Voting Aid Applications between charlatanism and political science: the effect of statement selection

Stefaan Walgrave, University of Antwerp
Michiel Nuytemans, University of Antwerp
Koen Pepermans, University of Antwerp

ABSTRACT

Vote Aid Applications (VAA) help voters make their decision at election times. They typically consist of a number of statements similar to likert items that are used to match voters with parties. The paper establishes that the specific selection of statements forming the VAA has a large impact on the vote advices that are produced: some configurations favour certain parties, other configurations benefit other parties. Drawing on a large-scale simulation of 500,000 different configurations of 36 statements and on a random sample of Belgian voters we show that many of these combinations produce advices that are not at all consistent with the real electoral strength of the parties. Whether statement are weighed or not does not make a lot of difference, the gap between the real world and the output of the VAA remains very large. The paper ends with a plea for a careful selection of VAA statements and for a proper process of benchmarking based on survey data. Without appropriate calibrating, VAAs produce invalid results.

Voting Aid Applications between charlatanism and political science: the effect of statement selection

INTRODUCTION

A new phenomenon in modern election campaigning Voting Aid Applications (VAA) help voters cast their vote. Drawing upon an internet application VAAs link a voter's preferences—most of the time answers to precise issue-related questions—to parties' policy proposals or general preferences. VAAs produce a sort of "advice", or at least an aid, for the participating voter by highlighting the party that stands closest to the voter's preferences. VAAs are spreading quickly to ever more countries. In some countries, for example in The Netherlands, VAAs have even become one of the most important players in the election campaign. Not only are VAAs spreading to most European democracies, also the number of voters participating in a VAA and getting a voting advice grows year after year (for an overview see Walgrave, Van Aelst, and Nuytemans 2008b). VAAs' popularity raises important questions as to the validity of the voting advice that is being produced by VAAs' underlying algorithms. What is the value of the advice voters get? Is it a "good" and reliable hint for voters to follow or is it just crap, at most an entertaining game producing random and messy advices?

Together with the spread of VAAs to ever more countries, the diversity of the applications has increased. Although basically doing the same thing—matching voters with parties based on both their preferences—the differences between VAAs abound. Some VAAs rely on parties' explicit 'authorization' concerning their stances, others draw on a detailed analysis of parties' official documents; some VAAs weight the statements according to the saliency attributed to the issue by voters or parties, other VAAs just add up all statements giving them an equal weight; some VAAs categorize the political space in different dimensions, other VAAs consider the political space as one-dimensional; some VAAs work with statements that tap very precise policy proposals, other VAAs rely on general and ideological statements (for an overview see: Laros 2007). So, not all VAAs are they based on the same principles. This

diversity of VAA-systems begs for a scientific reflection on VAAs' outputs and the status of the advice they produce.

In some of the countries with popular VAAs an intense debate broke out. Some maintained that VAAs are a fraud that can never be able to give correct and neutral vote advice; others contended that these applications must be commended as they focus people's attention on the party programmes and to policy issues compelling parties to discuss substance instead of personalities, images, and campaign events (see for example for the discussion in Belgium: Deschouwer and Nuytemans 2005; Swyngedouw and Goeminne 2005).

In sharp contrast to their amazing success the *scientific* debate about VAAs has hardly commenced. Political scientists seem to have been taken by surprise and have only just started to think about VAAs. Many political scientists have themselves been heavily implied in designing VAAs. For some VAAs, political scientists even provide the basic party preference information: as experts, they estimate where parties are situated in the political spectrum and what parties *would* answer when confronted with a certain statement (Teepe 2005). This involvement of political scientists in designing VAAs places them in an awkward position to critically evaluate VAAs and to feed the debate about VAAs' consequences (Ladner, Felder, and Fivaz 2008). Sure, there have been a handful of studies tackling the effect of VAA-advices on the voter but there have barely been studies focusing on the heart of the matter: How do VAAs work? What kind of "advice" do VAAs produce? And to what extent is the advice VAAs deliver sound and reliable?

In this paper, we want to start this debate by pinpointing just one aspect of VAAs make-up: the selection of the statements that are incorporated in the system. VAAs typically link a voter's answers to +30 very specific statements to parties' stances on these same topics. The statements are most of the time typical likert items asserting a certain position and participants have to indicate whether they (totally) agree, agree nor disagree, or (totally) disagree with it. The topic, the exact wording, the direction, and the formulation of a statement may all produce differences in answers. Some statements may benefit some parties, while other statements may work to the advantage of other parties. Hence, we hypothesize that selecting topics and

formulating statements is key to devise an equilibrated VAA-system that does justice to parties' program and gives them a fair chance to persuade voters to cast their vote for them. Yet, theoretically, the number of potential statements is infinite and any statement incorporated in a VAA might be replaced by any other statement on the same or on a different topic. There are no natural or fix statements that are per definition part of a VAA; constructing a VAA unavoidably means going through a selection process and choosing one statement above others. Apart from statement selection, there are of course many other aspects of VAAs internal machinery that are worth studying and that most likely affect their output. But we claim statement selection to be the key process.

The statements in a VAA, also, determine the look and feel of the system; the statements form the most visible aspect of a VAA and they have sparked most discussion in countries where VAA have led to a debate. This is no coincidence as VAA statements form, so to say, the software of the system; they determine what kind of voter preference information gets fed into the system that calculates the best fitting party. Different statements entail that different information is gathered from participants and this, as we will test in this paper, may lead to different outputs. Therefore the question we tackle in this paper is: *to what extent is the output of Voting Aid Applications—that is: the individual voting advice that participants get—determined by the specific selection of statements incorporated in the system*?

The data we use here have been used to build the Belgian[1] VAA *Doe De Stemtest* for the general elections in June 2007. *Doe De Stemtest* consisted of 36 statements. To arrive at this final selection, a short-list of 50 statements was included in a survey applied to a random sample of 1,000 Belgian citizens. We use the answers of this sample of Belgian citizens to test whether selecting 36 statements out of a list of 50 makes a difference. We will show that the range of possible Vote Aid Application output distributions is sheer endless: any selection of 36 out of 50 statements yields different results. That is to say, the share of advices to vote for a party differs dramatically across the many billions of possible 36-statement-configurations.

1 In fact, *Doe De Stemtest* was not a real Belgian VAA. It only targeted the Dutch-speaking part of the Belgian population living in Flanders (North) and forming 60% of the Belgian population.

The paper starts with documenting the spectacular spread of VAAs throughout Europe. Then, we recapitulate the limited scientific literature about the workings of VAAs. We especially focus on the problem of selecting issues and statements. Next, we present our evidence. Then, we analyze our data and discuss the results. We wrap up with a conclusion and discussion section.

BOOMING VOTE AID APPLICATIONS ACROSS EUROPE[2]

Vote Aid Applications have become increasingly popular in a large number of European countries over the last few years. Media companies or independent agencies have set up popular websites giving voters advice about which party program comes closest to their own preferences. In some countries VAAs have even been launched by at major TV-shows giving viewers the chance to participate and to receive their voting advice immediately while watching TV. Both the websites and the TV-shows have often had impressive participation rates and viewing figures suggesting that large numbers of people have been exposed to them.

To get a raw pictuire of the spread of the VAA phenomenon, in 2007, we set up a small expert survey of sample of political scientists from all European countries (via a web survey) to gather information on the existence of Vote Aid Applications in their country. 38 academics responded, giving us a first impression of the situation in 22 countries. In 15 of the 22 countries there has been an operational VAA at the last general elections. This is the case in as diverse countries as Belgium, Bulgaria, the Czech Republic, Denmark, Finland, France, Germany, Greece, Italy, Hungary, Latvia, Norway, Slovakia, Sweden, Switzerland and the Netherlands. In countries as Ireland and the United Kingdom, by the time of our survey in 2007, no VAAs were operational (yet).

The Netherlands was the VAA pioneer, with a first VAA in use as early as in 1989. The *Stemwijzer* only reached 500 participants with a print-based VAA. The numbers radically changed with the rise of the Internet. In 1998 the *Stemwijzer* went online

[2] This section draws on AUTHOR (2008).

for the first time and several others immediately followed. While in 1998, 250,000 Dutch voters asked for voting advice via a VAA this figure exploded to 2,600,000 users in 2002, an astonishing 25% of the Dutch electorate. For the 2003 general elections, no less than *eight* different VAAs were online. Finland is another early adopter with the first and still most popular VAA starting as early as in 1995. Four year later there were four television channels broadcasting TV-shows each launching another VAA.

At the start of the new century, the successful Dutch *Stemwijzer* was exported to several other countries. The German *Wahl-o-mat* and the Swiss *Politarena* are licensed versions of the *Stemwijzer*. In 2002, the German *Wahl-o-mat* was used for the first time, and to date it has attracted more than ten million users. It was most successful in 2005, when more than five million people participated (Marschall and Schmidt 2008). In Switzerland, *Politarena* and *Smartvote* both started in 2003, reaching 600,000 people. Since then more than 20 different vote applications were designed for several local and national elections. While Switzerland has a so called "panachage" system, were voters can vote for candidates from more than one party, *Smartvote* is able to create a list of candidates (of different parties) that are closest to one's opinions. Voters can even modify this list, print it and take it with them when they go out to vote (Ladner, Felder, and Fivaz 2008). Apart from Switzerland and Germany, also Belgium (Flanders) has been inspired by the Dutch example. In 2003, 2004, and 2007 the Belgian public broadcaster VRT launched its *Doe De Stemtest* TV-show annex VAA which was used by 840,000 people in 2004 (Walgrave and Van Aelst 2005). Together with the success of the VAA by the daily *De Standaard* the number of users amounted to about 1,000,000 in 2004 in Belgium.

The growth of the VAA phenomenon clearly is not yet finished. In countries where VAAs have been successful, the original VAAs have been joined by competitors. Also geographically VAAs are further spreading across Europe. In countries such as Sweden, Norway, Slovakia, and the Czech Republic VAAs have been launched recently, always with high participation rates. VAAs' success is probably due to the international trend of dealignment, and rising number of floating voters throughout Europe (Dalton and Wattenberg 2000). Some specific features of political systems stimulate or hinder successful VAAs. First, VAAs do not make much sense in two-

party systems in which parties' positions and voters' identifications are clear (Teepe and Hooghe 2005). This might explain why in the United Kingdom or the United States, usually trendsetters in campaign practices, VAAs are absent. It is no coincidence that VAAs are most popular in countries with a (highly) fragmented party-system. Another reason for the absence of a VAA in some countries is related to the mass media. VAAs are mostly sponsored by a newspaper or television station. A VAA without media support would have difficulties financing the system and attracting users. So, in countries in which press freedom is limited and state intervention a common feature, such as Russia and Belarus, we would not expect to see VAAs. The same goes for countries with a highly partisan media landscape: a VAA sponsored by a partisan newspaper or television channel will not attract the floating voter.

Statement Selection for Vote Aid Applications

The spectacular rise of VAAs—both in terms of the spread across countries and of the rise in the number of advices that have been given—stands in sharp contrast to the limited scientific literature on the topic. There has been some attention to the potential *effect* of VAAs on citizens' voting behavior (see for example: Kleinnijenhuis and Krouwel 2007; Kleinnijenhuis et al. 2007; Ladner, Felder, and Fivaz 2008; Laros 2007; Marschall and Schmidt 2008; Van Praag 2007; Walgrave, Van Aelst, and Nuytemans 2008a). Some other work has been done about the *reasons* people participate in VAAs (Boogers 2006; Hirzalla and Van Zoonen 2008) and about who the people are that use VAAs (Hooghe and Teepe 2007; Ladner, Felder, and Fivaz 2008; Teepe and Hooghe 2005). But the scientific literature has remained largely silent about how VAAs work, about their internal *mechanic* and make-up. The contributions focusing on the internal machinery of VAAs can be counted on the fingers of one hand. The most important work has been done by Kleinnijenhuis and Krouwel (Kleinnijenhuis and Krouwel 2007; Kleinnijenhuis and Krouwel 2008). They do not focus on the issue and statement selection process, though, but rather on the dimensionality of the political space in which voters are associated with political parties and on the added value of using weighted statements. In short, they focus on

the internal decision rules of VAAs and their effect on the output. What they call the best "Multi Attribute Utility Decision" (MAUD) model "predicts" a certain vote best (Kleinnijenhuis and Krouwel 2007). For our purpose here, it is interesting to see what they consider as the *best* models. The best models are the ones that better than other models predict a *past* vote for a party. So, the more a voting advice concurs with previous voting behavior, the better the VAA performs, they contend. Kleinijenhuis and Krouwel rely on aggregate percentage distributions: do the advices given by a VAA following a specific MAUD-model result in a distribution of advices that more or less resembles the real electoral strength of the parties? The more a VAA and its internal decision rule manage to generate an election result that approximates the real election outcome, the better it performs. They argue that, although most individual citizens are poorly informed about the issue positions of parties, all these individual idiosyncrasies tend to cancel each other out so that, on an aggregate level, the real elections outcome *does* reflect the issue preferences of the population as a whole (Kleinnijenhuis and Krouwel 2008: 3-4). That is why, according to Kleinnijenhuis and Krouwel, aggregate election results can be used to test for the quality of VAAs' advice. We will come back to that later.

That political scientists have barely bothered to examine how exactly VAAs work and how this affects the advices they produce is surprising as there are many examples showing that VAA outputs are not consistent but differ across time and across VAA. Let us give two examples from the Netherlands, the country with the longest VAA-tradition. First, Van Praag compared the vote advices given by the Dutch VAA *Stemwijzer* in 2002, 2003, and 2006 (Van Praag 2007: 5). He found dramatic differences in parties' share in voting advices between these three adjacent general elections. The leftist party *SP* got 12% in 2002, dropped to 8% less than a year later in 2003, and got 15% of the advices in 2006. The Christian-democrats of *CDA* witnessed an even bumpier ride: 11% in 2002, hardly 3% a year later in 2003, and 15% in 2006. The *PvdA* saw its share of voting advices varying between 10% (2006) and 17% (2003) and the Christian *SGP* between 2% (2002) and 7% (2003). It is hard to believe that, in merely four years, the Dutch parties and the Dutch electorate would have changed so dramatically; it is much more likely is that the particular configuration of statements in the *Stemwijzer* varied strongly between 2002 and 2006 leading to oscillating results advantaging one party in a first election, another in a second, and

yet another one in a third election. Probably not the real world changed but rather the *Stemwijzer* changed in an inconsistent way.

Second, there is the comparison of the advices given to a sample of the same citizens by two different Dutch VAAs, *Stemwijzer* and *Kieskompas,* both for the 2006 Dutch elections (Kleinnijenhuis and Krouwel 2007: 4). Differences are large: 13% of the advices of *Stemwijzer* were for *CDA*, 21% in *Kieskompas*; while *Stemwijzer* sent 28% of the participants to *SP*, this was only 12% in *Kieskompas*; *Stemwijzer* advised 9% of the participants to vote for the *CU*, *Kieskompas* did the same with 17% of its participants. "*Only 43% of the respondents who used both Kieskompas and Stemwijzer received identical advises*" (Kleinnijenhuis and Krouwel 2007: 5). Again the same picture emerges: large differences between the outputs of different VAAs. The point we want to make is simple: VAAs' outcome is not stable but seems to be haphazardly changing for one year to another or from one VAA to the other.

How come that VAAs yield such inconsistent results? Careless and non-systematic statement selection is the most probable culprit, as we will show in the next sections. But therefore, we need to explain how VAAs are built. A VAA building process typically consists of several steps. The most crucial step is the selection of the statements that form the heart of the system. VAA-builders most of the time use different criteria to assess the suitability of statements and to make a selection out of an in principle infinite number of possible statements (for a description of this process in several countries see: Deschouwer and Nuytemans 2005; Krouwel and Fiers 2008; Laros 2007; Marschall and Schmidt 2008). First, statements should be politically relevant, they should deal with important political topics. Second, statements should be diverse and tackle a large amount of different issues. Third, statements should discriminate parties; they should be able to distinguish parties from one another. Fourth, in some cases, parties are given the chance to veto some statements or to negotiate with the VAA-designers about precise statement wording (Van Praag 2007: 5-8). Typically, the statement selection process starts with generating a long-list of statements that is then gradually narrowed down to a short list and then to a final list of statements. In this selection process, the potential outcome—that is the distribution of advices to vote for certain parties—is not taken systematically into account. In other words: in the VAA building process there is no

in-built reality check testing whether the given advices approach in any way the political reality. Consequently, VAAs may advise a disproportionate part of the participants to vote for an extreme or marginal party on the fringes of the party landscape.

To our knowledge, only a single piece in the scientific literature refers to the crucial process of issue and statement selection for VAAs. Kleinnijenhuis c.s. explicitly tackle the matter asserting that some statements in the two most important Dutch VAAs for the 2006 general elections were differentially benefiting left-wing and right-wing parties (Kleinnijenhuis et al. 2007: 42-52). They contend that a non-equilibrated "left-wing" or "right-wing" formulation of statements and a disproportional under- or overrepresentation of issues in these VAAs structurally advantaged some Dutch parties at the expense of others. This can be explained by referring to the literature on issue competition and issue ownership (Budge and Farlie 1983; Klingemann, Hofferbert, and Budge 1994; Petrocik 1989). Some issues favour certain parties at the expense of others. This is the case because these parties "own" the issue at stake: they have acquired a strong reputation on the issue and their stance on the issue is popular (Walgrave and De Swert 2007). On other issues, the same party may have much more difficulty in convincing the electorate. Consequently, during the campaign, parties try to put forward the issues they are strong at and to ignore the issues about which their stance is not really popular. Statement selection for a VAA inevitably implies a selection of issues and thus a mix of favourable and less favourable issues for each party. In the end, it is very unlikely that a certain selection of statements perfectly balances favourable and unfavourable issues for a party and, consequently, the statement selection process almost inevitably leads to an expansion of advices for party A while diminishing the advices for party B. Added to that comes the fact that not only the underlying issues but also statement wording each time advantages some parties and disadvantages others. In sum, statement selection (and statement wording) have the potential to substantially affect the output of a VAA and to skew the distribution of advices.

Naturally, when we speak of a "skewed" output of VAAs the logical question is of course: skewed compared to what? How can we evaluate the statement selection process? The only option is to follow Kleinnijenhuis and Krouwel (2007) and to take

the real election results as benchmark. Elections form the only possible reality check for the output produced by VAAs; we will follow that track in the empirical section of the study. However, we are fully aware of the fact that using real election outcomes to test the validity of the 'virtual' election outcome produced by a VAA's statement configuration is not unproblematic. One can challenge the idea that VAAs should produce a result that resembles the real election outcome. Real votes are affected by many motivations only part of which are based on substantial issues of ideology and program content. People vote for a party because they are used to do so, because they like the party leader, because they think the party will defend their interests best, and, maybe most importantly, because of strategic reasons. VAAs typically only take issues and party programs into account and, hence, it is not more than natural that VAA advices and real election results diverge to some extent. The goal of VAAs is precisely to draw voters' exclusive attention to parties' ideas and to explicitly ignore all other possible vote motivations. If VAAs would really reconstruct the entire and multi-motivational voting process, they would be conservative instruments making people vote for the party they always voted for and they would miss the point. Yet, on the other hand, the distribution of VAAs' advices should not deviate too strongly from the real world of electoral strengths. If voting advices want to be effective they should be realistic and approach the real world choice situation voters face. VAAs that advice large groups of citizens to vote for a tiny party on the fringe of the political spectrum, for example, cannot be considered to be credible and to deliver a reliable vote advice. Hence, we contend that real election outcomes can serve as a useful proxy to test the validity of VAAs. But VAAs should by no means deliver an advice that entirely coincides with the real electoral results. There should at least be some resemblance between what VAAs advice voters to do and what voters have really done in the past. In the empirical part of this paper, therefore, we will test whether some configurations of statements lead to a distribution of voting advices that approximates the real electoral results more than other configurations.

To take up these issues and to test whether statement selection makes a difference in terms of party advice output, we carried out a survey among a random sample of 1,000 Belgian citizens[3]. Data are weighed on party preference (2004) and socio-demographics. A short-list of 50 potential VAA statements was presented to these 1,000 respondents. The 50 statements had been produced by the builders of the VAA *Doe De Stemtest* in 2007—political scientists from four Belgian universities—and were equilibrated in terms of issues and direction. The statements had already been answered to by the political parties. They had been tested on students and were unambiguously formulated. The 50 statements tap a wide range of different issues with more important issues, for example welfare or immigration, being covered by more statements than less important issues, for example EU affairs. Here are a few examples of the statements: '*Also in bars and cafes smoking should be forbidden*'; '*Companies should be forced to hire a certain amount of foreign immigrants*'; '*Gay couples should be allowed to adopt children*'; '*High pensions should be taxed more*'. In sum, these 50 statements match all the criteria that VAA-builders typically use to design a VAA: they are relevant, discriminate parties, cover a wide range of issue domains, and are accepted by the parties. We asked our respondents to agree or disagree (or neutral) with the statements.

A typical VAA consists of +30 statements. We decided to test a large number of configurations of 36 statements by running simulations. Within a total of 50 statements there are 937 *billion* unique combinations of 36 statements. It is of course impossible to test all these configurations. Therefore, we take a large random sample of 500,000 configurations of 36 statements.

[3] The computer assisted telephone survey (CATI) was carried out by *TNS-Media* on a random sample of 1,000 +18 year old Dutch speaking Belgians between 10 and 20 April, 2007. The survey was ordered by a consortium of four universities: University of Antwerp (Stefaan Walgrave), Free University Brussels (Kris Deschouwer), Catholic University Leuven (Marc Hooghe) and University of Ghent (Carl Devos). We thank these colleagues for letting us use these data. The survey was used to benchmark the 2007 version of the VAA *Doe De Stemtest* that was aired by the public broadcaster VRT in May-June 2007. The authors thank the consortium for putting the data at their disposal.

Then, we first run simulations of a simple VAA system on the basis of all the random 36-statement-configurations: statements are *unweighed*. We do not distinguish different dimensions in the statements. We simply calculate distances between parties and respondents: if a respondent agrees with a party on all 36 statements there is a 100% match between them; if a respondent is disagreeing with a party on every statement there is a 0% match. For each respondent we only take into account his or her "first" party, that is the party that most closely approximates his or her own opinions. We consider this as being the "vote advice" that is been given by the VAA. We acknowledge that only considering the first party is a strict criterion; one might think about considering the first two or even the three first parties as the "advice" of a VAA. But for the sake of clarity we limit the output here to the first party, the party that the system calculates to be the closest to a respondent's preferences. Each simulation consists of "pushing" the 1,000 respondents through the basic VAA—consisting of a random sample of 36 random statements—generating a certain vote advice output. So, each simulation generates a virtual election result for the 1,000 respondents with each party getting a share of the votes—that is: a share of the vote advices (first party in the party rank-order). We aggregate all results of all 500,000 randomly sampled simulations into a distribution of advices per party. Finally, these distributions—the virtual elections outcome as produced by a VAA based on a certain configuration of statements—are compared with the real 2004 and 2007 election results.

Second, we follow an identical procedure with the same 500,000 random configuration of statements but this time we *weigh* the statements resulting in a more realistic simulation of what most VAAs really do. Indeed, most VAAs use some kind of weighing procedure; trying to simulate real voting, some statements get more weight than others. Weighing can be done by letting participants decide themselves what topics or statements are more important or by assessing the emphasis parties put on certain issues and topics in their manifestoes. We decide to follow the last track here. All party manifestoes of the Flemish parties were entirely coded following the procedure of the Comparative Party Manifesto project in which each (semi)sentence gets a topic code (Budge et al. 2001). This proportional issue scores are then used to weigh the potential VAA statements about that issue. If a party in its manifesto devotes a lot of attention to the welfare issue, for example, the welfare

statements in the VAA will get more weight; if a respondent agrees with that party on the welfare statements this will boost that party's score in the VAA output this respondent gets. In a nutshell, the weighing procedure we use here consists of an issue specific weigh per party generating a complex system of specific weights per statement and per party. This more complicated VAA system is tested 500,000 times too based on the same 500,000 random configurations of statements. We expect this procedure to generate different results than the simple "add-up" version of the VAA.

In the analyses below we only take into account the five major Flemish parties or cartels competing the 2007 general elections. Apart from the ecological party *Groen!* and the extreme-right *Vlaams Belang* the three main stream parties—socialists, liberals and christian-democrats—entered the electoral arena in a cartel with a smaller party: the *CD&V-N-VA* (christian-democrats), *SP.A-Spirit* (socialists) and *VLD-Vivant* (liberals). Our VAA simulations produce voting advices for each of these parties separately. Yet, to be able to compare with the real election results of 2004 and 2007—we do not know what share of people voting for the CD&V-N-VA cartel actually voted for the CD&V or for the smaller N-VA, for example—we aggregate the advices for both cartel partners into one cartel advice. This also means that we will not deal with *Lijst Dedecker*, a newly founded 2007 liberal party that gained more than 6% of the votes, nor with some other smaller parties competing in 2004 and 2007.

As mentioned above, the core of our exercise is to compare our simulated VAAs' output with the real election results. Our benchmark is the real election result of the parties (or cartels) at the preceding 2004 regional elections and the immediately following 2007 national elections. Table 1 documents the real electoral strength of the Belgian (Flemish) parties in 2004 and 2007.

<Table 1 about here>

We contended earlier that a "good" VAA produces outputs that more or less approximate parties' electoral strengths. What does "approximating the real electoral strength" mean then in operational terms? We calculate two different "target zones" and test how many of the outputs per party are situated in this target zone. The first

"nonstandardized" target zone simply is the range of party scores between both election results; for example for CD&V-N-VA the target zone is 26.1-29.6%. Yet, as one can see, some parties' results in 2004 and 2007 lay very close together (Groen! with 7.6% and 6.3%) while others are much more distant (Vlaams Belang with 24.2% and 19.0%). Moreover, the size of the target zone should vary with the size of the party. That is why we also calculate a "standardized" target zone: that is the zone between a quarter above and a quarter below the average of the 2004 and 2007 results per party. Taking a quarter is of course arbitrary, one might consider different distances but we think it yields a realistic—although admittedly broad— interval that is useful to compare VAA outputs with (see Table 1).

In the next section we examine the distribution of advices to vote for a party as produced by 500,000 different VAAs each time consisting of a different configuration of 36 statements. The two main operational questions are the following:

(1) What are the basic parameters of the distribution per party? What is the average score a party gets and how stretched is the distribution? Does the distribution approach Normality? In other words: how large is the variation in vote shares per party as produced by the 500,000 36-statement configurations? And what are the differences between unweighed and weighed VAAs?

(2) To what extent do VAA outputs differ from the real election outcome in 2004 and 2007? How many of the configurations generate vote advices that are situated in the target zone? The more the real election results per party are distant and marginal in the distribution for this party, the more we can conclude that the average statement configuration tends to over- or underrate a certain party and thus to advantage or disadvantage a party.

RESULTS

We work in first instance with the simple VAAs containing unweighed statements. In the graphs below, we mark the real electoral strength of a party with a green (2007) or a blue (2004) line. We shade the area between both elections. This is the

nonstandardized “target zone”: the further the produced advices are distant from this real world election result, the more the advice produced by the VAA, a configuration of statements, is unrealistic. We also mark the output of the actual selection of statements that was in reality used for *Doe De Stemtest* in 2007 on the graph with a red line. *Doe De Stemtest* worked with *weighed* statements based on the party manifestos and on people's preferences. In the statement selection process the weight of the statements played a decisive role. So, what we indicate on the graphs below is the output of the *Doe De Stemtest* statement configuration in case the statements would not have been weighed, which was not the case.

Each graph contains, per party, the distribution of the advices to vote for that party—the party would appear as “the most close” party if the advice would be based on that particular 36-statement configuration—for all 500,000 simulations. For example, the first graph shows how many of the 36-statement combinations would advice a certain proportion of participants to vote for the party *Groen!*. The average of the distribution is 3.8%. This means that, over all possible 36-statement configurations in our sample, on average 3.8% of the participants would get the advice to vote for the Flemish green party. Comparing that figure with the real elections results in 2007 (6.3%) and 2004 (7.6%) shows that a large majority of configurations result in a smaller amount of advices for Groen! than the real election strength of Groen! would warrant. The histogram also shows the descriptors of the distribution, standard deviation and kurtosis.

<Graph 1 about here>

(1) The graphs show that the distribution of vote shares per party is largely Normal; this is what was to be expected according to the Central Limit Theorem (a large amount of independent statements will aggregate in a Normal distribution). The Normality is grasped by the kurtosis statistic. If the kurtosis approaches zero the distribution is Gaussian. For all parties the kurtosis approximates zero, except for *Groen!* (where it is 7.34). The reason simply is that the distribution of the Flemish green party is heavily skewed due to its average close to zero and the impossibility to have negative values. So the distribution's left tail is cut off leading to a fairly high kurtosis. The right tail of the distribution approaches Normality. That all

distributions are almost perfectly Normally distributed is important: it establishes that for each party there is a central tendency in the 50 statements. The 50 statements tend to produce a voting advice output clustered around an average share of vote advices. Of course, 50 other statements would most likely lead to different distributions with a different mean but they would almost certainly be Normally distributed too with an inbuilt dominant tendency to yield a certain party a certain share of the votes.

(2) The spread of the advices around the mean is relatively narrow—standard distributions are small (between 2.5 and 7.2). The distributions are fairly highly peaked with steep slopes and short tails; only for *SP.A-Spirit* outputs are more dispersed but even there the distribution is relatively compressed and compact. This finding reinforces the point above: there is not only a central tendency in the batch of 50 statements but this central tendency is quite strong and consistent. There are relatively few configurations of 36 statements that deviate very strongly from the mean; any combination of 36 statements produces fairly similar results. On the other hand, there is of course large variation in results. Statement configurations yield very different outputs per party. Consider the results for CD&V-N-VA for example. In some configurations the party hardly gets 5% of the advices in other configurations the party can boast with 32% of advices. Statement selection makes a substantial difference and can boost or curtail a party's advices.

(3) The most important point is that there is a large gap between the real election results of 2004 and 2007 and the advices produced by the simulations. Differences are remarkably large. The mean vote advice output for *none* of the parties approaches the real election results. More, the real electoral strength of the parties seems to be totally unrelated to the central tendency in the advices given by the 500,000 simulations. This is shown in the graph by the fact that the real election results are always situated in the far ends of the tails of the distributions. Moreover, differences between parties are huge. Some parties tend to be heavily overrated; other parties tend to be seriously underrated. *SP.A-Spirit*, for example, gets a strikingly higher amount of advices than what its electoral strength would warrant; parties as *CD&V-N-VA* and *VLD-Vivant* are electorally much stronger in reality than what the simulations suggest. In other words: an overwhelming amount of configurations produce advices that are very far away

from what we called the nonstandardized “target zone”, that is the grey area between.

(4) Is the real *Doe De Stemtest* 2007 statement configuration that has been used in reality to advice a few hundred thousand Belgian voters at the 2007 elections a satisfying selection? We mentioned already that the real *Doe De Stemtest* was based on weighed statements and that this produces significantly different results, but the graphs make it clear that, would it not be for statement weighing, the 36 statements of *Doe De Stemtest* 2007 cannot really be considered as a successful selection approaching the real world election results. The actual 2007 configuration is always situated close to the average, sometimes it produces results that are slightly closer to the real election results (*Groen!*, *CD&V-N-VA*, *Vlaams Belang*) but sometimes the specific configuration that has been used in 2007 is even further away from the elections than the mean (*SP.A-Spirit* and *VLD-Vivant*).

So far we presented results of the unweighed version of the simulated VAAs. Do the more realist simulations based on weighed statements lead to results that approach the real electoral strengths of the parties any better? The answer is negative. We do not show all five graphs per party again but summarize the basic statistics of each distribution per party in Table 2.

<Table 2 about here>

The weighed VAAs produce results that are equally Normally distributed. The means do not systematically approach the real elections results better than the unweighed VAAs. Weighed outputs for *CD&V-N-VA* and *Vlaams Belang* are even more distant from their real electoral strength. Only for *SP.A-Spirit* the mean of the weighed dataset comes a little bit closer than for the unweighed dataset to a realist figure. Interestingly, weighed VAAs lead to slightly flatter distributions with longer tails and a larger standard deviations: the average standard deviation goes up from 4.3 in the unweighed to 5.5 in the weighed simulations. This was to be expected. If some statements get more weigh for some parties chances are high that their presence or absence in the configuration affects the result of this party more and leads to more variation. Together with the finding that the real election results in the unweighed

dataset were situated in the tails of the distributions weighing seems to suggest that more configurations approach real electoral results. But on the other hand, as the spread increases and entirely unrealistic configurations increase in number too, it becomes even more difficult to find a realistic configuration of statements.

The distributional analysis convincingly shows that the output of a VAA—more concretely a random selection of 36 statements—does not produce a result that approximates electoral reality; weighing or not weighing the data does not seem to make a large difference. Average outputs largely differ from the real world. We can make that point as well by calculating what share of the 500,000 configurations produce vote shares that approximate the real elections. In graphical terms: how many of the configurations lay in the nonstandardized target zones or in the standardized target zones? Table 3 contains the results per party.

<Table 3 about here>

On average, for the nonstandardized target zones, only one out of 40 configurations yield a result that, at least for one party, produces a vote share that falls between the real 2004 and 2007 results. For some parties these optimal solutions are as good as absent; see the nearly inexistent target zone configurations for *SP.A-Spirit* (0.2) and *VLD-Vivant* (0.4). For the standardized target zones that are considerable broader (see Table 1) the number of configurations that yield "on target" outputs is higher. But even when we rely on this less strict benchmark the number of "correctly placed" outputs never surpasses ten percent on average. For some individual parties the number of correct configurations reaches 15% (*Vlaams Belang*) but for others it remains very low. Weighing the statements, as it increases the spread of the output, leads in both cases to *less* successfully placed parties in the nonstandardized or standardized target zone. The configurations of statements that do justice to parties' electoral strength are very rare. Picking out 36 statements without taking into account the real electoral strengths and based on a random procedure almost certainly leads to a distribution of advices that diverges starkly from electoral reality.

So far we considered advices per party separately. We found that a large majority of configurations—be it weighed or unweighed—produces outputs that do not at all

approach reality as operationalized in both nonstandardized and standardized target zones. However, in reality VAAs produce outputs for al parties *at the same time*. To what extent are the configurations able to give each party a realistic share of the votes at the same time? To test for this, we calculate for each party and for each unweighed statement configuration the average (absolute) distance in percent between the advice produced and the average of the real elections results of 2004 and 2007. Graph 2 shows the results.

<Graph 2 about here>

On average, the sampled statement configurations produce a voting advice that differs 11.7% from the real (average) election results per party. So, the mean configuration give parties a share of advices that differed strongly—±12% of the votes is a hell of a difference—from what this party gets in real elections. The standard deviation of the distribution is small indicating that the +11% error is a dominant trend and that "better" or even "worse" configurations do not deviate much. Most interestingly, the graph shows that *none* of the 500,000 configurations approaches the 0% error. The single best configuration still generates an output with an average error of more than 4%. This means that there simply is no optimal configuration. None of the configurations generates a result that approximates the real electoral result. As can be seen on the histogram, the real VAA *Doe De Stemtest* was based on a selection that deviated even slightly more from the real world than the average configuration. Weighing the statements does not improve the results, even on the contrary. The average difference of the weighed statements is even larger (12.2%) while the standard deviation remains identical (2.5).

The reason that the optimal configuration for all parties combined does not exist, is straightforward: VAAs are zero-sum games; if a configuration increases the advice share of one party by containing certain statements the advice shares of all other parties are affected too. So, slightly different configurations increasing a party's share of advices disturbs the equilibrium and may remove the other parties further from their target zone. This can easily be demonstrated by correlating the distance between advice and real (average) election outcome for all parties (unweighed dataset). Table 4 has the evidence.

<Table 4 about here>

All correlations between all the parties are highly significant—not really a surprise with an N of 500,000. Many of the correlations are substantial. All parties have at least one direct competitor that competes for the same advices: for all parties there is at least one other party with whom the correlation is larger than .30. The left cartel SP.A-Spirit competes directly with all other four parties. These high correlations, hence, imply for VAA builders that selecting another configuration to increase or decrease a party's share of advices to bring it closer to the real election results immediately affects the whole system and creates unwanted effects pushing other parties further way from their target zone. In sum: statement configurations are instable systems and it is almost impossible to find a satisfying equilibrium far away from the non-realistic mean.

CONCLUSION AND DISCUSSION

The aim of the study was to test empirically whether specific statement selections incorporated in Vote Aid Applications (VAAs) makes a difference. We wanted to start the debate about the make-up and the quality of VAAs among political scientists. To what extent is the output of VAAs—that is: the voting "advice" they produce— affected by the specific statements incorporated in the system? Drawing upon a random sample of the Belgian population and on the answers of thousand citizens to 50 typical VAA statements, we ran 500,000 simulations of VAAs with diverging 36-statement configurations. We alternatively tested unweighed and weighed VAA configurations. Each time we calculated how many Belgian citizens would get the advice to vote for a certain party. We analyzed the distribution of advices per party and compared these with the real election results in 2004 and 2008 and with a standardized "target zone" calculated on the basis of these election results.

We established that statements matter, and that they matter a lot. The outputs of voting advices to vote for certain party differ extensively across configurations. Some

configurations generate hardly any advices to vote for a certain party, other configurations boost a party's advices. Some parties' vote shares more than *sextuple* between the least and the most favourable statement configuration. The real electoral strength of parties does not seem to matter. Distances between the vote advice distributions and real electoral distributions are large. Our simulations show that the real electoral score of a party is mostly situated in the extreme ends of the advice distribution indicating that the mean output of a random configuration would be entirely different than what happens in the real world. The great majority of simulations did not approximate the electoral results at all. There is a central tendency in the statements: there is variation in advice distributions but the variation is clustered around a dominant mean. This suggests that an indiscriminate batch of statements has an inbuilt tendency to favour certain parties in contrast to others. Another batch of statements may have an entirely different built-in partisan bias. It does not seem to be a good idea for VAA-builders to neglect these effects and to act as if statements are just statements and that all selections would inevitably lead to the same or to a similar advice. This clearly is not the case.

Interestingly, weighing statements does not really solve the problem. When weighing the statements based on the parties' attention for issues in their manifestos the output does not really alter strongly. The mean of the distributions hardly moves; and for the parties for which it moves it often moves in the wrong direction—that is: further away from reality. Yet, the spread of the output distributions does increase. Weighing statements make that the number of configurations approaching the real election results increases. Consequently, selecting an optimal statement configuration is at the same time enhanced—chances are larger that there is a configuration that yields outputs closer to reality—as it is made more difficult. Indeed, as there is more variation the chance that one selects a configuration that is even further away from the real world increases too. As the number of "good" configurations rises, so does the number of "bad" configurations.

It may be the case, of course, that our findings are the result of our particular batch of 50 statements we have been working with. These 50 statements—they have all been used to devise *Doe De Stemtest* in Belgium in 2007—may have been particularly biased and other statements and other statement configurations may have produced

results that approximate the real world to a much larger extent. This possibility cannot be excluded. Maybe, other batches of statements are less biased. Yet, the list of 50 statements we have worked with here was not just put together haphazardly, they were not randomly or carelessly selected. In contrast, they had been carefully chosen by a team of five political scientists; they were balanced across issues; they were formulated in a balanced way trying to avoid bias in favour of left- or right-wing parties; they were agreed upon by the Belgian political parties. In short, the 50 statements were selected according to all criteria that are typically used by VAA builders when choosing VAA statements (see above). This does not exclude the possibility that this particular batch of 50 statements was extremely biased, but we do not believe this to be very likely.

Another challenge to our findings is that we compared the output distributions with the real election results. As mentioned earlier, it can be disputed whether election results are the best benchmark to test for a VAA's quality. After reviewing the evidence, we are even more convinced of the fact that there is no alternative than to use real elections. A specific configuration X of statements produces a result for Party A of 20% and for Party B of 10%. Another configuration Y produces exactly the possible advice. This is what we substantiated in the simulations: advices differ strongly across configurations. How to choose then between X and Y? What configuration yields the best VAA? The only feasible and transparent option is to choose the configuration that approximates the elections best. We simply see no other option. If not, the output of a given VAA is arbitrary, random, and based on sheer (bad) luck.

What do our findings mean for VAAs? Are they per definition invalid and unreliable, just an entertaining game that does not inform the voter about the match of his preferences with parties' offer? And, should political scientists refrain from being engaged in building VAAs altogether? Our results show that statement selection is the crux of the VAA-building exercise and that it should be undertaken with the largest possible care; statements are too important to be selected light-heartedly. The carefulness with which political scientists design their scientific surveys stands in sharp contrast to the lightheartedness with which some of them engage in devising

VAAs. As VAAs may have real world consequences one might expect rather the opposite to be the case.

We do not think our results entirely obliterate the idea that VAA can contribute to better informed voters. Our test—comparing VAAs outputs with real election results—also contains a solution to the statement selection problem. Indeed, we plead for a procedure in which the statement selection process is guided by a benchmarking process. Population surveys are excellent tools to benchmark and calibrate the statement selection process. Indeed, one of the best ways to benchmark the statements incorporated in a VAA is to conduct a survey on a representative sample of the population containing the potential VAA statements. If one knows what the population at large thinks about a series of specific statements, it is possible to put together a selection of statements forming a VAA that yields a more equilibrated, more reliable, and more realistic vote advice output. Responsible VAA-builders—to a large extent political scientists—need to know exactly what the statements incorporated in their VAA entail in terms of advice output. If it is unclear and remains untested what voting advices a VAA produces, it is unjustified and even irresponsible to present a VAA as an instrument that helps voters to make their choice. Any other VAA produces entirely different results. The advice given is then just an arbitrary and subjective output that has no scientific grounding and that cannot claim reliability or validity. If VAAs are not benchmarked their output is not valid or, at the very least, it cannot be considered a valuable advice. One might as well ask *Madame Soleil* what party to vote for in stead of asking for a VAA advice. VAAs without proper benchmarking are closer to charlatanism than to political science.

REFERENCES

Boogers, Marcel. 2006. Enquête bezoekers Stemwijzer. In *Tilburgse school voor politiek en bestuur*. Tilburg.

Budge, Ian, and Dennis Farlie. 1983. Party Competition - Selective Emhasis or Direct Confrontation? An Alternative View with Data. In *West European Party Systems. Continuity and Change*, edited by H. Daalder and P. Mair. London: Sage.

Budge, Ian, Hans-Dieter Klingemann, Andrea Volkens, Judith Bara, and Eric Tanenbaum. 2001. *Mapping Policy preferences. Estimated for Parties, Electors and Governments 1945-1998*. Oxford: Oxford University Press.
Dalton, J. Russel, and Martin Wattenberg. 2000. *Parties without partisans. Political change in advanced industrial democracies*. Oxford: Oxford University Press.
Deschouwer, Kris, and Michiel Nuytemans. 2005. De Stemtest van de VRT: een kijk in de keuken. *Samenleving & Politiek* 12:13-31.
Hirzalla, Fadi, and Liesbeth Van Zoonen. 2008. Usage of online voting aid among young people in the Netherlands. Paper read at Voting Advice Applications between Charlatanism and Political Science, May 16th, 2008, at Antwerp.
Hooghe, Marc, and Wouter Teepe. 2007. Party Profiles on the Web. An Analysis of the Logfiles of Nonpartisan Interactive Political Internet Sites in the 2003 and 2004 Election Campaigns in Belgium. *New Media & Society* 9 (6):965-985.
Kleinnijenhuis, Jan, and André Krouwel. 2007. The nature and influence of party profiling websites. Paper read at Politicologenetmaal, 31 May-June 1 2007, at Antwerpen.
Kleinnijenhuis, Jan, and André Krouwel. 2008. Simulation of decision rules for Party Advice Websites. Amsterdam.
Kleinnijenhuis, Jan, Otto Scholten, Wouter van Atteveldt, Anita van Hoof, André Krouwel, Dirk Oegema, Jan de Ridder, and Janet Takens. 2007. *Nederland vijfstromenland. De rol van de media en stemwijzers bij de verkiezingen van 2006*. Amsterdam: Bert Bakker.
Klingemann, H, R Hofferbert, and I Budge. 1994. *Parties, Policies and Democracy*. Oxford: Westview.
Krouwel, André, and Stefaan Fiers. 2008. From text to position. Text-extraction and authorisation of party positions along dimensions of political competition. Paper read at Voting Advice Applications between charlatanism and political science, May 16th, 2008, at Antwerp.
Ladner, Andreas, Gabriela Felder, and Jan Fivaz. 2008. Are Voting Advice Applications more than toys? First findings on Impact and Accountability of VAAs. Paper read at Voting Advice Applications between charlatanism and political science, May 16th, 2008, at Antwerp.
Laros, Jasper. 2007. Kiezen voor Stemadvies. Over de effecten van het gebruik van stemadviesprogramma's, Politicologie, Free University of Amsterdam, Amsterdam.
Marschall, Stefan, and Christian Schmidt. 2008. Preaching to the converted or making a difference? Mobilizing effects of an internet application at the german geberal election 2005. In *Non-party Actors in Electoral Politics*, edited by d. Farrell and R. Schmitt-Beck. Baden-Baden: Nomos.
Petrocik, John. 1989. The Theory of Issue Ownership: Issues, Agendas, and Electoral Coalitions in the 1988 Elections. In *Presented at the annual meeting of the American Political Science Association*. Atlanta.
Swyngedouw, Marc, and Bart Goeminne. 2005. Doe de stemtest: potentieel het machtigste campagne-instrument! *Samenleving en politiek* 12 (3):57-60.
Teepe, Wouter. 2005. Persoonlijke partijprofielen: een analyse vanuit de kennistechnologie. *Samenleving en politiek* 12 (3):2-12.
Teepe, Wouter, and Marc Hooghe. 2005. Interactief Internetgebruik in tijden van verkiezingskoorts. Een analyse van de gebruikers van 'Wij kiezen partij voor u' in 2003 en 2004. *Samenleving & Politiek* 12 (3):73-88.

Van Praag, Philip. 2007. De Stemwijzer: hulpmiddel voor de kiezers of instrument van manipulatie? Paper read at Politicologenetmaal, 31 May-1 June 2007, at Antwerpen.

Walgrave, Stefaan, and Knut De Swert. 2007. Where does issue ownership come from? From the party or from the media? Issue-party identifications in Belgium (1991-2005). *Harvard International Journal of Press and Politics* 12 (1):37-67.

Walgrave, Stefaan, and Peter Van Aelst. 2005. Much ado about almost nothing. Over de electorale effecten van Doe de Stemtest 2004. *Samenleving en Politiek* 12 (3):61-72.

Walgrave, Stefaan, Peter Van Aelst, and Michiel Nuytemans. 2008a. 'Do the Vote Test': The Electoral Effects of a Popular Vote Advice Application at the 2004 Belgian Elections. *Acta Politica* 43 (1):50-70.

Walgrave, Stefaan, Peter Van Aelst, and Michiel Nuytemans. 2008b. Vote Advice Applications as new Campaign Players? The Electoral Effects of the "Do the Vote test" during the 2004 Regional Elections in Belgium. In *Non Party Actors in Electoral Politics. The Role of Interest Groups and Independent Citizens in Contemporary Election Campaigns*, edited by D. Farrell and R. Schmitt-Beck. Baden-Baden: Nomos.

Table 1: real electoral strength of major Flemish parties in 2004 and 2007 (in %)

	2004	2007	Standardized "target zone"
Groen!	7.6	6.3	5.3 – 8.7
CD&V-N-VA	26.1	29.6	20.9 – 34.8
SP.A-Spirit	19.7	16.3	13.6 – 22.4
Vlaams Belang	24.2	19.0	16.3 – 26.9
VLD-Vivant	19.8	18.8	14.6 – 24.1
Other parties	2.7	10.0	—
Total	100.0	100.0	—

Table 2: Weighed versus unweighed simulations (N=500,000); summary statistics of distributions

	Mean		Standard deviation	
Party	Unweighed	Weighed	Unweighed	Weighed
Groen!	3.8	3.5	2.5	2.0
CD&V-N-VA	16.0	12.9	4.1	5.1
SP.A-Spirit	39.2	35.6	7.2	8.2
Vlaams Belang	12.5	11.1	3.6	3.7
VLD-Vivant	6.8	7.0	3.9	8.4
Average	—	—	4.3	5.5

Table 3: Share of configurations (N=500,000) producing outputs per party in both target zones (in %)

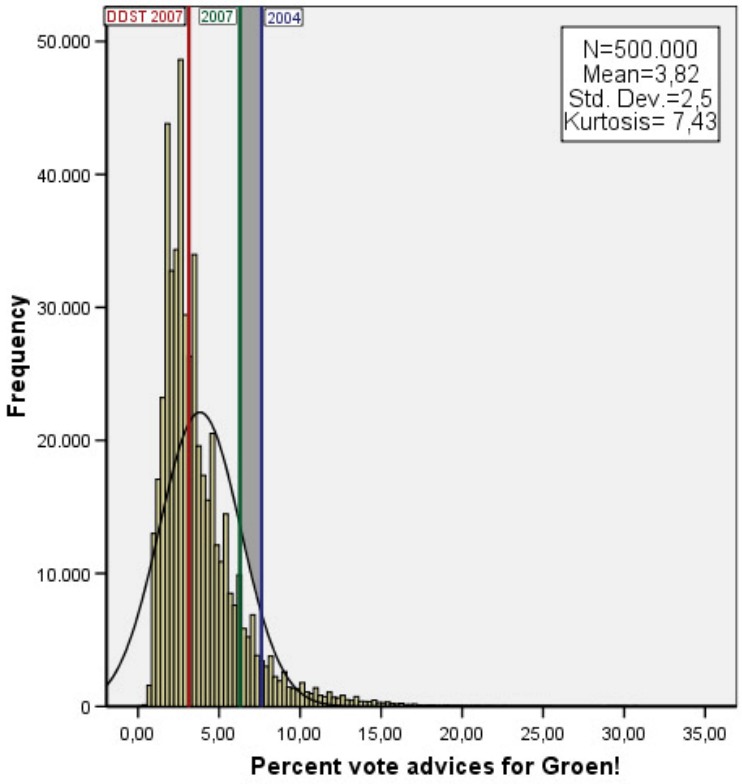
	Non-standardized "target zone"		Standardized "target zone"	
Party	Unweighed statements	Weighed statements	Unweighed statements	Weighed statements
Groen!	5.3	3.0	15.1	11.3
CD&V-N-VA	1.3	0.6	11.8	4.7
SP.A-Spirit	0.2	0.8	0.7	4.2
Vlaams Belang	4.9	2.9	15.3	9.2
VLD-Vivant	0.4	1.2	4.5	12.0
Average	2.4	1.7	9.5	8.3

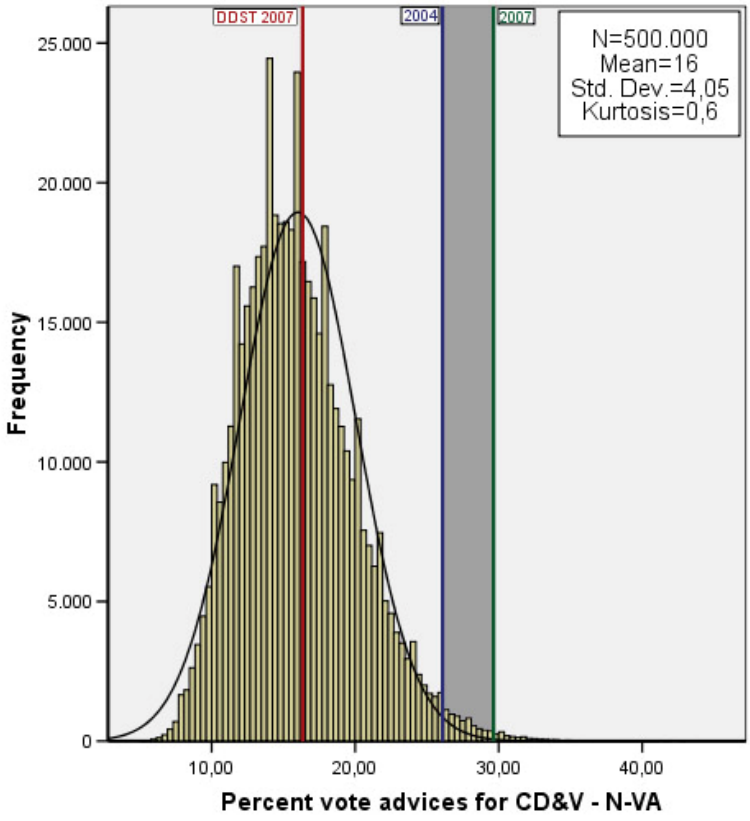
Table 3: Correlations between distance (in %) between advice share and real election result per party (N=500,000)

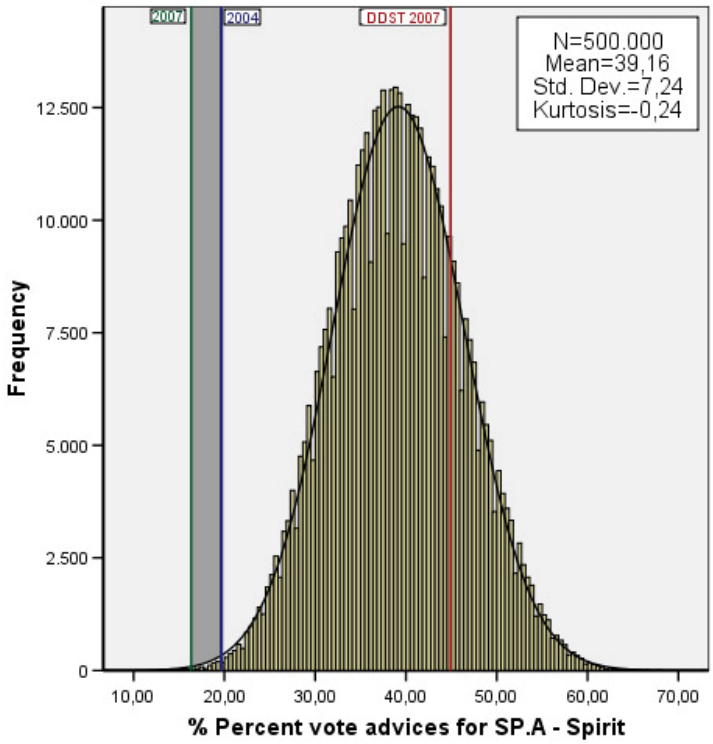
	Groen!	CD&V-N-VA	SP.A-Spirit	Vlaams Belang	VLD-Vivant
Groen!	—	-.036***	-.309***	.188***	-.091***
CD&V-N-VA	-.036***	—	-.492***	.100***	-.192***
SP.A-Spirit	-.309***	-.492***	—	-.393***	-.322***
Vlaams Belang	.188***	.100***	-.393***	—	-.068***
VLD-Vivant	-.091***	-.192***	-.322***	-.068***	—

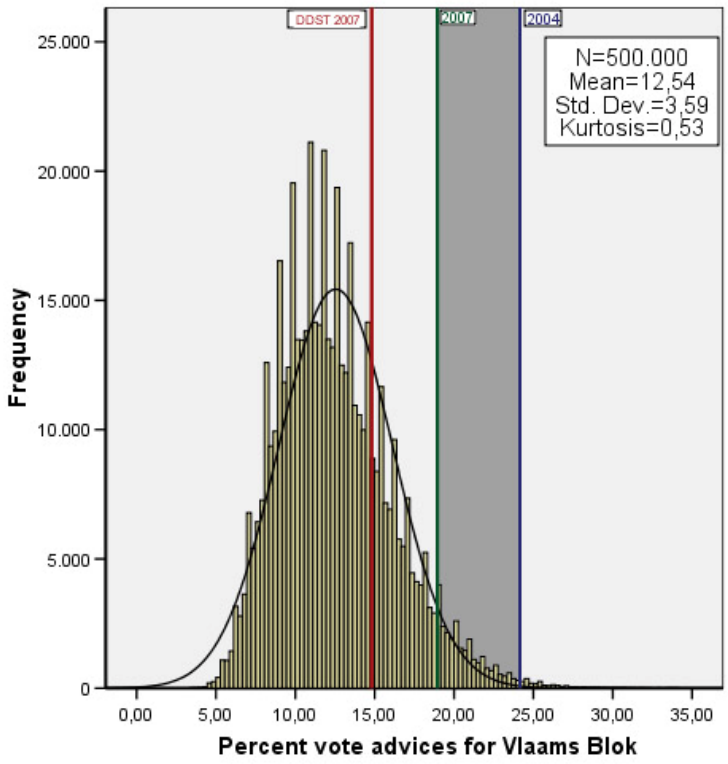
Note: entries in the table are Pearson correlations and significance (***= p<.000)

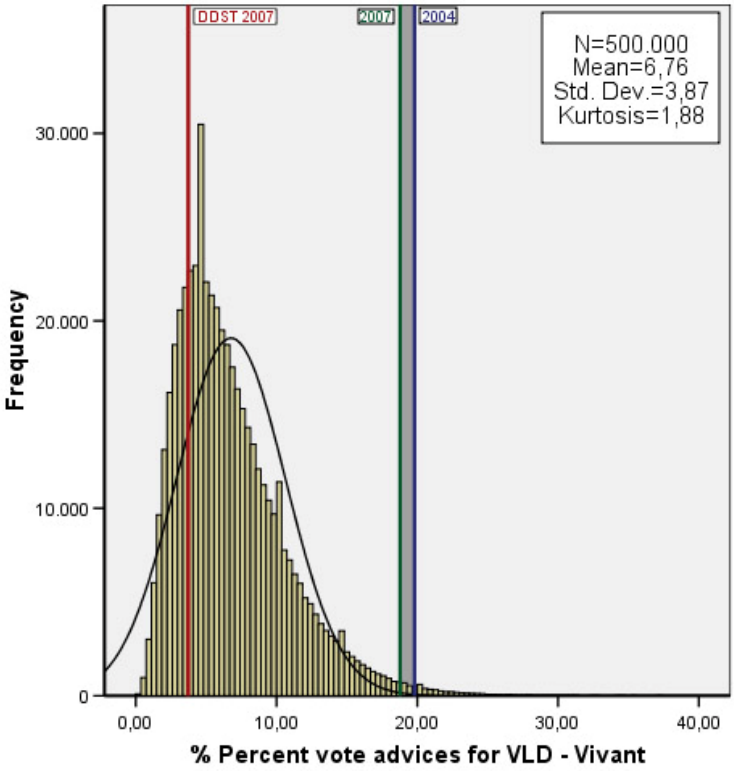
Graph 1: Vote advice output per party






2007
2004
DDST 2007
N=500.000
Mean=39,16
Std. Dev.=7,24
Kurtosis=-0,24
Frequency
12.500
10.000
7.500
5.000
2.500
0
10,00
20,00
30,00
40,00
50,00
60,00
70,00
% Percent vote advices for SP.A - Spirit


DDST 2007
2007
2004
N=500.000
Mean=12,54
Std. Dev.=3,59
Kurtosis=0,53
Frequency
25.000
20.000
15.000
10.000
5.000
0
0,00
5,00
10,00
15,00
20,00
25,00
30,00
35,00
Percent vote advices for Vlaams Blok


DDST 2007
2007
2004
N=500.000
Mean=6,76
Std. Dev.=3,87
Kurtosis=1,88
30.000
20.000
10.000
0
Frequency
0,00
10,00
20,00
30,00
40,00
% Percent vote advices for VLD - Vivant

Graph 2: Average distance in percent between the advice produced and the average of the real elections results of 2004 and 2007