Electoral Studies 36 (2014) 281–289

Contents lists available at ScienceDirect

Electoral Studies

journal homepage: www.elsevier.com/locate/electstud

Voting Advice Applications and electoral turnout

Kostas Gemenis [a, *], Martin Rosema [b]

[a] Department of Public Administration, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
[b] Centre for the Study of Democracy, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

ARTICLE INFO

Article history:
Received 18 December 2012
Received in revised form 19 December 2013
Accepted 21 June 2014
Available online 28 June 2014

Keywords:
Voting Advice Applications
Voting behaviour
Electoral turnout
Entropy covariate balancing

ABSTRACT

In the last two decades Voting Advice Applications (VAAs) have become popular tools among voters, especially in several countries with a multi-party system. In this paper we test if the use of VAAs stimulates electoral participation. We use survey data from the Netherlands, where such tools are widely used. In order to overcome methodological problems of earlier studies, we use techniques that model the effect of confounding variables as a problem of selection into the treatment (VAA usage). We estimate that VAA usage accounted for about four per cent of the reported turnout in the election. The mobilising effect was largest among groups that typically vote in relatively small numbers, such as young voters and those less knowledgeable about politics.

1. Introduction

The rise of the internet paved the way for the widespread use of Voting Advice Applications (VAAs), especially in countries with a multi-party system. By providing voters with information and facilitating their decision making, applications such as *smartvote* or *Wahl-O-Mat* may stimulate electoral participation and thus increase the level of turnout. Because VAAs are still a relatively new phenomenon, not much is known about their effects on electoral turnout. Moreover, most of the few studies conducted so far have neglected some important methodological challenges in estimating the purported causal effects. In this paper we present a more thorough analysis of the impact of VAAs on electoral participation by focussing on the Netherlands, a country where these tools have become very popular.

VAAs can be designed in different ways, but most follow a similar procedure (see Cedroni and Garzia, 2010; Garzia and Marschall, 2014). The developers first formulate a series of statements about policy issues that are expected to be salient in the campaign or that are associated with the main dimensions of political conflict (Lefevere and Walgrave, 2014; van Camp et al., 2014). Next, they estimate the party or candidate positions on those items. This is accomplished by just asking political parties or individual candidates for their positions, by analysing documents such as manifestos or speeches, or by employing expert surveys (Gemenis, 2013; Gemenis and van Ham, 2014). Developers may opt for one of these methods, or combine elements from several of these. When the application is made available to the public, users can indicate their own preferences with respect to the same statements. These answers are then compared to the party or candidate positions and the degree of match or mismatch is calculated and reported to the user. The most frequently adopted approaches are rank ordering the parties in terms of the degree of match between party and user or plotting both parties and users in a two-dimensional political space (Louwerse and Rosema, 2013; Mendez, 2012; Wagner and Ruusuvirta, 2012).

VAAs first appeared as paper-and-pencil tests in the late 1980s and did not immediately attract many voters. This changed after the tools were made available on the internet in the late 1990s (de Graaf, 2010; Ruusuvirta, 2010; Marschall and Garzia, 2014). In several European

* Corresponding author.
E-mail addresses: k.gemenis@utwente.nl (K. Gemenis), m.rosema@utwente.nl (M. Rosema).

http://dx.doi.org/10.1016/j.electstud.2014.06.010
0261-3794/

countries VAAs have become an important element of the election campaign, with usage figures exceeding one quarter of the electorate in Belgium, Finland, Germany, the Netherlands, and Switzerland (Marschall, 2014).

The developers of VAAs have often mentioned a positive effect on electoral turnout as a motive for their launch and several studies appear to support their claim. Nevertheless, the estimated effects of VAA usage on turnout vary. Several studies based their estimates on the self-report by voters. In the 2005 German federal election 8% of survey respondents said that *Wahl-O-Mat* motivated them to vote (Marschall and Schmidt, 2008, 270). In other elections in Germany and Switzerland self-reported mobilisation figures were somewhere up to 15% (Fivaz and Nadig, 2010, 184, Ladner and Pianzola, 2010, 219). Such figures should be interpreted with much care, however, because of methodological problems that we discuss later in this paper. Furthermore, citizens do not choose to use a VAA at random, which means that the causal linkage between VAA usage and casting a ballot can be confounded by other factors.

In this paper we therefore test the hypothesis that the usage of VAAs facilitates vote decision making, and thereby increases the chance of voting vis-à-vis abstention, by taking these methodological challenges into account. Furthermore, we test if these effects occur among specific groups. From here the paper proceeds as follows. In the next section we discuss the potential electoral effects of VAAs with a particular focus on voter turnout. We then turn to methodological issues surrounding the estimation of electoral effects and propose a solution, which we apply to data from the 2006 Dutch Parliamentary Election Study (DPES, 2006). Having discussed the data and method, we proceed with estimating the effect of using a VAA on turnout and investigate whether VAAs are more likely to have an impact on particular strata of the electorate. The paper concludes with a summary of the main findings and discussion of their implications for the functioning of representative democracy.

2. The potential electoral effects of Voting Advice Applications

VAAs can have different sorts of effects on voting behaviour. One potential effect concerns the choice of a particular party or candidate: if voters consult such a website they may remember the outcome of the test and vote accordingly (Ladner et al., 2012; Wall et al., 2012). Another type of effect concerns the question whether citizens vote at all. It is this effect that we focus on in this paper. Does using a VAA increase the chance that a citizen will vote? Research suggests that the internet can active citizens to become politically active (Hirzalla et al., 2010). To understand why it might with respect to voting, two theoretical approaches provide a rationale. The first approach is rational choice theory, which explains the act of voting in terms of a cost-benefit analysis. The second concerns insights from social psychology, in particular the use of heuristics.

Ever since the seminal contribution by Riker and Ordeshook (1968), it has been customary to explain turnout by using a calculus of voting (Blais, 2000). According to this theory, the utility of voting R, is a function of the benefit associated with voting B (effectively, whether the citizen's vote matters for the outcome), conditional to the probability p that the citizen will bring out the benefit B, minus the cost of voting C, plus the psychological benefits of voting D. If we have theoretical reasons to believe that the use of VAAs might have a mobilisation effect, then the use of VAAs must be linked theoretically with either of the right-hand side components of equation (1).

$$R = pB - C + D \tag{1}$$

Firstly, VAAs might increase the perceived utility of the benefit of voting (B in equation (1)). When taking such tests, citizens may become better aware of the differences between parties or candidates and hence realise that it matters who wins the election. The benefits of voting can also become clearer to voters if VAAs show them that a particular party or candidate matches well with their own preferences, which may motivate them to vote (Dinas et al. 2014; Lefkofridi et al., 2014). Secondly, the amount of easily available information provided by VAAs can reduce the costs of collecting information (C in equation (1)), which in turn increases the likelihood of voting (Lassen, 2005). Thirdly, VAAs may increase turnout by strengthening the sense of citizen duty that motivates citizens to vote (Blais, 2000), and therefore increase the perceived psychological benefits of voting (D in equation (1)).

The rational choice approach can lead to relevant insights about human behaviour, but research in psychology shows that the human mind often follows different paths towards a decision than such a cost-benefit analysis. One of the most important insights concerns the use of shortcuts or heuristics (Kahneman et al., 1982). VAAs can be viewed in this light. They provide voters with a simple heuristic to decide for whom to vote and thereby facilitate the decision making, which consequently has a positive effect on turnout. This is relevant in particular for undecided voters, since the inability to reach a decision may be a reason to abstain.

Some people may consider any mobilising effect of VAAs a good thing for democracy, whereas others might argue that the value of this effect depends on what segments of the electorate are affected. Low turnout is considered a problem because certain types of voters are not well represented (Lijphart, 1997; but see Rosema, 2007). If VAAs are able to mobilise such groups, this increases the value of a mobilisation effect. In other areas of research about public opinion it has been shown that some groups are more open to persuasion than others (Zaller, 1992). In the field of VAA research, too, it has been shown that certain types of citizens are more likely to follow the advice of VAAs than others (Ladner et al., 2012; Wall et al., 2012). More specifically, on the basis of previous research we may expect these effects to be largest among younger voters and those with low levels of political interest, since these voters are more often undecided. These are precisely the type of voters that abstain in relatively large numbers (Smets and Van Ham, 2013). If VAAs are capable of mobilising such voters in particular, they reduce unequal participation and thus foster the quality of democracy.

3. Estimating the causal effects of Voting Advice Applications

There are good reasons to believe that the use of VAAs stimulates electoral participation. The question for researchers, then, is how to study if these effects are indeed present and how to estimate their size. The analyses that have been published should be interpreted with much care because of three methodological problems, which are related to the notions of construct validity, external validity and internal validity.

Firstly, the problem of measurement, or what Shadish et al. (2002, 64–82) call 'construct validity', is omnipresent when researchers use data on users' intentions generated by VAAs themselves. The figures reported by Marschall and Schmidt (2008), for example, were derived from a survey among *Wahl-O-Mat* users, which they completed immediately after using the VAA. One may doubt whether the reported effects really occurred. Moreover, psychologists have long disputed that introspection is a valid method to assess reasons for holding certain attitudes or displaying particular behaviour. In the electoral context, too, scholars have found that voters have limited insight in the reasons for their preferences (Rahn et al., 1994). Indeed, Walgrave et al. (2008) compared the impact reported by respondents with their actual voting intentions before and after the use of VAAs. They concluded that often the reported impact is not matched by actual changes in voting intentions, while also observing that effects on voting intentions did not always carry over to actual voting behaviour. Hence, they concluded voters' self-reports are 'a rather unreliable measure of real electoral impact' (Walgrave et al., 2008, 66).

Secondly, even if we take self-reported figures at face value, we should take into account that the data collected by VAAs is largely non-representative of the population of interest. This is the problem of external validity (Shadish et al., 2002, 83–93). Thirdly, if we are interested in making claims about the causal effect using VAAs, we need to compare VAA users to non-users. This leads to another methodological challenge, namely that of internal validity. Such a comparison is impossible with data from VAAs, as these do not include non-users. Alternatively, one could use data from probability samples that include questions about VAA use. But the problem of internal validity remains, because VAA users differ from non-users: the typical VAA user is younger, more highly educated and politically interested than the average citizen. We might observe that VAA users are more likely to vote than non-users, but it will be unclear whether we should attribute this difference to the use of the VAA or to pre-existing differences.

To solve the problem of self-selection we would, ideally, randomize the assignment of citizens to the 'treatment' (use of a VAA) and control groups in a field experiment. However, denying a group of citizens access to a VAA while forcing others to use one, would be unethical and impractical. Field experiments that attempt to circumvent such practical and ethical issues often encounter problems in treatment implementation and compliance (Imai, 2005). Laboratory experiments could be seen as an alternative, but their artificiality leads to problems of construct validity (Shadish et al., 2002, 77–79) as they merely measure intentions regarding turnout and not actual turnout.

Previous studies have responded to these methodological problems in varying ways. Some have chosen to extrapolate from self-reported figures in non-representative samples and report the estimates with a note of caution (Fivaz and Nadig, 2010), others have estimated effects by using representatives samples of the online population and controlling statistically for possible confounders (Marschall and Schultze, 2012), while still others propose to compare individuals in non-representative VAA user samples to individuals in randomly sampled surveys using Heckman selection models (e.g. Pianzola, 2014). We argue that such approaches are not satisfactory. Statistical controls do not tackle the problem of self-selection effectively, whereas the Heckman selection model typically employed in VAA effects on turnout research do not satisfy the 'exclusion restrictions' (the assumption that the variables that affect selection to the treatment should *not* affect the outcome variable) necessary for correct estimation under a Heckman selection model (Sartori, 2003, 115).

We argue here that the statistical techniques that fall under the label of 'matching', that have been employed extensively in estimating the causal effects in observational and experimental studies on voter turnout (e.g. Imai, 2005; Lassen, 2005), can be a good way to deal with these methodological challenges. In its very basic form, matching works by comparing like-with-like individuals from subgroups which resemble treatment and control groups of an experiment. As such, however, matching techniques (just like regression techniques) make the assumption that selection depends on the observable (and measurable) variables (Sekhon, 2009, 496–497). Unlike statistical controls, however, matching is used to explicitly account for selection as a threat to internal validity. Matching therefore does not simply control for confounders (i.e. other factors that influence turnout), but models the underlying process of selection into the treatment by balancing the treatment and control groups on a number of relevant covariates (Shadish et al., 2002, 161–162). Contrary to Heckman selection models and instrumental variables approaches, the matching variables do not need to satisfy any exclusion restrictions. To the contrary, matching variables are selected by virtue of *jointly* affecting the treatment and outcome variables. Moreover, matching estimators are non-parametric (Sekhon, 2009; pp. 492–493, Stuart, 2010, p. 11), a desirable property, since researchers are not required to model the non-linear and interaction effects we would normally expect in the context of turnout.

4. Data and method

We analyse the effect of the use of VAAs on electoral turnout with data from the Netherlands. There are at least two reasons why the Dutch case is interesting and suitable for this study. The first is the widespread use of VAAs in the Netherlands and the second is the availability of national election studies that include measures about VAA use and voting behaviour. The main VAA in the Netherlands is *StemWijzer*, which was introduced in 1989 as a paper-and-

pencil test and went online in 1998. The user figures rose from only 6500 in 1998 to about two million in 2002 and 2003 (de Graaf, 2010). The most visible alternative has been *Kieskompas* (Wall et al., 2012). The main difference between these VAAs is the calculation method and presentation of the advice: whereas *StemWijzer* presents a graph in which parties are listed in the order of the match, *Kieskompas* plots users and parties in a two-dimensional political space (Louwerse and Rosema, 2013). In 2006 *StemWijzer* and *Kieskompas* reported 4.7 million and 1.7 million visitors of their websites, respectively, who completed the test (Aarts and van der Kolk, 2007, 835). Since the total number of votes cast in the 2006 election equalled 9.4 million, if these user figures are accurate, they are quite substantive.

The popularity of VAAs in the Netherlands implies that it is possible to employ a random sample of the Dutch electorate and estimate the effect of VAA use on turnout. We therefore base our analysis on the DPES 2006, a two-wave survey based on face-to-face interviews with a large random sample of the Dutch electorate ($N = 2623$). The survey had a response rate of 67% for the first wave, while 60% of the respondents participated in both waves (Schmeets, 2007). The first interview was held in the weeks before the election of November 22 and the second interview shortly after. The pre-election interview contained questions that can be used as predictors of turnout and VAA usage. Information about the use of VAA as such was obtained in the post-election interview, which asked voters whether they knew any VAA (referred to in the survey as 'vote tests on the internet'), whether they used them, and what party or parties the test(s) advised.

Differences in response rates resulted in modest biases for age, urbanisation, and ethnic background (see Schmeets, 2007). In terms of party choice in the election, on the other hand, the sampled voters were virtually a mirror image of the electorate at large. A more serious bias concerns electoral participation, since abstention as reported by the survey respondents (7%) was much lower among the sampled voters than in the actual election (20%). Voter over-representation is a well-known problem in post-election surveys (Selb and Munzert, 2012) and can be attributed to the three factors, namely selective response (non-voters are more likely to refuse participation in the interview), a stimulus effect (participating in a pre-election interview increases the chance of voting), and social desirability (Schmeets, 2007; Voogt and van Kempen, 2002). Because of the large sample size, however, non-voters can still be meaningfully analysed on the basis of this data set.

Since we are interested in estimating causal effects via propensity score matching, we start with identifying the most relevant variables that are associated with VAA usage and voter turnout. Previous studies have shown that VAA usage is related to demographic characteristics age, gender, and education (Fivaz and Nadig, 2010, 181–182), as well as political interest and party identification (Marschall and Schmidt, 2008, 269). For estimating electoral turnout we rely on the comprehensive meta-analysis of Smets and Van Ham (2013, 356), who concluded that the variables that 'have a consistent effect on turnout […] are: age and age squared, education, residential mobility, region, media exposure, mobilisation (partisan and non-partisan), vote in previous election, party identification, political interest, and political knowledge.'

The DPES 2006 not only includes measures of our two key variables, use of VAAs and voting behaviour, but also contains measures for age, gender, and education, as well as index scores for political interest, political knowledge, and strength of party identification.[1] Comparisons between VAA users and non-users showed no statistically significant differences for gender and education, political interest and strength of party identification, which suggests that the wide use of VAAs in the Netherlands has led to the disappearance of differences observed in other countries (see also Ruusuvirta, 2010, 56–57). We nevertheless include these three variables as the standard advice is to include non-statistically significant variables in the propensity score specification if they are known to be associated with the outcome and the treatment (Caliendo and Kopeining, 2008, 38–39, Rubin and Thomas, 1996, 253, Stuart, 2010, 5). However, we refrain from using an over-parameterised specification by including age squared, residential mobility, region and vote in previous election, since the theoretical and empirical literature specifies that we should not expect these covariates to influence turnout and selection into the treatment.

There are different techniques that can be used for matching on the basis of these variables. Nearest neighbour matching without replacement, where each individual in the treatment group (VAA users) is matched to an individual in the control group (non-users) once some individuals from the treatment group are dropped (individuals outside common support), did not result in a satisfactory matching between our treatment and control groups and left 249 out of the 875 individuals in the treatment group outside common support. Matching with replacement, that results to estimators with the lowest conditional bias (Sekhon, 2009, 497), by keeping most of the observations in the treatment group through matching them to already matched individuals in the control group did not perform satisfactory on various diagnostics either.[2] For these reasons, we employ a pre-processing technique to achieve balance between the treatment and control groups in the covariates of interest. The advantage of this approach is that it does not rely on the iterative process of calculating

[1] The treatment variable (VAA use) was constructed by using the variables tapping the awareness of existence of VAAs (V590) and VAA usage in the 2006 election (V591). We use the variable measuring the highest followed level of education by respondents (V431) to account for the presence of higher education students who are regular users of VAAs. The other variables are gender (V420), turnout (V510), age (recoded from year of birth V421), and three index variables measuring political knowledge on a 0–12 scale (V255), political interest on a 0–4 scale (V015) and the strength of party identification on a 0–7 scale (V065). For details regarding the construction of index variables see DPES (2006).

[2] The diagnostics included t-tests on the equality of means, pseudo-R^2 from the logistic regression, the criterion for less 5% standardized bias after matching (Caliendo and Kopeining, 2008, 48–49), and the Kolmogorov-Smirov test on the similarity of the distributions in the propensity score between treated and controls (Diamond and Sekhon, 2013). Details can be found in the replication files available in the online supplementary material of this article.

Table 1
Assessing covariate balance before and after entropy reweighting.

	Mean			Variance			Skewness		
	Treatment	Control		Treatment	Control		Treatment	Control	
		Before	After		Before	After		Before	After
Age	38.91	52.99	38.91	183.6	272.4	183.1	.4836	.01692	.6013
Female	.4743	51.63	.4743	.2496	.2499	.2495	.103	−.06524	.103
Education	4.177	3.47	4.177	.8462	1.641	.8613	−1.213	−.4848	−1.301
Political knowledge	7.466	6.865	7.466	8.025	8.775	8.77	−.2027	−.1503	−.2624
Political interest	1.798	1.735	1.798	1.045	1.089	1.043	.6758	.5458	.6036
PID strength	2.36	2.731	2.36	4.707	5.619	4.856	.7448	.4184	.6621

Note: Reweighting was calculated on the first moments of covariates using the ebalance package (Hainmueller and Xu, 2013).

the propensity score via logistic regression and checking whether balance has been achieved through a host of diagnostics.

We use the recently proposed entropy balancing technique (Hainmueller, 2012), which can be applied to observational studies with a binary treatment. Entropy balancing has similarities to other recently proposed pre-processing methods (see Ho et al., 2007; Diamond and Sekhon, 2013, Iacus et al., 2012) and works as a reweighting scheme which incorporates covariate balance at a level specified by the researcher. Entropy balancing then 'searches for the set of weights that satisfies the balance constraints but remains as close as possible (in an entropy sense) to a set of uniform base weights to retain information' (Hainmueller, 2012, 26). This technique has been shown to perform well in Monte Carlo simulations and to replicate the causal effects of experimental benchmarks, while it has the attractive property of retaining important information as it does not discard units from either the treatment or control groups (Hainmueller, 2012). Table 1 compares the means, variances and skewness of the treatment to the control group before and after entropy reweighting. The two groups have identical means and very similar variances and skewness after the reweighting, which suggests that a sufficient degree of balance has been achieved.

5. Results

Given the large number of site visits for *StemWijzer* and *Kieskompas*, it is no surprise that many sampled voters indicated that they were aware that such tests existed and that many reported to have used at least one such test. To be more precise, 61% of the respondents in our data set said they knew such tests and 38% indicated they had consulted one or more. As expected, users of VAAs reported lower levels of abstention than non-users, as shown in Table 2. Whereas among non-users the abstention equalled 9%, among users it was only 3%. These figures are in line with our hypothesis that usage of VAAs stimulates electoral participation. However, we cannot exclude the possibility that other factors have been at work. For example, political interest may influence the use of VAAs while also boosting electoral turnout without any direct linkage between VAA use and turnout.

To estimate the causal effect of the use of VAAs on turnout, and hence account for self-selection into the treatment, we regress turnout on VAA usage by using the entropy balancing weights calculated on the basis of the covariates listed in Table 1. The results in Table 3 suggest that there is a clearly discernible effect of VAA usage on electoral turnout, which is also statistically significant. In terms of odds ratios, for users the odds of voting among VAA users are 4.2 times larger than the odds for individuals who did not use a VAA. This suggests that the use of VAAs does indeed stimulate electoral participation.

Since entropy balancing allows for the use of standard statistical techniques in the estimation of causal effects, we implement simulations (King et al., 2000) to assess turnout under different VAA usage scenarios in Table 4. The simulated turnout under a scenario of VAA usage set to 38.3% (as reported in the DPES, 2006) is estimated at 93.3%, a value very close to the DPES 2006 reported turnout in Table 2. If no one used a VAA, turnout would be down to 88.9%, a decrease which is statistically significant. This implies that the usage of VAAs in 2006 increased turnout by 4.4% compared to a scenario where VAAs would not exist. If VAA usage, however, was to be increased to 60% we would not see much of a difference (at least not a statistically significant difference), which implies a ceiling effect. VAA usage would have to be increased to something close to 100% to discern any substantial differences.

We also investigate whether the effect of VAA usage on electoral turnout is conditional on other variables of interest. Could it be that VAAs disproportionately affect

Table 2
VAA usage and electoral turnout in the 2006 Dutch parliamentary election.

	Voted	Abstained	Total
VAA users	97.0 (876)	3.0 (27)	(903)
Non-users	90.6 (1317)	9.4 (136)	(1453)
Total	31.1 (2193)	6.9 (163)	(2356)

Note: Entries are percentages; frequencies in parentheses.
Source: DPES (2006).

Table 3
The effect of VAA usage on electoral turnout.

	Coefficient (linearised std. error)	Odds ratio [95% confidence interval]
VAA usage	1.435 (.248)	4.2 [2.6, 6.8]
Constant	2.091 (.142)	8.1 [6.1, 10.7]

Note: Logistic regression estimates were calculated by using entropy balancing weights on the covariates of Table 1.

Table 4
Simulating the effect of VAA usage under different scenarios.

VAA usage (%)	Turnout (%)	95% Confidence interval
0	88.9	[86.1, 91.4]
38.3 (2006 DPES scenario)	93.2	[91.8, 94.6]
60	95	[93.6, 96.1]
100	97.1	[95.8, 98]

Note: Turnout estimates were calculated based on 1000 simulations using Clarify (Tomz et al., 2003).

younger or politically uninterested voters? To answer such questions we examined the interaction of VAA usage with the covariates found in Table 1 as modifying variables. Because interaction effects cannot be directly assessed by looking at the statistical significance of the interaction term, we follow the recommendation of Brambor et al. (2006) and plot the marginal effect of VAA use on voter turnout across the different values of the modifying variables. The plots in Fig. 1 suggest that the marginal effect of VAA usage on electoral turnout is larger among people at lower educational levels, among younger people, and among those with a weak to no party identification and limited political knowledge. The plots for gender and political interest (not shown here) did not reveal such effects. Of particular interest is that the marginal effect of VAA use on turnout for people over 50, for highly knowledgeable citizens and for party adherents (defined as party identification strength >3 in the DPES, 2006) becomes indistinguishable from zero. This means that VAA effects occur only among particular segments of the electorate.

In order to check the robustness of our results we conducted two additional tests. A placebo test showed that the estimated causal effect is unlikely to be confounded by the presence of unobservable variables. Moreover, the replication of our analysis in low salience election (the 2009 election to the European Parliament) showed that our estimated causal effect is unlikely to be considerably biased due to turnout over-reporting. The results of these analyses are available in the appendix.

6. Conclusions

When the level of turnout in an election is low, or substantially lower than in the previous election, this is considered a serious problem for democracy (e.g. Lijphart, 1997). In this context the widespread use of VAAs, especially in multi-party systems, is an interesting development. These online tools enable citizens to compare their own policy preferences with the competing parties or candidates and find out which of them provides the best match. Does the use of such websites facilitate the decision making and hence boost electoral turnout as is often assumed by the organisations that develop them? We found that we do indeed observe a mobilising effect for the use of VAAs. We estimated by means of simulation that the presence of VAAs was responsible for 4.4% of the reported turnout in the election (as estimated by the DPES, 2006).

This figure is considerably lower than those reported in previous studies. In theory it is possible that we have a unique case with the 2006 Dutch parliamentary election, where effects are smaller than elsewhere. However, given the arguments provided and literature reviewed in this paper, it seems more likely that methodological problems account for this. Estimates of the mobilising effect of VAAs were often based on internet-based opt-in surveys among users, with low response rates. By relying on a large random nation-wide sample, measures of voting behaviour instead of voting intentions, and comparisons between similar respondents through covariate balancing,

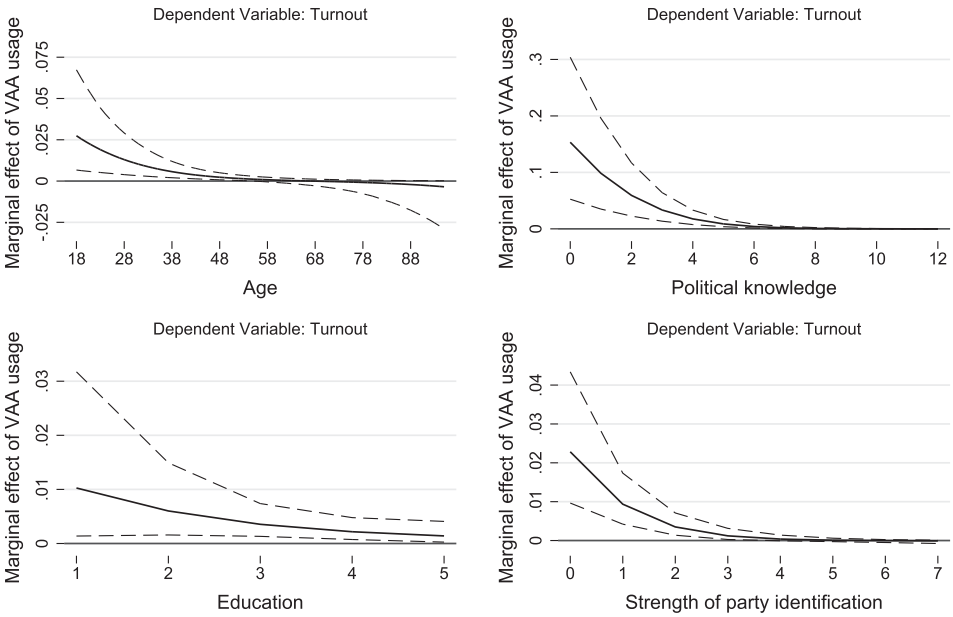

Fig. 1. The marginal effect of VAA usage across different respondent characteristics (95% confidence intervals indicated by dashed lines).

we have created a study that has a higher internal, external and construct validity compared to previous research.

That said, we do not argue that our study is without problems or has no limitations. First, although we have attempted to include the most relevant factors in the matching analysis, we cannot exclude the possibility that other relevant factors, which have perhaps not been measured in the survey, played a role. Our estimated effect rests on the assumption that selection into the treatment is based on the observable variables we have included in our model (Sekhon, 2009). If this assumption does not hold, then there could be potential threats to the internal validity of our inferences. Even though we have assessed the robustness of our estimates, the question remains to what extent the sense of determination to cast a vote has driven citizens to VAAs. Future studies may seek to further improve the quality of the matching analysis by better modelling the selection into the treatment. In addition, the external validity of our estimates is limited in terms of treatments and settings (Shadish et al., 2002, 19–20). We therefore hope that other scholars will pick up the challenges that still remain and examine to what extent mobilisation effects occur and also explore the impact of contextual variation (Garzia et al., 2014). One should realise that in the Netherlands the use of VAAs is widespread. Self-evidently, the effects observed in this case should not be expected in countries where VAAs are hardly used.

In our final words, let us return to the heart of the matter. The presumed mobilising effect of VAAs is among the reasons that developers justify their enterprise and this is why governments may support (the organisations that develop) them. Furthermore, when asked about the effects of VAAs on their own voting behaviour, voters have in many different contexts reported that they feel stimulated in casting their vote. Although the size of the effects may not be as large as the developers have hoped for, or as voters themselves have sometimes reported, our analysis suggests that those considerations and claims are not built on quicksand. Indeed, VAAs do appear to facilitate the decision making and foster electoral turnout. Furthermore, our analysis has shown that these effects are restricted to the younger age cohorts and to citizens with low levels of political information. This is an important fact for debates about internet and political participation. A key question is whether citizens' use of the internet increases the gap between those who are more and those who are less politically active (Hirzalla et al., 2010). It appears that by means of VAAs the use of internet increases electoral participation among young voters and those less interested in politics. Because these are voters that appear to be more difficult to mobilise in elections, the widespread use of VAAs seems to contribute to the functioning of representative democracy, for which the level of electoral turnout is considered an important indicator.

Table 5
Placebo test: the effect of VAA knowledge on electoral turnout.

	Coefficient (linearised std. error)	Odds ratio [95% confidence interval]
Unbalanced model		
VAA knowledge	.673 (.208)	2 [1.3, 2.9]
Constant	2.061 (.105)	7.9 [6.4, 9.6]
Balanced model		
VAA knowledge	.488 (.286)	1.6 [.9, 2.9]
Constant	2.264 (.218)	9.6 [6.3, 14.8]

Note: Logistic regression estimates for the balanced model were calculated by using entropy balancing weights on the covariates of Table 1.

Appendix. Robustness tests

In order to enhance the plausibility of our results we conduct a number of robustness checks. We begin with a placebo test which examines the strength of a purported causal effect in a situation where we would expect this to be zero. A way to do this is to examine an alternative outcome. In the context of VAA usage, a placebo test on alternative outcomes involves the assumption that VAA usage in 2006 cannot predict turnout in the previous 2003 election. If we detect such an effect then we can conclude that it cannot be logically attributed to the VAA use, so it must be attributed to the presence of some unobservable variable that was not accounted for by the matching estimator (Sekhon, 2009, 501). Unfortunately, the context in the Netherlands does not allow us to conduct such a placebo test as we cannot credibly assume a zero causal effect of 2006 VAA usage on 2003 turnout: individuals who used a VAA in 2006 may well have done so in the previous election as well. For this reason, we turn to an alternative treatment placebo test. In the context of VAA usage, non-users can be distinguished between those who are unaware of the existence of VAAs and those who are aware of their existence but choose not to use them. This is akin to distinguishing between two control groups with unoffered and unaccepted treatments respectively (Rosenbaum, 1987). Since the 'aware non-users' and 'unaware non-users' are similar in many ways but not in several others, we treat VAA awareness as an alternative treatment in the placebo test. We therefore regress voter turnout against VAA awareness in Table 5. As we cannot logically assume that VAA knowledge has an effect on turnout, then we can conclude that the detection of such an effect must be attributed to the presence of some unobservable variable that was not accounted for by the matching estimator. Table 5 shows that our choice of matching covariates easily passes the placebo test. The (unlikely) effect of VAA knowledge on turnout which was originally statistically significant becomes statistically non-significant after the inclusion of entropy weights. Encouragingly, this loss of statistical significance is due to the shrinking of the effect and not to an increase in uncertainty as shown by the similar standard errors between Tables 3 and 5. In all, the evidence from this test shows that the estimated causal effect is unlikely to be confounded by the presence of unobservable variables.

An additional robustness test looks at the effect of turnout over-reporting in the 2006 DPES. This is a well-known pattern, that is caused by selective non-response, a stimulus effect, and social desirability (Voogt and van Kempen, 2002; Selb and Munzert, 2012). With regards to

the internal validity of our inferences, turnout over-reporting is important inasmuch it can lead to an over-estimation of the effect of variables that are related in the same direction to both over-reporting and turnout (Bernstein et al., 2001). Unfortunately, the data at hand do not allow us to deal with this problem effectively. Instead, we offer an alternative validity test. Considering that the effect of over-reporting bias is considerably lower in low salience elections (Górecki, 2011), we replicated our analysis using data from the Netherlands 2009 election to the European Parliament (van Egmond et al., 2013). After matching on a similar set of covariates to those in Table 1, the results show a considerable effect (odds ratio: 6.7 [3.75, 11.95]) in the context of the low salience election where over-reporting bias should be less of a concern.

References

Aarts, K., van der Kolk, H., 2007. The parliamentary election in the Netherlands, 22 November 2006. Elect. Stud. 26, 797–837.

Bernstein, R., Chadha, A., Montjoy, R., 2001. Overreporting voting: why it happens and why it matters. Public Opin. Q. 65, 22–44.

Blais, A., 2000. To Vote or Not to Vote: the Merits and Limits of Rational Choice Theory. University of Pittsburgh Press, Pittsburgh.

Brambor, T., Clark, W.R., Golder, M., 2006. Understanding interaction models: improving empirical analyses. Polit. Anal. 14, 63–82.

Caliendo, M., Kopeining, S., 2008. Some practical guidance for the implementation of propensity score matching. J. Econ. Surv. 22, 31–72.

van Camp, K., Lefevere, J., Walgrave, S., 2014. The content and formulation of statements in Voting Advice Applications: a comparative analysis of 26 VAAs. In: Garzia, D., Marschall, S. (Eds.), Matching Voters with Parties and Candidates: Voting Advice Applications in a Comparative Perspective. ECPR Press, Colchester, pp. 11–32.

Cedroni, L., Garzia, D. (Eds.), 2010. Voting Advice Applications in Europe: the State of the Art. ScriptaWeb, Naples.

Diamond, A., Sekhon, S.S., 2013. Genetic matching for estimating causal effects: a general multivariate matching method for achieving balance in observational studies. Rev. Econ. Stat. 95, 932–945.

Dinas, E., Trechsel, A.H., Vassil, K., 2014. A look into the mirror: preferences, representation and electoral participation. Elect. Stud. 36, 290–297. http://dx.doi.org/10.1016/j.electstud.2014.04.011.

DPES, 2006. Dutch Parliamentary Election Study 2006 [Electronic File]. Centraal Bureau voor de Statistiek, Stichting Kiezersonderzoek Nederland, H. van der Kolk, K. Aarts, M. Rosema, M. Brinkman [producers], Data Archiving and Networked Services (DANS) [Distributor]. Persistent Identifier: urn:nbn:nl:ui:13–4zd-x4e.

van Egmond, M., van der Brug, W., Hobolt, S., Franklin, M., Sapir, E.V., 2013. European Parliament Election Study 2009, Voter Study [electronic file]. GESIS Data Archive, Cologne [distributor]. ZA5055 Data file Version 1.1.0. http://dx.doi.org/10.4232/1.11760.

Fivaz, J., Nadig, G., 2010. Impact of voting advice applications (VAAs) on voter turnout and their potential use for civic education. Policy Internet 2 (4), 167–200.

Garzia, D., Marschall, S. (Eds.), 2014. Matching Voters with Parties and Candidates: Voting Advice Applications in a Comparative Perspective. ECPR Press, Colchester.

Garzia, D., De Angelis, A., Pianzola, J., 2014. The impact of VAAs on political participation. In: Garzia, D., Marschall, S. (Eds.), Matching Voters with Parties and Candidates: Voting Advice Applications in a Comparative Perspective. ECPR Press, Colchester, pp. 105–114.

Gemenis, K., 2013. Estimating parties' policy positions through voting advice applications: some methodological considerations. Acta Polit. 48, 268–295.

Gemenis, K., van Ham, C., 2014. Comparing methods for estimating parties' positions in Voting Advice Applications. In: Garzia, D., Marschall, S. (Eds.), Matching Voters with Parties and Candidates: Voting Advice Applications in a Comparative Perspective. ECPR Press, Colchester, pp. 33–47.

Górecki, M.A., 2011. Electoral salience and vote overreporting: another look at the problem of validity in voter turnout studies. Int. J. Public Opin. Res. 23, 544–557.

de Graaf, J., 2010. The irresistible rise of the StemWijzer. In: Cedroni, L., Garzia, D. (Eds.), Voting Advice Applications in Europe: the State of the Art. ScriptaWeb, Naples, pp. 35–46.

Hainmueller, J., 2012. Entropy balancing for causal effects: a multivariate reweighting method to produce balanced samples in observational studies. Polit. Anal. 20, 25–46.

Hainmueller, J., Xu, Y., 2013. Ebalance: a stata package for entropy balancing. J. Stat. Softw. 54 (7), 1–18.

Hirzalla, F., Van Zoonen, L., de Ridder, J., 2010. Internet use and political participation: reflections on the mobilization/normalization controversy. Inf. Polity 27, 1–15.

Ho, D.E., Imai, K., King, G., Stuart, E., 2007. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Polit. Anal. 15, 199–236.

Iacus, S., King, G., Porro, G., 2012. Causal inference without balance checking: coarsened exact matching. Polit. Anal. 20, 1–24.

Imai, K., 2005. Do get-out-the-vote calls reduce turnout? The importance of statistical methods for field experiments. Am. Polit. Sci. Rev. 99, 283–300.

Kahneman, D., Slovic, P., Tversky, A., 1982. Judgement under uncertainty: heuristics and biases. Cambridge University Press, New York.

King, G., Tomz, M., Wittenberg, J., 2000. Making the most of statistical analyses: improving interpretation and presentation. Am. J. Polit. Sci. 44, 347–361.

Ladner, A., Pianzola, J., 2010. Do voting advice applications have an effect electoral participation and voter turnout? Evidence from the 2007 Swiss federal elections. In: Tambouris, E., Macintosh, A., Glassey, O. (Eds.), Electronic Participation. Springer, Berlin, pp. 211–224.

Ladner, A., Fivaz, J., Pianzola, J., 2012. Voting advice applications and party choice: evidence from smartvote users in Switzerland. Int. J. Electron. Gov. 5, 367–387.

Lassen, D.D., 2005. The effect of information on voter turnout: evidence from a natural experiment. Am. J. Polit. Sci. 49, 103–118.

Lefevere, J., Walgrave, S., 2014. A perfect match? The impact of statement selection on Voting Advice Applications' ability to match voters and parties. Elect. Stud. 36, 252–262. http://dx.doi.org/10.1016/j.electstud.2014.04.002.

Lefkofridi, Z., Giger, N., Gallego, A., 2014. Electoral participation in pursuit of policy representation: ideological congruence and voter turnout. J. Elect. Public Opin. Parties 24, 291–311.

Lijphart, A., 1997. Unequal participation: democracy's unresolved dilemma. Am. Polit. Sci. Rev. 91, 1–14.

Louwerse, T., Rosema, M., 2013. The design effects of voting advice applications: comparing methods of calculating matches. Acta Polit. http://dx.doi.org/10.1057/ap.2013.30.

Marschall, S., 2014. Profiling users. In: Garzia, D., Marschall, S. (Eds.), Matching Voters with Parties and Candidates. ECPR Press, Colchester, pp. 93–104.

Marschall, S., Garzia, D., 2014. Voting Advice Applications in a comparative perspective: an introduction. In: Garzia, D., Marschall, S. (Eds.), Matching Voters with Parties and Candidates. ECPR Press, Colchester, pp. 1–10.

Marschall, S., Schmidt, C., 2008. Preaching to the converted or making a difference? Mobilizing effects of an internet application at the German general election 2005. In: Farrell, D.M., Schmitt-Beck, R. (Eds.), Non-party Actors in Electoral Politics. Nomos, Baden-Baden, pp. 259–278.

Marschall, S., Schultze, M., 2012. Voting advice applications and their effect on voter turnout: the case of the German Wahl-O-Mat. Int. J. Electron. Gov. 5, 349–366.

Mendez, F., 2012. Matching voters with political parties and candidates: an empirical test of four algorithms. Int. J. Electron. Gov. 5, 264–278.

Pianzola, J., 2014. Selection biases in voting advice application research. Elect. Stud. 36, 272–280. http://dx.doi.org/10.1016/j.electstud.2014.04.012.

Rahn, W., Krosnick, J.A., Breuning, M., 1994. Rationalization and derivation processes in survey studies of political candidate evaluations. Am. J. Polit. Sci. 38, 582–600.

Riker, W.H., Ordeshook, P.C., 1968. A theory of the calculus of voting. Am. Polit. Sci. Rev. 62, 25–42.

Rosema, M., 2007. Low turnout: threat to democracy or blessing in disguise? Consequences of citizens' varying tendencies to vote. Elect. Stud. 26, 612–623.

Rosenbaum, P.R., 1987. The role of a second control group in an observational study. Stat. Sci. 2, 292–316.

Rubin, D.B., Thomas, N., 1996. Matching using estimated propensity scores: relating theory to practice. Biometrics 52, 249–264.

Ruusuvirta, O., 2010. Much ado about nothing? Online voting advice applications in Finland. In: Cedroni, L., Garzia, D. (Eds.), Voting Advice Applications in Europe: the State of the Art. ScriptaWeb, Naples, pp. 47–63.

Sartori, A.E., 2003. An estimator for some binary-outcome selection models without exclusion restrictions. Polit. Anal. 11, 111–138.
Schmeets, H., 2007. Opzet en uitvoering van het nationaal kiezersonderzoek. In: Aarts, K., van der Kolk, H., Rosema, M. (Eds.), Een Verdeeld Electoraat: De Tweede Kamerverkiezingen Van 2006. Spectrum, Utrecht, pp. 259–274.
Sekhon, S.S., 2009. Opiates for the matches: matching methods for causal inference. Annu. Rev. Polit. Sci. 12, 487–508.
Selb, P., Munzert, S., 2012. Voter overrepresentation, vote misreporting, and turnout bias in postelection surveys. Elect. Stud. 32, 186–196.
Shadish, W.R., Cook, T.D., Campbell, D.T., 2002. Experimental and Quasi-experimental Designs for Generalized Causal Inference. Houghton Mifflin, Boston.
Smets, K., Van Ham, C., 2013. The embarrassment of riches? A meta-analysis of individual level research on voter turnout. Elect. Stud. 32, 344–359.
Stuart, E., 2010. Matching methods for causal inference: a review and a look forward. Stat. Sci. 25, 1–21.
Tomz, M., Whittenberg, J., King, G., 2003. Clarify: software for interpreting and presenting statistical results. J. Stat. Softw. 8 (1), 1–30.
Voogt, R., van Kempen, H., 2002. Nonresponse bias and stimulus effecs in the Dutch national election study. Qual. Quant. 36, 325–345.
Wagner, M., Ruusuvirta, O., 2012. Matching voters to parties: voting advice applications and models of party choice. Acta Polit. 47, 400–422.
Walgrave, S., van Aelst, P., Nuytemans, M., 2008. 'Do the vote test': the electoral effects of a popular vote advice application at the 2004 Belgian elections. Acta Polit. 43, 50–70.
Wall, M., Krouwel, A., 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 Polit 20, 416–428.
Zaller, J.R., 1992. The Nature and Origins of Mass Opinion. Cambridge University Press, New York.