAA users tend to cluster in the middle of distributions in scales consisting of multiple items (as in the case of the two-dimensional political space of *Kieskompas* and the VAAs in this family), or in the middle of response scales in questions which have been framed so as to present a dilemma between two different policies (Baka et al. 2012).

2.2 Conventional expert surveys

Ever since Castles and Mair (1984) popularised this method in political science, expert surveys have been regularly used to estimate parties' positions, including VAAs such as *Pick-Your-Party* (Ireland), *HelpMeVote* (Greece), and *Votematch* (Italy). In its simplest form, an expert survey consists of a questionnaire sent to political scientists asking them to place certain political parties using the provided scales. Conventional expert surveys do not ask experts to justify the given placements. It is assumed that their expertise is enough to produce valid

estimates. Budge (2000) argued that expert surveys are problematic inasmuch different experts might evaluate different aspects of the party, may use different criteria, and different time frames. Steenbergen and Marks (2007) countered this by arguing that errors emanating from such problems tend to 'cancel each other out'. This cancelling out, however, may not necessarily lead to better estimates. As Tilley and Wlezien (2008) have shown, the 'cancelling' of errors by simple mathematical aggregation via some measure of central tendency (mean/median) may lead to implausible party placements near the centre of the scales. Experts are not as sophisticated as expert survey designers usually assume, and should not necessarily know how to place small parties on very specific statements such as the ones typically asked in VAAs. Instead of mathematical aggregation one needs to select and use the responses of the most sophisticated experts for each question posed in the expert survey, something that is impossible to do in conventional expert surveys since we have no means to evaluate experts' expertise.

Even though we cannot evaluate their expertise we know that experts estimate party positions with much uncertainty, as evident by their disagreement. Steenbergen and Marks (2007: 353-355) and Hooghe et al. (2010: 693) showed that expert disagreement seems to be a function of party differentiation, issue salience, internal dissent, party size, and party extremism. Experts are least in agreement when placing smaller parties on very specific issues, especially if the parties in question are not 'owners' of the issues. These findings imply that experts' estimates are reliable in some cases but not in others. Unfortunately, the latter are exactly the cases VAAs often attempt to estimate: very specific issue positions for very small parties. This is why VAAs often use expert surveys intended to capture party positions on more general policy scales (environment, social policy, immigration) than the more specific questions used by the same VAAs to capture voter preferences (Wall et al. 2009). If such general scales are used at the party level, it follows that similar scales need to be used to measure users' preferences since the general logic of VAAs is to match voters to parties by using common questions and scales. Asking users to self-place on such general scales, however, might compromise the measurement of their attitudes as the perception of their content varies considerably among respondents (Evans et al. 1996).

Finally, we should note that expert survey estimates of party positions might be biased since a considerable majority of political scientists are known to be leftist or liberal in their own preferences (Mariani and Hewitt 2008). Curini (2010) investigated this hypothesis and found that experts who are unsympathetic towards extreme right and conservative parties would sometimes place such parties (statistically) significantly more to the right compared to experts that are indifferent in terms of sympathy. Therefore, the switch from party self-placement to an expert survey does not necessarily imply an absence of bias. The parties' strategic manipulation of self-placement may be replaced by implicit bias coming from political scientists own partisan sympathies and levels of expertise.

2.3 The Kieskompas method

The pitfalls of party self-placement and expert survey methods have prompted VAA designers to opt for a hybrid method (Krouwel et al. 2012). This hybrid method has been used by *Kieskompas* (The Netherlands), and its variants, like *La Boussole Présidentielle* (France), Aftonbladets Valkompass (Sweden), *Bússola Eleitoral* (Portugal), as well as the EU Profiler. This method can be best described as iteration between party self-placement, and party placement by a small team of experts. The VAA questionnaire is sent to parties that are asked to position themselves on the given statements and provide some factual evidence of their placement, while a small team works concurrently but independently to place parties based on their manifestos and public statements. The two placements are compared to one another and, in cases of disagreement, parties are asked to reconsider their initial placement. After several rounds of iteration between the team and the parties, the percentage of statements in which parties and the coding team agree with regard to the placement, rises from 70-80% to around 95% (Krouwel and van Elfrinkhof 2013: 14). In the remaining statements where disagreement between parties and the team persists, despite the iteration over several rounds, the team makes the final decision about how party positions should be coded.

Krouwel and van Elfrinkhof (2013) argue that the *Kieskompas* method is an improvement over previously used methods such as party self-placement and expert surveys, as it combines their strengths in order to counter their weaknesses. The positions provided by parties aim to help the team (of experts) in cases where policy positions are not clearly stated in the publicly available documents, while the positions provided by the team aim to counter-balance the possibility of strategic manipulation by parties. The development of this hybrid method has been a particularly welcome development in the VAA literature, but it is not entirely unproblematic. For one, the *Kieskompas* method still requires the cooperation of parties. As already noted, however, the majority of parties in Europe are not willing to respond to questionnaires, and some even turn hostile when they realize that the placements they provided can be challenged by VAA designers (see Trechsel and Mair 2011: 13-15). Without the full and unfettered cooperation of political parties the *Kieskompas* method cannot work as originally intended.

Secondly, while it has been shown that the iteration between parties and teams of experts leads to a consensus position for the vast majority of the cases, we know little if anything about how the team reaches consensus for their own part of the estimation process. The *Kieskompas* method tries to ensure that the members of the team will be on the same 'page' and minimize inter-coder disagreement, by establishing a 'hierarchy of (document) sources' (Krouwel et al. 2012: 227-228). Yet disagreements can emerge as coders might be using the same piece of information but interpret or weight it differently (Bolger and Wright 1992: 61-63). Gemenis (2013: 278-279), for instance, found extensive disagreement among student coders when they were asked to code parties on selected EU

Profiler statements using their Euromanifestos as the sole piece of information. Trechsel and Mair (2011: 13) mention that such inter-coder disagreements were resolved through discussions among team members and the team leader. This implies that consensus in the team was reached through a process which can be characterized as 'unstructured behavioural aggregation' (Ferrell 1985: 135). Nevertheless, methodologists often advise against the use of such discussions as a mean to achieve consensus (Armstrong 2006), as such processes are known to be affected by the personalities and prestige of those involved in the discussion (Ferrell 1985: 136; Krippendorff 2004: 217).

2.4 The Delphi method

The problems with inter-expert/coder agreement in conventional expert surveys and the *Kieskompas* method have prompted VAA researchers to use an alternative method of eliciting and aggregating expert opinion (Gemenis 2012a). This approach uses the so-called Delphi method originally developed to forecast technological change (Dalkey and Helmer 1963). This Delphi method has been used to estimate parties' and candidates' positions in VAAs developed by the 'Preference Matcher' consortium starting with *Choose4Greece* (Greece), and continuing with *Xmamkvlevi* (Georgia), *VotulMeu* (Romania), and *Choose4Cyprus* (Cyprus).[2] The Delphi method is an interactive forecasting technique that relies on the judgmental input of a panel of experts through a process of 'structured behavioural aggregation' (Ferrell 1985: 140) characterized by anonymity and controlled feedback. In its simplest form, a 'moderator' selects a panel of experts who work independently of each other and asks them to provide estimates on parties' policy positions, and justify them by providing a piece of information. Subsequently, the moderator collects the individual estimates and associated pieces of information and feeds them anonymously back to the panel for a new round of estimation. The panelists are then asked to update their initial estimates based on the new information. Once sufficient consensus is reached, the responses are aggregated mathematically (by taking a measure of central tendency) for establishing final estimates.

A considerable body of evidence (for a comprehensive meta analysis, see Rowe and Wright 1999) has shown that the Delphi method gives more accurate estimates compared to mere mathematical aggregation via conventional expert surveys or unstructured behavioural aggregation via unstructured group discussions. Anonymity plays a crucial role as it guarantees that consensus is reached due to the quality of information associated with the estimates and is not affected by the personalities (and biases) of individual panelists.[3] The mechanism is simple: knowledgeable panelists will stick to their original estimates, whereas those with little information will revise their estimates towards the group average

[2] The consortium website can be accessed at: http://www.preferencematcher.org.

[3] Of course, it is possible for personalities to become manifest in the estimation via the provided justifications. We contend that this does not pose a serious challenge to the method, as anonymity allows panelists to focus on the usefulness of the information provided and disregard any possible personality effects.

(Parenté and Anderson-Parenté 1987). Gemenis (2012a) argues that the Delphi method performed well in eliciting consensus among panelists in *Choose4Greece*, including many 'difficult' cases, after two rounds of iteration. An additional advantage of using the Delphi method is that VAA designers need not rely on the cooperation of political parties, although party self-placement may be solicited and incorporated in the estimation process as an additional piece of feedback information between rounds. The main disadvantage of this method regards its cost. As the panel of experts cannot consist in its entirety of the VAA design team members, external expertise must be solicited. Panelists need to be remunerated for their involvement in the estimation process since the Delphi method is considerably cognitively and time-taxing compared to a conventional expert survey.

3 Data

In order to conduct a fair 'shoot-out' among the four methods we sought to compare their efficiency using a common set of VAA statements. For practical reasons, we drew six statements from the 2012 Dutch parliamentary election *Kieskompas* as we would be unable to replicate the exact process of estimation (discussions within the *Kieskompas* team and iteration with political parties) should we have chosen statements from another VAA. Our choice of statements (see Table X.2) attempted to reflect the typical differences in policy areas, complexity, and framing strategies found in VAAs.[4] For all statements the response was a 5-point scale ranging from 'completely disagree' (1) to 'completely agree' (5) with a 'no response' option.[5] Choosing the Netherlands as the country for comparison ensured the fairness of the shoot-out as the country is known for having high response rates in both conventional expert surveys and requests for party self-placement. For practical reasons associated with time constrains and cost, we chose to survey the eight largest parties from the eleven that are currently represented in *Tweede Kamer*.[6]

The party self-placement data were kindly provided by André Krouwel and consist of the initial party responses to the request for self-placement made by *Kieskompas*. For the expert survey data we conducted a conventional expert survey by contacting 42 Dutch political scientists specializing in party politics and electoral research through the on-line Lime Survey platform. In addition to

[4] The original wording of the statements is: (1) Het belastingtarief voor de hoogste inkomens moet omhoog; (2) De overheid moet meer ingrijpen in de economie; (3) Ambtenaren van de burgerlijke stand mogen weigeren homostellen te trouwen; (4) Alle volwassenen zijn automatisch orgaandonor, tenzij zij expliciet hebben aangegeven dat niet te willen; (5) Nederland moet in de euro blijven; (6) Het dragen van een boerka moet worden verboden.

[5] Original wording: (1) Helemaal niet mee eens; (2) Niet mee eens; (3) Neutraal; (4) Mee eens; (5) Helemaal mee eens; (missing value) Geen mening.

[6] It would have been interesting to see how the methods would perform in the estimation of smaller parties' positions, which is generally considered to be more difficult. Our choice of issues and parties combined, however, presents enough variation to allow us to make comparisons across different levels of estimation difficulty.

placing the eight parties on the six statements, we asked experts to place the same parties on a 10-point left-right scale.[7] To gauge experts' sympathy towards the parties, we asked them to indicate the degree to which each parties' policies corresponded to their own, using 10-point scales.[8] We received 25 valid responses (59.2% response rate), more than those received by the Chapel Hill team (Bakker et al. 2012) and Benoit and Laver (2006) expert surveys. For the *Kieskompas* method data, we extracted the final estimates from the 2012 Dutch parliamentary election *Kieskompas* website. Finally, for the Delphi method data we solicited the help of a panel of 14 experts. Since the Delphi method has been designed for use with disparate experts, our panel included four faculty members holding a doctorate, four PhD researchers, and six master's students. The panelists were assigned through a block haphazard procedure to estimate the positions of four parties each, so as to have seven panelists for each party (see Appendix). Estimation took place over two rounds with controlled feedback from one round to another through the use of J. Scott Armstrong's on-line Delphi platform. The panelists received instructions regarding the Delphi method and the on-line platform. More specifically, they were asked to refer to the 2012 election manifestos of the parties as much as possible when they were justifying their estimates, although they were also told that they could use other pieces of information or provide a personal justification in case the election manifesto was not helpful enough. After the second round of estimation, the panelists were remunerated for their participation.

4 Analysis

We begin the analysis with an evaluation of agreement among experts/coders. Since the party self-placement method provides only a single estimate, and since we were unable to fully replicate the unstructured group discussions in a way that would be fair to the *Kieskompas* method, we limit our examination of inter-expert/coder agreement to the expert survey and Delphi methods. To measure agreement, we use van der Eijk's (2001) coefficient A. As van der Eijk (2001: 328) demonstrated, in ordinal rating scales such as the 5, 7 and 10-point response scales used in VAAs and expert surveys, the standard deviation (which is typically used to evaluate expert surveys, e.g. Benoit and Laver 2006: 162-164; Steenbergen and Marks 2007: 353-355; Hooghe et al. 2010: 693) is inappropriate as measure of agreement or consensus because it reflects the skewedness of a distribution in addition to dispersion. Conversely, coefficient A mitigates this problem as it conceptualizes agreement as a function of dispersion and deviation from unimodality. Table X.1 presents the A coefficients for each of the six statements. As can be seen from the figures in the table, agreement among the experts is quite high for both the expert survey and the Delphi method estimation.

[7] Original question wording: 'Plaatst u alstublieft alle partijen op een algemene links-rechts dimensie. Neem daarbij zoveel mogelijk beleidsposities van de partij in overweging.'
[8] Original question wording: 'Als u alle aspecten van het partijbeleid in overweging neemt, geef dan alstublieft voor iedere partij aan hoe dicht deze bij uw eigen opvattingen staat.'

Table X.1: Perceptual agreement among experts

	PvdA		**VVD**		**CDA**		**PVV**		**CU**		**D66**		**GL**		**SP**		*Mean*	
	Round		Round		Round		Round		Round		Round		Round		Round		Round	
Statement	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd
1. Taxes	0.93	1	1	1	0.4	0.79	0.55	0.86	0.22	0.71	0.55	0.48	1	1	1	1	0.71	0.86
2. State/economy	0.86	1	0.86	1	0.71	0.93	0.71	0.93	0.55	0.86	0.57	0.71	0.38	0.93	1	1	0.71	0.92
3. Gay couples	1	1	0.93	1	0.86	0.79	0.93	1	1	1	1	1	0.93	1	0.45	1	0.89	0.97
4. Organ donors	0.53	1	0.93	1	0.31	0.57	0.79	0.79	0.86	1	0.86	0.93	0.93	1	0.92	1	0.77	0.91
5. Euro	1	1	0.93	0.93	1	1	1	1	0.64	0.79	1	1	0.86	1	0.75	0.86	0.9	0.95
6. Burqa	0.75	0.79	0.93	1	0.63	0.7	1	1	0.29	0.86	0.83	1	0.86	0.93	0.42	0.79	0.71	0.88
Mean	0.85	0.97	0.93	0.99	0.65	0.80	0.83	0.93	0.59	0.87	0.80	0.85	0.83	0.98	0.76	0.94	**0.78**	**0.92**
1. Taxes	0.84		0.98		0.76		0.48		0.73		0.66		0.86		1		0.79	
2. State/economy	0.86		0.92		0.82		0.5		0.52		0.66		0.78		0.96		0.75	
3. Gay couples	0.79		0.68		0.68		0.65		0.92		0.98		0.94		0.54		0.77	
4. Organ donors	0.71		0.48		0.82		0.63		0.89		0.93		0.75		0.44		0.71	
5. Euro	0.86		0.8		0.76		0.85		0.6		0.96		0.82		0.52		0.77	
6. Burqa	0.73		0.51		0.61		0.98		0.68		0.72		0.72		0.5		0.68	
Mean	0.8		0.73		0.74		0.68		0.72		0.82		0.81		0.66		**0.75**	
Left-right	0.79		0.74		0.79		0.53		0.73		0.86		0.75		0.72		**0.72**	

The level of perceptual agreement in the first Delphi round is similar to that of the expert survey, although the figures are higher in the Delphi in four out of six statements. The comparison to the second Delphi round, however, is unequivocal. The anonymous iteration with feedback of the Delphi method leads to an even higher agreement among the panelists, higher than any figure observed in the expert survey. The examination of the figures for individual parties confirms this pattern although it also shows that the average figures mask considerable cross-party variation, especially for the expert survey where 17 out of the 48 A coefficients were under 0.7 indicating a high degree of disagreement. For the Delphi, only two A were under 0.7 after the second round estimation. Both of these were about the placement of the *Democraten '66* (D66) party on the two statements about the economy. Most likely, this difficulty in agreeing about the position in these statements, even after the iteration with feedback, can be attributed to the well-established centrist position of D66 in economic issues that can be a source of conflicting messages in its manifesto. In general, the results point to the efficiency of the Delphi method in terms of achieving consensus among experts in comparison to the conventional expert survey.

Of course, reliability, measured in terms of perceptual agreement among experts, may not necessarily lead to valid estimates, as reliability is generally considered to be a necessary but not sufficient condition for validity (Krippendorff 2004: 212-213). Unfortunately, when party positions are concerned, we lack the 'gold standard' benchmark against which estimates from various methods can be compared. This is why the attempts to validate party positions usually involve comparisons among different, yet imperfect, methods, and discussions as to which method positions parties according to well-established intuitions (e.g. Dinas and Gemenis 2010; Krouwel 2012; Marks et al. 2007). We follow a similar approach. We refrain from summarizing the relationship between different methods using statistical measures of association because the nature of the party placements in VAAs cannot be represented accurately through non-parametric measures of association based on rank order.[9] In Table X.2 we compare the placements of the eight parties on each of the six issues across the

[9] More specifically, Spearman's ρ and Kendall's τ_a cannot handle the ties in the dataset (two parties placed in the same position). Kendall's τ_b can handle ties by using a divisor term, but like the aforementioned measures assumes rank ordered data. VAA party position data are not rankings but continuous variables that are observed as ordinal measures through a process of discretisation (e.g. experts assigning parties to the 'agree', 'disagree', etc categories). Pearson's polychoric correlation coefficient ρ is not based on rank order and has attractive properties for VAA data as it assumes a continuous latent variable behind the ordered ratings. Its estimation, however, assumes bivariate normality that is clearly not satisfied by the skewed nature of the VAA party position data. Moreover, the use of correlation coefficients overestimates the degree of concordance between methods of party positioning in the presence of systematic error (Gemenis 2012b: 600-601). Consider the following example: 'Method A' places parties A=1, B=2, C=3 and 'Method B' A=3, B=4, C=5 respectively. In this case Spearman's ρ, Kendall's τ_b and Pearson's ρ all equal 1 indicating a perfect correlation, as the rank ordering of the parties remains the same. However, the correlations disregard the systematic difference between the two methods. While 'Method A' places party B as 'disagree', 'Method B' places it as 'agree'. Lin's concordance correlation coefficient ρ_c accounts for the presence of systematic error by using a bias correction factor, but assumes continuous data.

four methods. As evident from this table, the four methods generally agree on most placements but there are many disagreements as well. We discuss some of them below with reference to what parties argue in their 2012 election manifestos.

First of all, in line with previous research, agreement appears to be highest for issues that were salient in the campaign, and disagreement appears to be highest for centrist parties, and for parties that do not 'own' an issue, while party size seems to matter less. For example, while the economic crisis was the major issue in the 2012 election campaign, two issues that received quite some media attention were the proposal by the radical right Party for Freedom (PVV) to pull out of the Euro and return to the guilder (statement 5) and the discussion about whether or not registars should be allowed to refuse to marry same-sex couples if they had moral objections to do so (statement 3). On these two issues, all methods differentiate clearly between parties for and against: in the case of the Euro, only PVV was against remaining in the Eurozone, all other parties were in favour; and in the case of gay marriage, only the Christian-democratic (CDA and CU) parties were in favour of allowing registars to refuse to marry same-sex couples, with all other parties against. In the case of the PVV, the consensus among all the methods can be attributed to its ownership of the Euro issue. Another example of the effect of issue ownership on the clarity of party positioning is the position of D66 on automatic organ donation, a proposal put forward by D66 (statement 4).

Moreover, disagreement between methods appears to be higher for CU and CDA, and to a lesser extent the radical left Socialist Party (SP) and D66, on issues on which these parties have either a centrist or unclear position. For example, on the issues of increasing tax rates for high incomes, and state involvement in the economy (statements 1 and 2), CU gets positioned by the different methods on 'disagree', 'agree' as well as 'neither agree nor disagree'. Closer scrutiny of the manifesto shows that CU takes a rather centrist position on these issues, stating for example that taxes should be proportional to incomes, without directly proposing that taxes for higher incomes should be increased.[10] As another example, on the issue of organ donation CDA gets four different positions from the different methods, being positioned from 'completely disagree' to 'agree'. In fact, the manifesto states that CDA wishes to further stimulate organ donation, however there is no mention of an automatic registration system.[11] Moreover, in the run up to the election this was one of the issues of contention at the party members' congress, and hence it is not surprising that there was disagreement among the different methods, given its unclear position.

[10] CU election manifesto, *Voor de verandering: 7, Christelijk-sociale hervormingen. Verkiezingsprogramma 2013-2017*, 57.
[11] CDA election manifesto, *Iedereen, verkiezingsprogramma 2012-2017*, 55.

Table X.2: Party positions in the selected statements.

	Completely disagree	Disagree	Neither agree, nor disagree	Agree	Completely agree
1. Tax rates for the highest incomes should be increased.					
Self-placement	VVD	CDA, PVV, CU, D66		PvdA	SP, GL
Kieskompas	VVD, PVV	CDA, CU, D66			SP, PvdA, GL
Expert survey	VVD	CDA, D66	PVV, CU	PvdA, GL	SP
Delphi method	VVD	CDA, PVV, D66		CU	SP, PvdA, GL
2. The government should intervene more in the economy.					
Self-placement	VVD	PVV, CU, D66	CDA	PvdA, GL, SP	
Kieskompas	VVD, PVV	CDA, CU, D66		PvdA, GL	SP
Expert survey		VVD, PVV, CDA, D66	CU	PvdA, GL	SP
Delphi method	VVD	PVV	CDA, D66	PvdA, GL, CU	SP
3. Registrars should be allowed to refuse to marry gay couples.					
Self-placement	VVD, PvdA, D66, GL	PVV, SP		CDA, CU	
Kieskompas	VVD, PVV, PvdA, D66, GL, SP			CDA	CU
Expert survey	PVV, D66, GL	VVD, PvdA, SP		CDA	CU
Delphi method	VVD, PVV, PvdA, D66, GL, SP				CU, CDA
4. All adults are automatically registered as organ donors, unless they have explicitly declared otherwise.					
Self-placement		VVD, PVV, CU		CDA, SP, GL	PvdA, D66
Kieskompas	PVV	VVD, CDA, CU			PvdA, D66, GL, SP
Expert survey	CDA, CU	VVD, PVV	SP	PvdA, GL	D66
Delphi method	VVD, CU	PVV	CDA		PvdA, D66, GL, SP
5. The Netherlands should stay part of the Euro.					
Self-placement	PVV			VVD, CU, SP	CDA, PvdA, D66, GL
Kieskompas	PVV	CU		VVD, SP	CDA, PvdA, D66, GL
Expert survey	PVV			VVD, CU, SP	CDA, PvdA, D66, GL
Delphi method	PVV		CU	SP	CDA, PvdA, D66, GL, VVD
6. Wearing a burqa should be prohibited.					
Self-placement		PvdA, D66, GL, CU, SP			VVD, PVV, CDA
Kieskompas		PvdA, D66, GL, CU, SP			VVD, PVV, CDA
Expert survey	D66, GL	PvdA, SP	CDA, CU	VVD	PVV
Delphi method	D66, GL	PvdA	SP	CDA, CU	VVD, PVV

Now, while there is generally rather high agreement between the different methods, some differences do appear. As regards the self-placement of parties, parties appear to position themselves generally in accordance with other methods. However, there are several occasions where parties place themselves slightly less extreme than the other methods, which might indicate strategic behaviour to appeal more easily to the median voter. There is only one case in which a party places itself opposite to other codings, i.e. the CDA in the case of organ donation. Turning to *Kieskompas*, this method appears to use more dispersed party placements, and, more importantly, never uses the middle-category of 'neither agree nor disagree'. Arguably however, in some cases parties do take a centrist position, such as the position of the CU on the issues mentioned above. The expert surveys seem to have the reverse tendency to *Kieskompas*, i.e. to position parties more often on a centrist position or on less extreme positions. Also, on some issue positions experts may have been using ideological dimensions of party competition as a heuristic short-cut to position parties, rather than using information from the party manifesto. A clear example of the latter is the issue of the CDA organ donation, where the position of the party was at best unclear, while experts nevertheless placed the party on 'completely disagree'. It seems likely that experts did not actually know the party's position on this issue, but used its conservative position on the liberal-conservative dimension as a cue. Finally, the Delphi method appears to be placing parties in more dispersed positions than party self-placements or the expert survey, and yet uses the middle position more often than *Kieskompas*. Contrary to the other three methods, the Delphi method allows us to check why parties were positioned in a centrist position, or other positions. Moreover, experts can use more sources than just the party manifesto, which greatly helps to clarify potential party shifts in position during an electoral campaign.

For instance, in the case of the CDA's position on automatic organ donation the CDA manifesto was unclear about the party's position while its formed position as stated on the party website had been removed during the campaign. The panelists indicated (always referring to verifiable sources) that there had been a debate at the party members' congress, as the result of which the party eventually adopted the position to accept a system of automatic donor registration that would however, give citizens many more chances to opt out than the system originally proposed by D66. This example hints at the efficacy of the Delphi method in estimating parties' positions under uncertainty.

As discussed, a general pattern that emerges from Table X.2 is that some methods tend to give a more polarized picture of party placements. Since looking directly at 192 (6x8x4) placements makes it difficult to assess this pattern, we summarize the degree of polarization by using the (1-A)/2 formula, where A is van der Eijk's (2001) coefficient of agreement calculated among the placements of the eight parties on each of the statements. As evident in Table X.3, the expert survey (followed by party self-placement) is the method that gives a more centripetal picture of Dutch party positions on every statement with the

exception of the statement about the Euro where the Delphi method gives a slightly more centripetal picture due to the centrist placement of CU. The *Kieskompas* and the Delphi methods clearly give the most polarized picture compared to the other two methods. Although we have no benchmark to assess which scenario is the more plausible, we nevertheless contend that, for the expert survey, the observed centripetal tendency is related to the uncertainty associated with the estimates. When faced with uncertainty, respondents tend to pick the middle responses as a 'safe' value or as a proxy for 'I don't know' (see Baka et al. 2012). Alternatively, respondents might pick a less centrist response, but under uncertainty these will cancel each other out and bring the median estimate to the centre of the scale (see Tilley and Wlezien 2008). In the case of self-placement by parties, polarization is somewhat more pronounced. Here, centrist placements are less likely to be associated with uncertainty and either the result of a true centrist position, or the product of a strategy intended to give a position that would look appealing to the median voter. For these reasons we contend that the polarized picture presented by the *Kieskompas* and Delphi methods is likely to be more plausible.

Table X.3: Polarization in party positions

	Self-placement	**Kieskompas**	**Expert survey**	**Delphi method**
1. Taxes	0.46	0.66	0.41	0.64
2. State/economy	0.4	0.45	0.29	0.35
3. Gay couples	0.34	0.24	0.35	0.5
4. Organ donors	0.44	0.6	0.48	0.51
5. Euro	0.3	0.32	0.3	0.26
6. Burqa	0.53	0.53	0.41	0.55
Mean	0.41	0.46	0.37	0.47

5 Conclusions

We draw several conclusions from the comparison of the four methods in estimating party positions in VAAs. Firstly, we consider that the party self-placement method is impractical in many contexts since party response rates tend to be low. Nevertheless, even when parties do respond to VAA questionnaires, as is the case for the Dutch parties, strategic manipulation remains as a possibility.

Our empirical analysis noted the tendency of this method to portray parties as centrist, at least more so than alternatives such as the *Kieskompas* method. The degree to which these centrist positions can be attributed to strategic placement is not directly verifiable, yet we note that parties have been shown to be able to manipulate the direction of voting recommendations in VAAs by taking carefully calculated combinations of extreme positions (see Ramonaitė 2010). Our conclusion is that VAAs should not uncritically rely on party self-placements. Expert surveys transfer the responsibility of party positioning from parties to political scientists. Nevertheless, expert survey estimates of party positions exhibit much uncertainty. Although the average picture is one of modest to

considerable agreement among the experts, prompting researchers to unequivocally label expert surveys as 'reliable' (Hooghe et al. 2010; Steenbergen and Marks 2007), a closer examination reveals that agreement varies considerably from issue to issue. Disagreement often leads to centrist estimates, and our comparison to the remaining three methods showed that such estimates may be invalid. Our conclusion is that expert surveys cannot reliably estimate the positions of all parties on all statements in a VAA. The *Kieskompas* method promises to counter the weaknesses of the aforementioned methods by combining their strengths. Our comparison showed that this promise is generally satisfied although two concerns remain. To begin with, the method still requires parties' full cooperation. If parties respond, the aggregation of responses can be structured on the party/team interaction. In cases where parties do not cooperate, however, the estimation on the team side becomes somewhat of a black box. Although there is evidence that team members often disagree with each other as to how parties should be placed, the process of fostering consensus via unstructured team discussions cannot be validated empirically through replication (see Krippendorff 2004: 217-219) so claims that 'inter-coder reliability was maximized' (Trechsel and Mair 2011: 13) should be viewed with much caution.

The Delphi method aims to overcome these concerns associated with the *Kieskompas* method. Consistent with previous research (Gemenis 2012b), our application of the Delphi method to the case of the Dutch parties showed that the anonymous iteration with feedback over two rounds increased the consensus among experts to levels considerably higher than those observed in a conventional expert survey. Moreover, our comparison showed that this consensus gave rather plausible estimates of party positions even in several 'difficult' cases. The problems in applying this approach in the VAA context are largely practical. The first regards cost. For employing seven panelists to estimate the positions in an eight parties party system on 30 statements as in a typical VAA, the estimated cost for remunerating the panelists is about 4,000 euro. Of course, this could be reduced considerably if some of the panelists are members of the design team. We contend that this is not a prohibitive cost given that some VAAs receive generous government funding, while others receive media sponsorships. The second problem is of a technical nature. The only easily accessible on-line platform for Delphi estimation is dated and often exhibits technical problems. The Preference Matcher consortium is nevertheless in the process of providing a specialized platform for Delphi estimation that is specifically designed for coding party positions in the VAA context, and a beta version of this platform has been used successfully during the design of *Choose4Cyrpus*. With these issues practically solved, the next goal is to test the accuracy of the Delphi method in randomized experiments for determining, a) the optimal type of feedback (median position and/or justifications versus simple iteration), and b) the optimal composition of the panel in terms of size and expertise. Although such questions have been rigorously tested in other contexts (see Rowe and Wright 1999; Rowe et al. 2005), it is necessary to examine them

in the VAA/party positions context. With such issues addressed, we will be in possession of a method that is practical, as well as capable of producing reliable and valid estimates of parties' policy positions.

Bibliography

Armstrong, J. S. (2006) 'How to make better forecasts and decisions: Avoid face-to-face meetings', *Foresight: The International Journal of Applied Forecasting*, 5: 3-8.

Baka, A., Figgou, L. and Triga, V. (2012) 'Neither agree, nor disagree: A critical analysis of the middle answer category in voting advice applications', *International Journal of Electronic Governance*, 5: 244-263.

Bakker, R., de Vries, C., Edwards, E., Hooghe, L., Jolly, S., Marks, G., Polk, J., Rovny, J., Steenbergen, M. and Vachudova, A. M. (2012) 'Measuring party positions in Europe: The Chapel Hill Expert Survey trend file, 1999-2010', *Party Politics*, doi: 10.1177/1354068812462931.

Benoit, K. and Laver, M. (2006) *Party policy in modern democracies*, London: Routledge.

Bolger, F. and Wright, G. (1992) 'Reliability and validity in expert judgment,' in Wright, G. and Bolger, F. (eds.) Expertise and decision support, New York: Plenum pp. 47-76.

Budge, I. (2000) 'Expert opinions of party policy positions: Uses and limitations in political research', *European Journal of Political Research*, 37, pp. 103-113.

Castles, F. G. and Mair, P. (1984) 'Left-right political scales: Some "expert" judgments', *European Journal of Political Research*, 12: 73-88.

Curini, L. (2010) 'Experts' political preferences and their impact on ideological bias: An unfolding analysis based on a Benoit-Laver expert survey', *Party Politics*, 16: 299-321.

Dalkey, N. C. and Helmer, O. (1963) 'An experimental application of the Delphi method to the use of experts', *Management Science*, 9: 458-467.

Dinas, E. and Gemenis, K. (2010) 'Measuring parties' ideological positions with manifesto data: A critical evaluation of the competing methods', *Party Politics*, 16: 427-450.

Evans, G., Heath, A. F. and Lalljee, M. (1996) 'Measuring left-right and authoritarian-libertarian values in the British electorate', *British Journal of Sociology*, 47: 93-112.

Ferrell, W. R. (1985) 'Combining individual judgments,' in Wright, G. (ed.): *Behavioural decision-making*, New York: Plenum pp. 111-145.

Gemenis, K. (2012a) 'A new approach for estimating parties' positions in voting advice applications', paper presented at the XXVI Congress of the Italian Political Science Association, Rome, September 13-15.

Gemenis, K. (2012b) 'Proxy documents as a source of measurement error in the Comparative Manifestos Project', *Electoral Studies*, 31: 594-604.

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

Hansen, M. E. and Rasmussen, N. E. K. (2013) 'Does running for the same party imply similar policy preferences? Evidence from voting advice applications', *Representation*, doi: 10.1080/00344893.2013. 805161

Hooghe, L., Bakker, R., Brigevich, A., de Vries, C., Edwards, E., Marks, G., Rovny, J., Steenbergen, M. and Vachudova, M. (2010) 'Reliability and validity of the 2002 and 2006 Chapel Hill expert surveys on party positioning', *European Journal of Political Research*, 49: 687-703.

Krippendorff, K. (2004) *Content analysis: an introduction to its methodology*, second ed., Thousand Oaks, CA: Sage.

Krouwel, A. (2012) *Party transformations in European democracies*, Albany: State University of New York Press.

Krouwel, A. and van Elfrinkhof, A. (2013) 'Combining strengths of methods of party positioning to counter their weaknesses: The development of a new methodology to calibrate parties on issues and ideological dimensions', *Quality and Quantity*, doi: 10.1007/s11135-013-9846-0.

Krouwel, A., Vitiello, T. and Wall, M. (2012) 'The practicalities of issuing vote advice: A new methodology for profiling and matching', *International Journal of Electronic Governance*, 5: 223-243.

Ladner, A., Fivaz, J. and Pianzola, J. (2012) 'Voting advice applications and party choice: Evidence from Smart Vote users in Switzerland', *International Journal of Electronic Governance*, 5: 367-387.

Laver, M. (ed.) (2001) *Estimating the policy position of political actors*, London: Routledge.

Mariani, M. D. and Hewitt, G. J. (2008) 'Indoctrination U.? Faculty ideology and changes in student political orientation', PS: Political Science and Politics, 41: 773-783.

Marks, G. Hooghe, L., Steenbergen, M. and Bakker, R. (2007) 'Crossvalidating data on party positioning on European integration', *Electoral Studies*, 26: 23-38.

Parenté, F. J. and Anderson-Parenté, J. K. (1987) 'Delphi inquiry systems,' in Wright, G. and Ayton, P. (eds.) *Judgemental forecasting*, Chichester, England: Wiley pp. 129-156.

Ramonaitė, A. (2010) 'Voting advice applications in Lithuania: Promoting programmatic competition or breeding populism?', *Policy and Internet*, 2(1): 117-147.

Rowe, G. and Wright, G. (1999) 'The Delphi technique as a forecasting tool: Issues and analysis', *International Journal of Forecasting*, 15: 353-375.

Rowe, G., Wright, G. and McColl, A. (2005) 'Judgment change during Delphi-like procedures: The role of majority influence, expertise, and confidence', *Technological Forecasting and Social Change*, 72: 377-399.

Steenbergen, M. and Marks, G. (2007) 'Evaluating expert judgements', *European Journal of Political Research*, 46: 347-366.

Tilley, J. and Wlezien, C. (2008) 'Does political information matter? An experimental test relating to party positions on Europe', *Political Studies*, 56: 192-214.

Trechsel, A. H. and Mair, P. (2011) 'When parties (also) position themselves: An introduction to the EU Profiler', *Journal of Information Technology and Politics*, 8: 1-20.

Van der Eijk, C. (2001) 'Measuring agreement in ordered rating scales', *Quality and Quantity*, 35: 325-341.

Volkens, A. (2007) 'Strengths and weaknesses of approaches to measuring policy positions of parties', *Electoral Studies*, 26: 108-120.

Wagner, M. and Ruusuvirta, O. (2012) 'Matching voters to parties: Voting advice applications and models of party choice', Acta Politica, 47: 400-422.

Walgrave, S., van Aelst, P. and Nuytemans, M. (2008) '"Do the vote test": The electoral effects of a popular vote advice application at the 2004 Belgian elections', *Acta Politica*, 43: 50-70.

Wall, M., Krouwel, A. and Vitiello, T. (2012) 'Do voters follow the recommendations of voter advice application websites? A study of the effects of kieskompas.nl on its users' vote choices in the 2010 Dutch legislative elections', *Party Politics*, doi: 10.1177/1354068811436054.

Wall, M., Sudulich, M. L., Costello, R. and Leon, E. (2009) 'Picking your party online: An investigation of Ireland's first online voting advice application', *Information Polity*, 14: 203-218.

Wheatley, J., Carman, C., Mendez, F., Mitchell, J. (2012) 'The dimensionality of the Scottish political space: Results from an experiment on the 2011 Holyrood elections', *Party Politics*, doi: 10.1177/1354068812458614.

Appendix

Panelists: Loes Aaldering, Klaas Derks, Mark Hessels, Rens Hogeling, Elmar Jansen, Joyce Kuipers, Paul Lucardie, Jannine van de Maat, Daphne Van der Pas, Martin Rosema, Mariken Van der Velden, Cynthia Van Vonno, Annemarie Walter, Marc Van der Wardt.

Table X.4: Assignment of Delphi panellists to parties.

Panellist	PVV	CDA	D66	VVD	GL	PvdA	SP	CU
Faculty member 1	X	X			X	X		
Faculty member 2		X	X			X	X	
Faculty member 3			X	X			X	X
Faculty member 4	X			X	X			X
PhD researcher 1	X			X	X			X
PhD researcher 2	X	X			X	X		
PhD researcher 3		X	X			X	X	
PhD researcher 4			X	X			X	X
Master student/graduate 1	X		X		X		X	
Master student/graduate 2		X		X		X		X
Master student/graduate 3	X		X		X		X	
Master student/graduate 4		X		X		X		X
Master student/graduate 5	X		X		X		X	
Master student/graduate 6		X		X		X		X