Getting out the Vote with Voting Advice Applications

Micha Germann Kostas Gemenis*

Abstract

This paper investigates the mobilization potential of online voter information tools known as 'Voting Advice Applications' (VAAs). We argue that an observational approach utilizing survey data constitutes the best available method for causal inference where VAAs are popular—and we are thus most interested in VAA turnout effects—because randomized experiments are likely to run into double cross-over problems. We suggest that matching offers key improvements over existing methods to tackle self-selection into VAA use in observational studies. To improve confidence in selection on observables, we complement matching estimates with an extensive sensitivity analysis, including a placebo test. Empirically, we study the effect of *smartvote*, a popular VAA from Switzerland, on turnout in the 2007 Swiss federal election. We find that *smartvote* usage significantly increased the individual-level probability to vote. Our results suggest that *smartvote* was, on the aggregate, responsible for about 1.2 percentage points of the total tally with an estimated cost of nine Swiss Francs (7.5 US dollars or 1.4 Big Macs) per additional vote. Promising as well, we find that the mobilization effect was more pronounced among younger voters. Our findings point to the value of VAAs compared to traditional get out the vote tactics.

Keywords: Voting advice applications, electoral turnout, matching, get out the vote

*Micha Germann is a Postdoctoral Prize Fellow at the Department of Politics, Languages & International Studies, University of Bath, United Kingdom. Kostas Gemenis is Senior Researcher at the Max Planck Institute for the Study of Societies, Cologne, Germany.
Address correspondence to Micha Germann, 1 West North, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom. E-mail: micha.germann@outlook.com.
Supplemental materials for this article can be accessed on the publisher's website at [...]

1 Introduction

The declining electoral turnout rates across established democracies are often seen as a sign of citizen disengagement and have led to concerns about waning democratic legitimacy. Since some groups of voters are more likely to abstain than others, declining electoral turnout is also likely to increase representative deficits. Consequently, numerous proposals have been made on how to increase political participation, including various forms of electoral engineering and civic education. Another large strand of literature focuses on more short-term methods of voter mobilization: 'get out the vote' tactics such as door-to-door canvassing, telephone calls, and direct mailing campaigns (Green & Gerber, 2008). This paper investigates the turnout effect of a comparable short-term method of voter mobilization: online voter information tools known in the literature as 'Voting Advice Applications' (VAAs).

VAAs have become central features of electoral campaigns in several advanced democracies in recent years. Especially in Western Europe, increasingly large numbers of citizens turn to these applications before elections to find the party or candidate that best matches their preferences. In countries such as the Netherlands, Finland, Germany, and Switzerland, VAAs have therefore come to perform a function similar to the traditional media in terms of relaying information about the policy positions of parties and candidates to the voters.

VAAs could make a welcome contribution to the re-engagement especially of young voters (e.g. Dinas, Trechsel, & Vassil, 2014; Fivaz & Nadig, 2010). However, several methodological challenges render existing empirical estimates of the VAA mobilization effect based on observational data uncertain, given self-selection into VAA use. In response, the literature has recently turned to experimental designs. However, while experiments constitute a useful method to tackle self-selection in contexts where VAAs have limited popularity, the experimental gold standard is likely to stumble on problems with treatment implementation and control group resentment in countries where VAAs are widely used. The resulting 'double cross-over' problem constitutes a significant threat to the internal validity of randomized experiments (Freedman, 2006; Shadish, Cook, &

Campbell, 2002).

Given the likely double cross-overs in experimental studies, this paper argues that despite self-selection into treatment an observational approach utilizing survey data remains the best available method for causal inference in contexts where VAAs are popular—and, by implication, their electoral effects most policy-relevant. We make two suggestions for how to improve causal estimates in observational studies. First, we argue that matching offers key improvements over previously used methods to tackle selection bias in observational studies because it reduces model dependence and allows for straightforward checking of covariate balance between treatment and control groups. Second, we propose to complement matching estimates with an extensive sensitivity analysis so as to improve confidence in selection on observables. We illustrate our suggestions with an analysis of the effect of *smartvote*, a popular VAA from Switzerland, on turnout in the 2007 Swiss federal election. We find that *smartvote* usage significantly increased the individual-level propensity to vote. The effect estimate passes a large number of robustness checks, including a placebo test.

In addition, our paper contributes to the literature by framing the estimated VAA turnout effect more broadly in the context of the extant literature on voter mobilization. Compared to many traditional get out the vote tactics, VAAs provide much richer nonpartisan and, crucially, personalized political information. Moreover, VAAs are disseminated online and thus offer significant economies of scale. Accordingly, we find that what constitutes a small effect at the individual level translates to a relatively cost-effective tactic at the aggregate level. Promising as well, we find that the VAA mobilization effect was more pronounced among younger voters.

2 Voting Advice Applications and Electoral Turnout

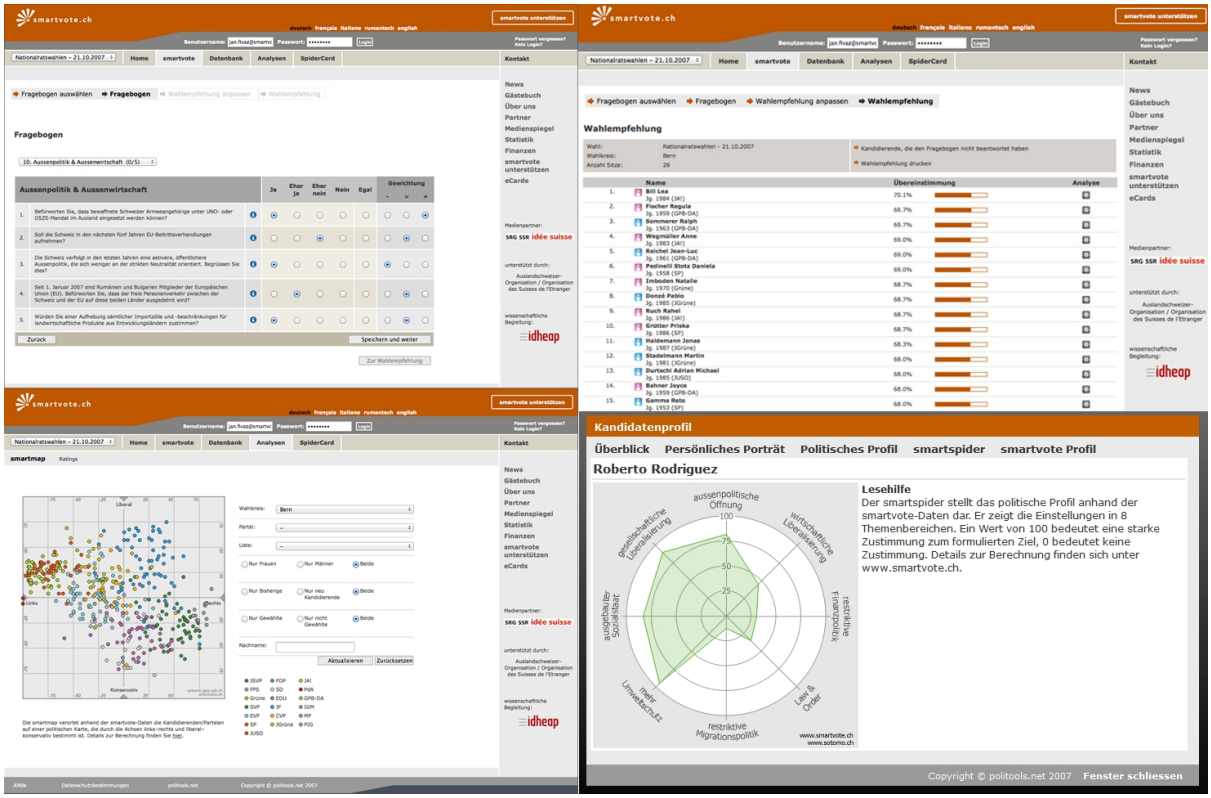
Early studies in political communication showed that the Internet may enhance voter information about candidates and elections, and in turn stimulate increased political participation (e.g. Tolbert & McNeal, 2003). Newer studies suggest that social media can

promote political participation, including electoral turnout (e.g. Gil de Zúñiga, Jung, & Valenzuela, 2012). With the possibility of conducting elections online, the Internet also promises to reduce the transaction costs involved in voting (e.g. Germann & Serdült, 2017). VAAs constitute an additional digital innovation that may promote voter turnout.

VAAs are nonpartisan online tools that are made available before elections to advise voters about their ideological congruence with political parties or candidates. They are typically designed by political scientists, often in cooperation with media outlets, NGOs, or national civic education agencies. Although there is some variation with regards to the methodological design of these tools, their basic mode of functioning tends to be similar. VAA users are asked to indicate their opinions on a number of policy items, typically in the form of questions with an agree/disagree type of response scale. Upon completion of the questionnaire, the application measures the extent to which a user's stated preferences match with previously estimated preferences of political parties or candidates. This is then presented to the user, typically in the form of a rank-ordered list of suitable vote choices and/or a political map showing the positions of both the user and the parties or candidates (see Figure 1).

VAAs have by now established themselves as an integral part of electoral campaigns in a number of countries, especially in Western Europe. As tools that are designed to assist citizens in navigating the political landscape and to gather information about politics, VAAs are located at the intersection of communication flows between political parties (or candidates) and the citizens. As such, VAAs increasingly complement the traditional mass media as a means of political communication in the context of election campaigns (Krouwel, Vitiello, & Wall, 2014). Perhaps the clearest evidence for the substantial and growing role of VAAs in electoral campaigns comes from user statistics. VAA usage ranges from a low of roughly 3% in the 2009 elections to the European Parliament over 10% to 20% in recent national elections in Canada, Finland, Germany, Greece, New Zealand, and Switzerland, to a third of voters in recent Dutch and Danish national elections (see Figure 2). In countries where longitudinal data is available, survey data suggests that the popularity of VAAs has generally increased over the years. VAAs have been

Figure 1: The 2007 version of *smartvote* (screenshots)

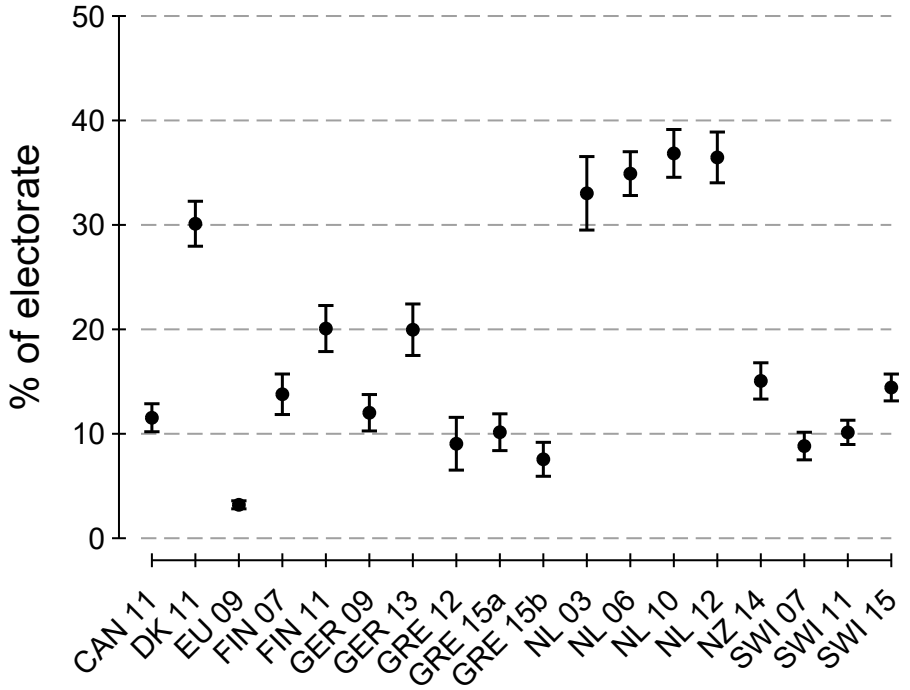


Note: Clockwise from top left the screenshots show the *smartvote* questionnaire and three types of voting recommendations.

made available in several other countries not included in the figure, but these tend to attract fewer users and no representative survey data on the number of users is available. Examples include France, Portugal, Spain, the UK, and the US.

There are several mechanisms by which VAAs may affect turnout (Dinas et al., 2014; Gemenis & Rosema, 2014). Two mechanisms relate to the information function of VAAs. Similarly to the traditional mass media, VAAs provide citizens with easily accessible information on the stances of political parties or candidates. However, VAAs go one step further in that they provide users with *personalized* information on their ideological congruence (or 'match') with different parties or candidates. Hence, VAAs may significantly reduce the cost associated with voting in terms of gathering relevant information. Second, the type of information displayed by VAAs may also increase the instrumental benefit of voting. By highlighting suitable choices to voters, VAAs may foster weak pre-existing preferences or even lead to the formation of new preferences. Moreover, VAAs may help

Figure 2: VAA usage in selected countries



Note: The dots (means) and spikes (95% confidence intervals) represent estimates of the share of eligible voters who used a VAA during selected national and supranational election campaigns. See Appendix A for more details.

voters realize that some parties or candidates are far removed from their ideological position, thus helping them see the relevance of different election outcomes and spark political interest.

Furthermore, the interaction with the tool may also increase the psychological benefit of voting. The psychological satisfaction resulting from adherence to the social norm of participation is widely recognized as a leading explanation for electoral turnout. Existing research shows that a treatment as simple as reminding citizens that voting constitutes a civic duty can significantly increase electoral turnout (Gerber, Green, & Larimer, 2008). Similarly, prolonged interaction with a VAA may remind voters that they are expected to vote in the upcoming election. Finally, VAAs may also give rise to discussions among peers, as illustrated by Fivaz and Nadig (2010, pp. 187–189) for the case of Switzerland. Political discussions stimulated by VAA use may lead to further information gains that VAA users would not have acquired otherwise. Moreover, social interactions could also

further reinforce the perception that voting constitutes a civic duty.

3 Existing Evidence

The question whether VAAs make citizens more likely to vote has been examined across several countries, using different data sources and research designs. However, extant research continues to grapple with a series of methodological challenges. A total of four generations of studies can be distinguished, each with increasing levels of methodological sophistication.

The first and most basic generation of studies drew inferences based on self-reported impact measures. Directly relevant to the Swiss case we focus on is Fivaz and Nadig (2010). Drawing on an opt-in survey administered to *smartvote* users, Fivaz & Nadig report that approximately 40% of respondents indicated that taking the VAA positively influenced their decision to vote. Extrapolating this figure to the general population, they come to the conclusion that '150,000 [*smartvote* users], or 6 percent of Swiss voters, were positively influenced in their decision to cast their vote in the 2007 elections' (p. 184). Marschall and Schmidt (2008) report results from a similar study in the German context.

Clearly, first generation estimates of the VAA turnout effect must be interpreted with great care. People are not known to be the best judges of their own behavior. Self-reported impact measures are likely to over-report the effects of VAAs (Walgrave, van Aelst, & Nuytemans, 2008). Furthermore, opt-in surveys involve self-selection. Thus, the second generation of studies turned to representative election surveys. Election studies allow for direct comparisons of participation rates between VAA users and non-users; and they also reduce the problem of self-selection into the survey. Nevertheless, self-selection continues to be a problem: survey respondents self-select into the treatment condition (VAA usage) and the determinants of the self-selection process are likely to be associated with turnout. For example, VAA users tend to be disproportionately young, well-educated, and politically interested (Marschall & Schmidt, 2008), and these factors

are likely to affect electoral participation as well (Smets & van Ham, 2013). To account for this, second generation studies resorted to multiple regression. Marschall and Schultze (2012), for example, found that after controlling for various socio-demographic and attitudinal variables, users of the German VAA *Wahl-O-Mat* were between 1.3 and 2.3 times more likely to turn out in a federal election. Employing an analogous design, Kruikemeier, van Noort, Vliegenthart, and de Vreese (2014) found that use of *StemWijzer* significantly increased participation in the 2010 Dutch national elections.

However, the multiple regression techniques used in the second generation of studies may be subject to two key problems. First, regression assumes selection on observables. Thus, we cannot exclude the possibility that results are driven by unobserved factors. Second, regression adjustment does not effectively tackle the problem of self-selection into treatment because it does not model the self-selection process directly. Standard regression techniques may therefore fail to account not only for unmeasured, but also for measured imbalances between VAA users and non-users and could lead to potentially severe problems with extrapolation and model dependence (Ho, Imai, King, & Stuart, 2007).

In an attempt to rectify this, the third generation of studies turned to selection models. Selection models involve the estimation of a two-equation system wherein the first equation models the self-selection process and the second equation models the outcome including a correction factor that accounts for the disturbances between the two equations. Selection models have two crucial advantages: i) they explicitly model the self-selection process and ii) the inclusion of the correction factor in the outcome equation allows to control for unobserved confounders. However, selection models require at least one valid instrument—a variable that is related to the treatment condition but exogenous to the outcome conditional on the treatment (exclusion restriction). We argue that finding a variable that plausibly satisfies the exclusion restriction in the context of election surveys, where it is known that everything hangs together with everything else, is close to impossible. Consider Garzia, de Angelis, and Pianzola (2014), who proposed to instrument for VAA usage with gender and left-right self-placement. Using this procedure, Garzia

et al. found that VAA usage had a positive impact on electoral turnout in Finland, Germany, the Netherlands, and Switzerland. However, even if not statistically significant in one-stage regressions, there are plausible theoretical arguments that gender and ideology have an independent effect on turnout.[1] In the absence of a valid instrument that satisfies the exclusion restriction, selection models are identified exclusively on the basis of strong distributional assumptions regarding the error terms of the two equations and have been shown to often yield increased rather than decreased bias (Stolzenberg & Relles, 1990). Given this, the suggestion to instrument for VAA use in the context of election study data is likely to be of little practical relevance.

Like elsewhere in the voter mobilization literature, problems with causal inference in observational studies motivated a turn towards experimental designs. The fourth generation of studies was pioneered by Vassil (2012), who conducted a randomized experiment in the context of the 2009 elections to the European Parliament among Estonian participants recruited through a university email list. Vassil found a small yet statistically significant effect on turnout in the 2009 European elections, which could be indicative of how small VAA effects are if stripped of selection bias. In a similar survey experiment, Garzia, Trechsel, and De Angelis (2017) addressed issues surrounding external validity by focusing on respondents recruited through the Italian election study. Garzia et al. reported significantly higher turnout among treated respondents. A similar survey experiment conducted in Hungary, however, found only weak evidence in favor of a VAA mobilization effect (Enyedi, 2016). Finally, a field experiment conducted before the 2014 provincial elections in Quebec using a convenience sample reported no effect on turnout (Mahéo, 2017).

The newest, experimental wave goes a long way towards alleviating concerns about self-selection into VAA usage. However, the countries where experimental designs have been relatively successfully deployed (Estonia, Hungary, Italy) are all countries with rather low rates of VAA usage. Implementing a similarly clean-cut experimental design is difficult, if not impossible, in countries where VAAs are highly popular. If a fifth or even more of the voting population uses a VAA before the election, it becomes likely that

many participants take a VAA even if they are assigned to the control group ('always takers'). Moreover, survey experiments can suffer from treatment delivery problems because it cannot be guaranteed that respondents assigned to the treatment group actually take the treatment; therefore, many respondents may not take the VAA even if they are assigned to the treatment group ('never takers'). Even in field experiments where researchers are actively monitoring treatment delivery, problems in treatment adherence can occur as VAA usage in less than ideal conditions (e.g. on the street, as in the Quebec experiment) does not allow respondents to fully reflect on the provided information. In combination, problems with treatment delivery and control group resentment are likely to lead to extensive double cross-overs, thus threatening the experiment's internal validity (Freedman, 2006; Shadish et al., 2002). In theory, instrumental variable techniques can be used to address double cross-overs by using treatment assignment as an instrument for VAA use. However, in countries where VAAs are popular treatment assignment is bound to be only weakly correlated with VAA use, giving rise to a weak instrument problem (Murray, 2006).

In sum, our review of the existing empirical literature suggests three conclusions. First, randomized experiments constitute the gold standard for causal inference: they offer the most powerful method to overcome self-selection into VAA use and estimate turnout effects free of confounding. However, and second, the gold standard is difficult to implement in countries where VAAs are popular because extensive double cross-overs are probable. Third, observational studies utilizing survey data therefore remain the best available method for causal inference in countries where VAAs are popular. However, observational studies have thus far been unsuccessful in finding a convincing solution to the problem of self-selection into treatment.

4 Using Matching to Estimate the VAA Turnout Effect

As the countries where VAAs are popular are also those where VAAs and their effects matter most, we believe that observational work remains valuable as a complement to well-designed experiments. We argue that matching, a method that has thus far largely been ignored in the VAA literature (though see Gemenis & Rosema, 2014), offers at least two key improvements over existing approaches to dealing with self-selection in the context of election surveys. First, contrary to regression adjustment, matching directly models self-selection into VAA usage. Specifically, matching imitates the experimental ideal in that it attempts to directly equate the distribution of measured confounders in the treated and control groups (Stuart, 2010). As a result, compared to regression adjustment matching often leads to higher unit homogeneity and reduced model dependence. Treatment and control groups become more similar, thus ensuring that subsequent comparisons between treated and untreated subjects (e.g. VAA users vs. non-users) are not confounded by differences in covariate distributions. Due to this, matching can significantly improve the quality of causal estimates (Dehejia & Wahba, 1999).

Second, simple diagnostic tests such as t-tests can be used to check the degree of covariate balance between treated and control groups after matching (Austin, 2011). If these tests suggest that important systematic differences remain, the researcher can modify the matching solution with a view to achieving acceptable balance, for example, by adding interactions between covariates to the model, or by reducing the maximum acceptable distance between treated and control groups. If in a given application no acceptable matching solution can be achieved, this alerts the researcher that there is insufficient overlap between treated and control groups—that they are effectively drawn from distinct populations and, by implication, that any causal estimate will necessarily be highly model-dependent. By contrast, it is much more difficult to determine the extent of covariate balance in the context of regression models and a researcher may proceed with a regression-based analysis without being aware that causal estimates are based on

extrapolation between distinct populations (Ho et al., 2007).

Of course, matching is not a magic stick. In particular, just like regression, matching cannot remove any hidden bias caused by unobserved confounders. Unobserved confounding could in theory be overcome by instrumenting for VAA use. However, as argued, the search for a valid instrument is likely to remain unsuccessful. Ignorability seems unavoidable. Accepting this, we suggest that the best option for researchers interested in VAA turnout effects is to subject causal estimates to an extensive sensitivity analysis. In particular, we argue that methods such as placebo tests can often be usefully deployed to assess the plausibility of selection on observables.[2] Similarly to sugar pills in medical research, placebo tests in the observational realm involve the testing of an alternative treatment or an alternative outcome for which there cannot logically be a treatment effect, assuming selection on observables holds (Sekhon, 2009). If the controls are able to wipe out the placebo effect, this increases confidence that the result is not due to hidden bias. If not, this indicates that the model is biased. In the context of VAA turnout research, Gemenis and Rosema (2014) implemented a version of a placebo test involving an alternative treatment—mere knowledge of the existence of VAAs, as opposed to actual usage—that should not affect turnout after conditioning on covariates. However, few election surveys include items on knowledge of the existence of VAAs. Below, we suggest an alternative placebo test involving an alternative outcome—past turnout in elections where no VAA was available—that may be more widely available.

Next, we proceed to illustrate our suggestions and apply matching to estimate the effect of usage of the Swiss VAA *smartvote* on electoral turnout (compare Figure 1 above for example screenshots of the *smartvote* application). In many ways, *smartvote* constitutes an ideal case for our undertaking. Most importantly, *smartvote* is one of the most established and popular VAAs (see Fivaz & Nadig, 2010). Moreover, several prior studies attempted to estimate *smartvote*'s effect on electoral behavior with diverse methods, including self-reported impact measures (Fivaz & Nadig, 2010), selection models (Garzia et al., 2014), and a (field) experiment that exemplifies the difficulties of the experimental approach in a country such as Switzerland where VAAs are popular and treatment non-

compliance thus prevalent (Pianzola, Trechsel, Gchwerdt, Vassil, & Alvarez, 2012). We address possible limitations of the Swiss case in terms of generalizability in the conclusion.

5 Data and Methods

We draw data from the Swiss Electoral Study (Selects) while focusing on the 2007 federal election as 2007 is the only election year that allows us to both i) conduct a placebo test using turnout in past elections as fake outcome and ii) draw inferences free of post-treatment bias (see below). Below we show that we get similar results for the 2011 and 2015 federal elections.[3] The data for Selects 2007 was collected through computer-assisted telephone interviewing (CATI) and includes a total of 4,392 respondents, with some of the cantons (regions) being oversampled. As is common in election studies, Selects overestimates electoral participation (69% vs. 48%). There is also a slight overestimation of leftist voters. Selects includes survey weights to correct for the oversampling of selected cantons, as well as turnout and party choice bias. As is customary in the literature, we begin by estimating the causal effect in the unweighted sample, but adjust for the sample weights in subsequent analyses that generalize to the general population.

The dependent variable in our study is self-reported turnout in Switzerland's 2007 federal election, as measured by Selects. Unfortunately, validated federal election turnout data is not available in Switzerland. As a result, our results could be affected by vote overreporting due to social desirability. That said, evidence from an exceptional case where a Swiss canton enabled vote validation in the context of a regional referendum shows that the overestimation of electoral turnout in Selects likely stems primarily from the overrepresentation of voters, and not vote overreporting (Sciarini & Goldberg, 2016). In addition, we counter threats to causal inference due to social desirability by accounting for known predictors of vote overreporting, such as political engagement and education (Ansolabehere & Hersh, 2012). We return to the issue of vote overreporting below in the context of the placebo test.

5.1 Covariate Selection

We selected covariates based on a literature review of the individual-level determinants of VAA use (Gemenis & Rosema, 2014; Marschall, 2014) and turnout (Smets & van Ham, 2013) and included all factors that have a consistent relation with both concepts: age, sex, education, student status, political interest, party identification, and political knowledge. In addition, we control for whether the respondent lives in an urban or rural context; canton dummies; turnout in the previous (2003) federal election; and the rate of participation in past referendum votes.[4]

We use the standard measures included in Selects 2007 for all covariates except for political interest and political knowledge (see Appendix B) since the latter are likely to be affected by post-treatment bias (Rosenbaum, 1984). As most election studies, Selects 2007 is conducted after the election. By implication, covariates may be affected by pre-election stimuli, such as VAAs. This is not a problem for variables that refer to pre-treatment events (such as past turnout) or for variables that are relatively stable in character (such as education). By contrast, political interest and political knowledge could be directly affected by VAA use.[5]

Take political interest first. Two of the mechanisms that may link VAA usage to increased turnout run via increased political interest (see above). To avoid under-estimating the effect of VAA usage, we opt for a proxy measure of political interest: Seven dummy variables indicating a respondent's political activities over the previous five years (signing citizen's initiatives, collecting signatures for citizens' initiatives, attending political conventions, donating to parties, activity in a party, activity in a political action committee, and taking part in protests). Political interest constitutes a prerequisite to engage in either of these activities, and the proxy measures have moderate-to-high correlations with the standard political interest measure.[6] At the same time, since the survey asks about activities in the past five years, it is unlikely that they are affected by usage of *smartvote* before the 2007 election. This measurement strategy is only possible in the 2007 election study because both the 2011 and 2015 surveys relegated the items on political activities to an additional, non-random survey component that was conducted by post and had a

low response rate.

Post-treatment bias also constitutes a problem for Selects' standard political knowledge scale. Several of the constitutive items have a more or less direct relation to the kind of information disclosed by *smartvote*. For example, one of the items asks respondents about the size of their districts. Given that *smartvote* provides voting advice on a district-level basis and highlights the number of candidates a respondent can vote for (which varies by district), VAA users can be expected to have an edge. To avoid post-treatment bias, we include only those two (out of five) knowledge items that are sufficiently orthogonal to the kind of information disclosed by the VAA: an item asking about the name of Switzerland's president and another asking about the signature requirement for citizens' initiatives. Below we show that while our effect estimate becomes smaller if the original political interest scale and the full range of political knowledge items are included in the specification (as should be expected in the presence of post-treatment bias), the substantive conclusion remains unchanged.

5.2 Matching Estimator

We use entropy balancing (Hainmueller, 2012) to balance the treatment (VAA users) and control (non-users) groups. Entropy balancing directly incorporates covariate balance into a weight function, taking into account the first, second, and possibly even higher moments of the covariate distributions. The resulting weights can then be passed to almost any standard estimator to estimate the average treatment effect on the treated (ATT). Entropy balancing has two important advantages over alternative matching procedures (Hainmueller, 2012, 26). First, it typically achieves higher covariate balance compared to popular alternatives, such as propensity score matching, because it directly balances the sample moments. Second, entropy matching is favorable in terms of efficiency because contrary to many other matching techniques, no units are discarded in the matching process. Units that are very different from treated units may receive a very low but never a zero weight. Below we also report tests using alternative matching estimators. The results remain similar.

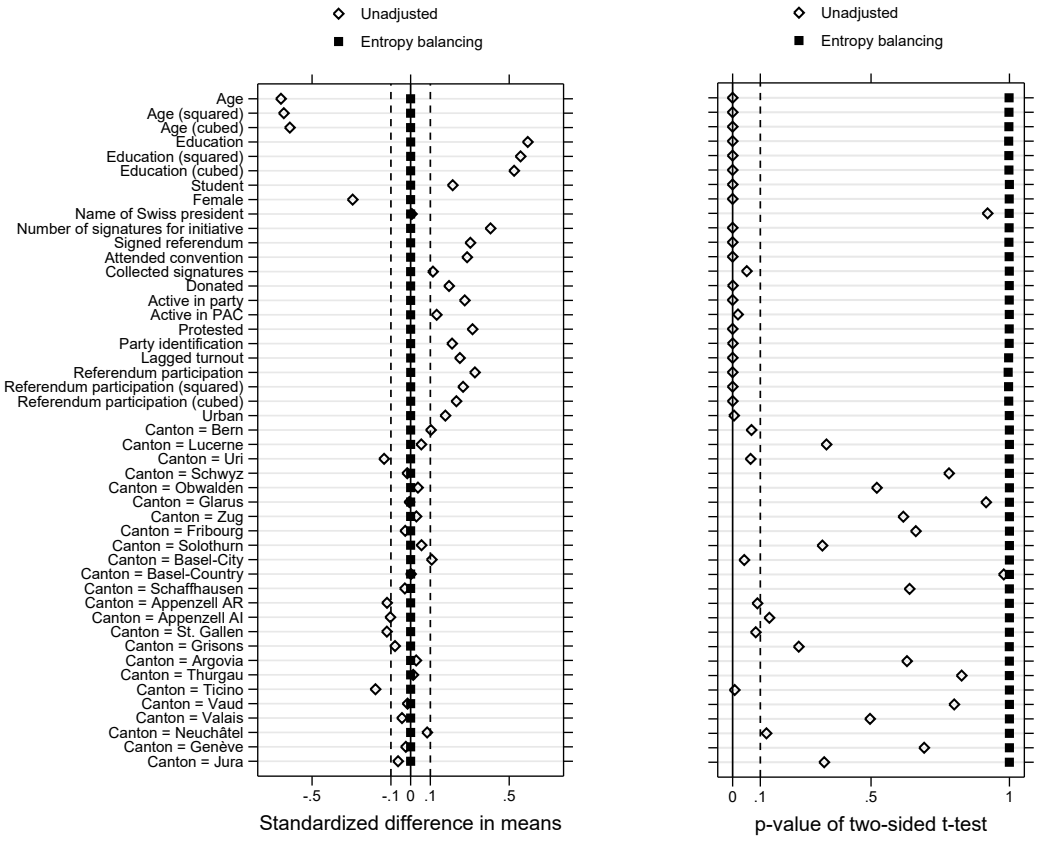
To maximize covariate balance, we balance the first, second, and third moments of the covariate distributions (i.e., the means as well as variances and skewness). Thus, in addition to the first-order terms we balance on the squared and cubed terms of our metric variables (age, education, and participation in past referendum votes), which adjusts for their second and third moments, respectively. For binary variables, it is sufficient to balance on the first-order terms as this automatically adjusts higher moments (Hainmueller & Xu, 2013).

6 Results

The quality of matching inferences hinges strongly on the quality of the matched samples (Stuart, 2010) and, as noted above, one of the distinct strengths of matching is that covariate balance can be straightforwardly assessed. Figure 3 shows the covariate balance before and after entropy adjustment, based on two standard balance statistics: the standardized difference in means (Austin, 2011, 411–412) and two-sided difference in means tests (Caliendo & Kopeining, 2008, 48–49). Unsurprisingly, there are considerable imbalances before the adjustment, as indicated by the hollow diamonds in Figure 3. In particular, *smartvote* users turn out to be disproportionately young, educated, politically knowledgeable, and interested in politics compared to non-users. As can be seen from the black squares, entropy adjustment effectively removes these differences. In other words, after entropy adjustment VAA use is orthogonal to all covariates.

To estimate the effect of VAA use on turnout, we run a logistic regression of turnout in the 2007 federal election on *smartvote* usage while weighting by the entropy balancing weights. Table 1 gives the results. According to our model, the odds of a *smartvote* user to vote are 1.8 times larger than those of a person who did not use the VAA. The effect is statistically significant at the 1% level. Since odds ratios are difficult to interpret, we calculate the implied average change in the probability of voting. According to the results, *smartvote* users were on average around 7.7 percentage points more likely to vote.

Figure 3: Covariate balance



Note: The left panel gives the covariate-by-covariate standardized difference in means between the treatment and the control group before and after entropy adjustment. A zero standardized difference in means indicates identical means; negative/positive scores indicate a lower/higher mean in the treatment group. A standardized difference of less than | .1 | means acceptable balance (Austin, 2011, 412). The right panel gives the p-value of covariate-by-covariate two-sided difference in means tests (t-tests) before and after entropy adjustment.

Table 1: VAA mobilization effect

		ATT	
	N	Odds ratio	Δ probability of voting (pp)
Entropy balancing	3578	1.816 [1.249, 2.639]	7.691 [3.554, 11.829]

Note: The table shows the average treatment effect on the treated (ATT) of *smartvote* usage on turnout in the 2007 federal election, based on a logistic regression weighted by entropy balancing weights and with robust standard errors. The ATT is given in both odds ratio and the change in the probability of voting (in percentage points). 95% confidence intervals are given in square brackets.

7 Sensitivity Analysis

7.1 Placebo Test

Matching shares with regression adjustment the same pivotal causal identification assumption: all confounders need to be measured and included in the model. We conduct a placebo test to strengthen confidence in selection on observables. In the present case an obvious placebo test involves the estimation of the treatment effect on past turnout. Turnout is highly correlated over time and logically, VAA usage should not have an effect on turnout in an earlier election; if we nevertheless find one, this would indicate that the estimated effect is in reality due to unmeasured confounders that predict both VAA use and citizens' general propensity to vote.

In implementing this placebo test we need to take into account that there were other versions of *smartvote* before the 2007 version we are looking at. VAA usage is likely to be correlated over time. As a result we cannot expect that VAA usage in 2007 has zero effect on turnout in the 2003 federal election, where another version of *smartvote* was available. Fortunately for us, however, 'regional' *smartvote* versions for cantonal elections were still relatively uncommon before 2007 (they have become much more prevalent in recent years). Therefore, we use as placebo outcome past turnout in those 16 cantonal elections between 2003 and 2007 where no VAA was available.[7] Turnout in the previous cantonal election and the 2007 federal election are correlated with 0.78, rendering it a good candidate for a placebo test.

Table 2 shows that VAA users were 8 percentage points more likely to vote in the previous cantonal election. The correlation is marginally statistically significant (p=0.07). This should be expected, given the strong correlations of VAA use with several stable drivers of individual-level turnout, such as education. Reassuringly, once we introduce our set of controls and employ entropy adjustment, the placebo effect disappears and becomes statistically insignificant (p>0.5). This improves our confidence that the effect estimate is not due to hidden bias.

It is worth noting that the placebo test also improves our confidence that the effect

Table 2: Placebo test

	N	Odds ratio	Δ probability of voting (pp)
Raw (bivariate)	1566	1.497	8.039
		[0.965, 2.32]	[-0.001, 16.083]
Entropy adjusted	1353	0.846	-2.839
		[0.508, 1.408]	[-11.725, 6.047]

Note: The table shows correlations of VAA usage with a placebo outcome (turnout in the previous cantonal election), based on logistic regressions with robust standard errors. 95% confidence intervals are in square brackets.

estimate is not owed to social desirability bias. In the present context, social desirability can be thought of as an unobserved variable that makes some respondents overreport voting. However, social desirability is likely to not only affect self-reported turnout in the current election, but also in previous elections. Therefore, if our effect estimate were due to VAA users having a higher propensity to lie about participation, we should see a placebo effect after adjustment—which we do not.[8]

7.2 Specification Checks

To further probe selection on observables, we add a number of additional control variables to the specification, including the standard political interest scale, the full range of political knowledge items, household income, news consumption, and the frequency of political discussions. The effect decreases by a third but remains statistically significant if covariates such as political interest or the frequency of political discussions are added. A plausible explanation is that these variables induce post-treatment bias. If only variables are added that are unproblematic in terms of post-treatment bias, such as household income, the effect size remains unchanged (see Appendix C).

7.3 Alternative Estimators

We also re-estimate our main model using a number alternative pre-processing techniques: nearest neighbor matching with and without exact matching on crucial covariates, coarsened exact matching, and three re-weighting methods similar to entropy adjustment (propensity score weighting, kernel matching, and local linear matching). We

consistently find a statistically significant positive effect. Nearest neighbor matching and coarsened exact matching suggest a significantly larger VAA mobilization effect (up to 12 percentage points), but this may be because they do not achieve a comparable level of covariate balance. Effect sizes for the alternative re-weighting methods are similar to entropy adjustment (see Appendix D for the results).

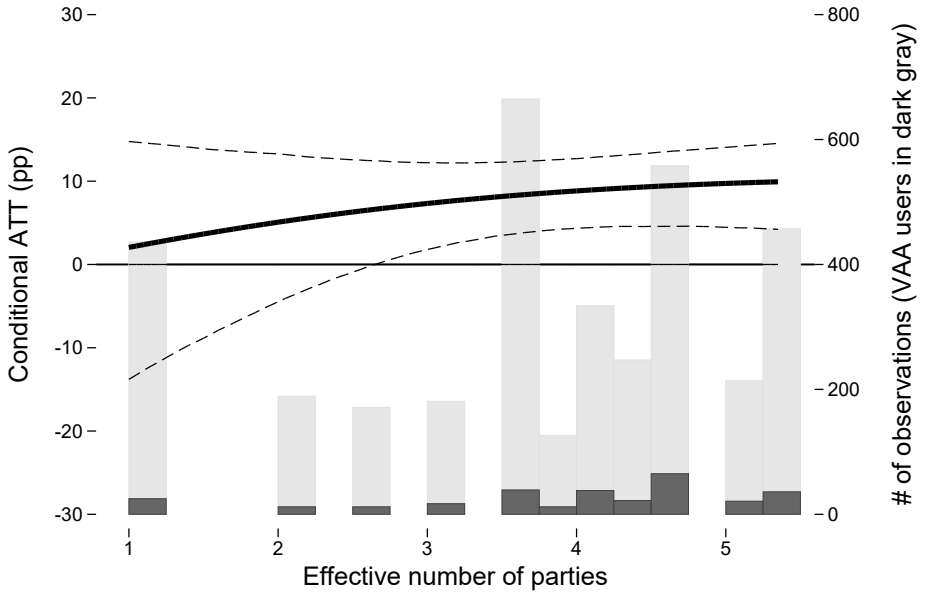
7.4 Plausibility Test

To check whether some of the theoretical expectations about the VAA mobilization effect can be borne out empirically, we exploit contextual variation in terms of the complexity of the voting decision. Based on our theory we would expect that VAAs have a higher impact in more complex electoral systems, given that the potential information gain that can be provided by a VAA is higher. Hence, the treatment effect should vary with the complexity of the voting decision.

In Switzerland, there is significant variation between the districts in national elections (i.e., the cantons) in terms of electoral complexity. To a large extent this variation is due to differences in terms of electoral competition. Switzerland has an open-list PR system and in some, especially the larger cantons, there are significantly more parties and candidates compared to others, implying a larger menu and thus a more complex decision for the citizen. Accordingly, we proxy for electoral complexity with a measure of electoral competition: the effective number of parties that were elected into the lower chamber in a canton in the 2007 election (using the Laakso and Taagepera 1979 index). As the effective number of parties is constant within cantons, we re-estimate the entropy weights while dropping the canton dummies from the specification. We then feed the new entropy weights to a logistic regression of turnout on *smartvote* usage, our proxy for electoral complexity, and an interaction term to capture the conditional relationship.

Since interaction effects cannot be directly interpreted in non-linear models, Figure 4 plots the effect of VAA use on turnout across all observed values of the effective number of parties.[9] In line with expectations, the point estimates suggest that *smartvote* usage increases turnout by a mere two to five percentage points if there are two or fewer effective

Figure 4: ATT conditional on electoral complexity



parties, but by seven to ten points if there are three or more effective parties. That said, the confidence intervals are wide, especially when the number of effective parties is low. Further analyses reported in Appendix E suggest that the second differences between the VAA effects are not statistically significant at conventional levels when the effective number of parties shifts from one (or two) to five (or four). However, it is worth noting that the lack of statistical significance could plausibly be owed to the low number of control and, in particular, treated observations at the lower end of the effective number of parties measure (as can be seen from the underlying stacked histogram). Overall, even if the second differences are not statistically significant, the conditional treatment effect behaves in a theoretically consistent way, thus increasing the plausibility of the information-based mechanisms we posit and therefore also the validity of the estimated effect.

7.5 Replication with 2011 and 2015 Data

As a final robustness check we re-estimate our model using data from the 2011 and 2015 federal elections. The specification remains the same, except that we now use the standard political interest scale instead of the 7 political activities proxies, which were not included

Table 3: VAA mobilization effect in the 2011 and 2015 federal elections

		ATT	
	N	Odds ratio	Δ probability of voting (pp)
2011 federal election	3657	1.945	7.986
		[1.359, 2.784]	[4.297, 11.674]
2015 federal election	2931	1.894	8.945
		[1.37, 2.618]	[4.757, 13.133]

Note: The table shows the average treatment effect on the treated (ATT) of *smartvote* usage on turnout in the 2011 and 2015 federal elections, based on a logistic regression weighted by entropy balancing weights and with robust standard errors. 95% confidence intervals in square brackets.

in the 2011 and 2015 election studies. Covariate balance after entropy adjustment remains excellent (see Appendix F). Table 3 shows that the VAA mobilization effects replicates for the 2011 and 2015 elections, with by and large similar effect sizes (though there is likely to be some downward bias due to the post-treatment measure of political interest).

8 Cost-Effectiveness

Having established the robustness of our effect estimate, we now turn to a closer investigation of the mobilization potential of VAAs as compared to other short-term voter mobilization techniques, such as door-to-door canvassing or mailing campaigns. Decades of research on get out the vote (GOTV) tactics, focused mostly on the American context, suggest two main conclusions (see Green & Gerber, 2008). First, the effects of many commonly deployed GOTV tactics are small and/or not statistically distinguishable from zero. Examples include leafleting, e-mail notifiers, and radio and TV ads. Second, even the most effective tactics are quite expensive. According to a review of a large number of studies (Green & Gerber, 2008, 139), an additional vote costs between 38 and 90 US dollars in telephone campaigns, depending on the coaching and motivation of the callers. For door-to-door canvassing, the corresponding figure is 29 US dollars, and for the most cost-effective tactic, election day festivals, it is 28 US dollars.

Similarly to other traditional GOTV tactics such as leafleting, VAAs can raise turnout by providing voters with political information about the upcoming election and reminding

Table 4: Effectiveness of *smartvote* in getting out the vote

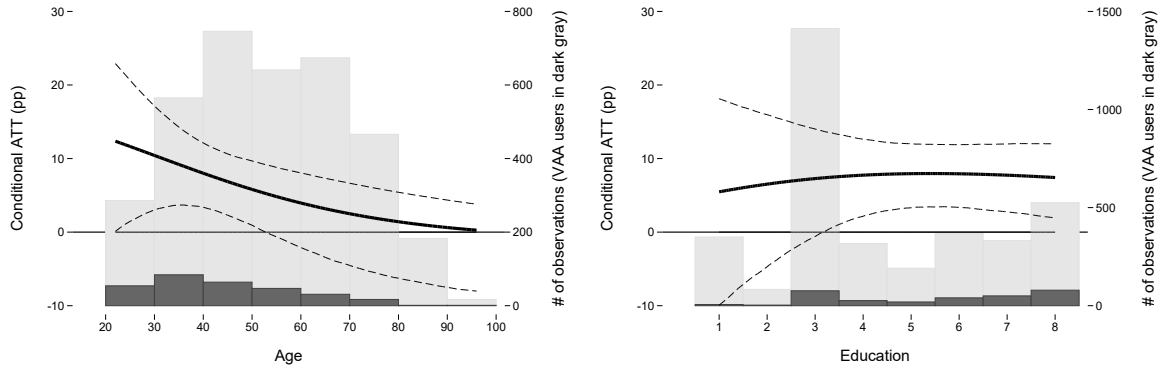
Δ number of voters	57660	[9879, 101258]
Δ total tally (pp)	1.173	[0.201, 2.06]
Cost per add. vote (CHF)	8.744	[3.814, 29.288]
Cost per add. vote (Big Macs)	1.388	[0.605, 4.649]

Note: All estimates are based on a variant of the main model that corrects for household and cantonal oversampling as well as turnout and party choice bias. 95% confidence intervals are given in square brackets.

users of the value of turning out. However, VAAs go beyond most other GOTV efforts in that they provide nonpartisan information about ideological congruence with parties or candidates that is tailored to the individual voter. Another advantage of VAAs is that there are significant economies of scale when VAAs attract large numbers of users. Setting up a well-designed VAA is costly, but the costs are more or less independent of the number of voters who end up using it. Given that *smartvote* was able to attract several hundred thousand voters in 2007, we find that it fares well in comparison with more 'traditional' GOTV tactics.

To estimate *smartvote*'s effectiveness in getting out the vote, we first re-estimated the main model (see Table 1) while accounting for Selects' in-built survey weight to get a nationally representative estimate. We find that the nationally representative individual-level effect is significantly higher while estimation uncertainty also increases (13.578 [3.452, 23.704] percentage points, see Appendix G). The increase in effect size is driven by the overrepresentation of voters in the unweighted sample and the oversampling of small cantons, where electoral complexity is lower and the VAA turnout effect thus smaller (see above).[10] Extrapolating to the electorate at large (see Table 4), we find that *smartvote* added an estimated 58,000 votes to the total tally in the 2007 federal election and was responsible for an estimated 1.2 percentage points of the total tally (48.3%). These figures are significantly smaller if compared to an earlier, less rigorous study (Fivaz & Nadig, 2010); and the confidence intervals are relatively wide. Still, even the lower bound suggests that *smartvote* was responsible for 10,000 new voters, or 0.2 points of the total tally. Furthermore, *smartvote* achieved this at a very competitive price. According to information provided by the creators of *smartvote*, the design, maintenance, and

Figure 5: Interactions with age and education



advertising of *smartvote* cost about 400,000 Swiss Francs (CHF). This suggests a point estimate of about nine Swiss francs per additional vote (approximately 7.5 US dollars at the time). The comparison is even more impressive if one considers how little one can buy with nine Swiss francs in Switzerland, one of the most expensive countries in the world. Using the Big Mac index as a metaphor (The Economist, 2015), our results suggest that *smartvote* mobilized an additional voter for the cost of around 1.4 Big Macs. Again, the confidence intervals are wide, but even the upper bound (31 CHF or 27 USD) compares well to other 'traditional' GOTV tactics.

To further prompt the VAA mobilization potential, we finally look at the effect of *smartvote* usage in 2007 conditional on education and age. Again, we get promising results (see Figure 5 and Appendix H). On the one hand, we find that *smartvote* mobilized the well and the less well educated about equally (the effect is not statistically significant at very low levels of education, but there are few observations).[11] On the other hand, we find that the effect is stronger for young and middle-aged voters.[12] This may be because younger, less experienced voters have a bigger demand for political information and weaker ex-ante preferences, and/or because younger citizens more readily adopt new technology. Given that the low turnout problem is accentuated among younger voters, and that voting is habit-forming (Dinas, 2012), *smartvote* and VAAs more generally appear to present a unique opportunity for those interested in boosting turnout.

9 Conclusions

Randomized experiments represent the gold standard for causal inference, but experiments are not always feasible. Under such circumstances, researchers have to think about second-best approaches. In this paper, we argued that control group resentment and treatment implementation problems render experiments difficult, if not impossible, to implement where VAAs are highly popular—and we are most interested in VAA mobilization effects. We argued that an observational approach utilizing survey data in combination with matching represents the best available method for the study of VAA turnout effects in contexts where VAAs are popular.

Applying our suggestion empirically, we found a substantively and statistically significant VAA turnout effect in the context of the 2007 federal election in Switzerland. Our causal estimate suggests that an additional vote generated by *smartvote* cost around nine Swiss francs (7.5 US dollars or 1.4 Big Macs), and that *smartvote* was overall responsible for an estimated 1.2 percentage points of the total turnout. Promising as well, the mobilization effect seems strongest among younger voters.

Of course, matching is no panacea. Most importantly, while matching tends to decrease imbalances between treated and control groups and allows to straightforwardly test for the degree of covariate balance that was achieved, just like regression it assumes that all relevant confounders are measured and included in the model. Therefore, we suggested to complement matching estimates with an extensive sensitivity analysis. Strengthening confidence in selection on observables, we found that our estimate passed a placebo test using turnout in past electoral contests as fake outcome, as well as a large number of other robustness checks. Nevertheless, caution remains justified.

There are several other notable limitations. First, our use of survey data could threaten internal validity due to social desirability effects. However, as vote overreporting is likely to affect self-reported turnout in current and past elections to a similar extent, the placebo test helps mitigate concerns about social desirability. Moreover, existing evidence suggests that vote overreporting is not a big problem in Swiss postelection surveys (Sciarini & Goldberg, 2016). Second, and perhaps more importantly, our reliance on sur-

vey data could affect the generalizability of our inference in the sense that hard-to-reach individuals in terms of survey response may be less prone to use VAAs in the first place. For this reason, our cost-effectiveness calculations could be too generous.

Third, we focused on a single VAA in a single country. While our results suggest that *smartvote* usage had similar effects in the 2011 and 2015 federal elections, future research should cross-validate our result in other electoral contexts. Extant research, especially the experiments conducted in countries where VAAs are less popular, increases our confidence that the VAA turnout effect is not confined to Switzerland. However, there are good reasons to expect that the effect size and, by extension, the cost-effectiveness of VAAs as a mobilization strategy vary depending on the context. For example, given its low electoral turnout (around 50% in recent years), the potential for VAAs to increase turnout could well be higher in Switzerland compared to countries such as Germany or the Netherlands (>70%), let alone Belgium (around 90%). Given the proportional (open-list) electoral system, Swiss elections are also relatively complex, especially in large districts, and above we provided tentative evidence that VAA usage may be higher where voters face a more complex decision. On the other hand, with its eight million inhabitants Switzerland is also a relatively small country and there could be improved economies of scale in larger countries. We also focused on a national election and it could be that VAAs have a larger effect in low salience elections, where voters have a higher need for information. Future work should address in more detail whether the VAA mobilization effect is context-dependent.

Finally, as most existing studies on VAA effects, this study has treated causality as a black box. Do VAAs increase turnout by providing additional information and stimulating political interest; by fostering social norms of participation; by stimulating social interaction; or is it all of the above? Our findings regarding the interaction with electoral complexity could suggest that political information may play a role, but our study does not provide any evidence on the other potential mechanisms. In addition to strengthening theory, a closer examination of the intervening mechanisms is likely to reveal useful insights as to the kind of VAA design most likely to increase participation.

Future research should start hypothesizing and testing (e.g. in laboratory experiments) what constitutes the optimal VAA design in terms of boosting voter turnout. Is it the VAA that stimulates political discussions, or the one that increases the sense of political efficacy? What kind of VAA works best for different types of voters?

Although more work needs to be done, we believe that our study has several important implications. First, this study points to the value of VAAs for those interested in boosting turnout rates. Based on our results, VAAs are not only an effective tool to increase turnout, but can achieve this goal at a very competitive price. Of course, VAAs do not address the root causes behind declining turnout rates and cannot thus be expected to offer a panacea to voter apathy. For example, citizens with no interest in politics are unlikely to use VAAs in the first place. Furthermore, VAAs can only constitute a cost-effective tactic if they reach substantial numbers of voters and, as mentioned, they may also be better suited for more complex PR elections (or party primaries), as opposed to first-past-the-post elections. Within these constraints, however, VAAs appear to offer a valuable instrument to re-engage (especially young) voters.

Second, extant research on GOTV tactics suggests that personal mobilization tactics, such as door-to-door canvassing, tend to be the most effective. This has led many to conclude that it is the social interaction between canvasser and the targeted voter that makes the difference (e.g. Bedolla & Michelson, 2012; Green & Gerber, 2008). While we do not deny the importance of social contact, our results suggest that a 'GOTV method' can also be successful in raising turnout if it does not involve human interaction. Most likely, this is because VAAs go way beyond alternatives, such as leafleting, in terms of the wealth of political information they provide, information that is notably tailored to the individual voter; but also in terms of prompting users to actively engage (by themselves or in interactions with others) with political questions including their ideological standing.

Third, we believe there are some more general ramifications regarding the maintenance and design of VAAs. If VAAs affect turnout, as this study suggests, then it is likely that they also affect vote choice (Mahéo, 2016). In turn, this implies high ethical responsibilities for the designers of VAAs. Methodological choices can have large effects

on the voting advice that is produced, with existing work highlighting the role of statement selection and formulation (Gemenis, 2013; Walgrave, Nuytemans, & Pepermans, 2009), how party or candidate positions on political issues are estimated (Gemenis & van Ham, 2014; Gemenis, 2015), and the methods used to match voters to parties or candidates (Germann, Mendez, Wheatley, & Serdült, 2015; Germann & Mendez, 2016; Louwerse & Rosema, 2014; Mendez, 2012). VAA designers should take note, adhere to the highest methodological standards, and clearly communicate the limitations of their tools to the public.

Finally, there is an evident threat of electoral manipulation. To the best of our knowledge, VAAs have to this date maintained a strict policy of nonpartisanism. Yet increased awareness of the potency of VAAs may trigger partisan interest in setting up biased VAAs, or even in hacking established VAAs and manipulating their voting recommendations. Therefore, questions related to security and the societal and institutional control over the development of such tools, thus far largely absent from political debates even in those countries where VAAs are most popular, merit increased attention in future policy debates.

Acknowledgements

We thank Jens Olav Dahlgaard, Jonas Israel, Jan Fivaz, Silke Goubin, Marc Hooghe, Georg Lutz, Fernando Mendez, Martin Rosema, Uwe Serdült, Jonathan Wheatley, Yiqing Xu, the journal editor, as well as the anonymous reviewers for valuable comments at various stages of this project. Earlier versions of this paper were presented at a 2018 seminar at KU Leuven, the 2015 EPSA Annual Conference, the 2014 ECPR General Conference, the 2014 ECPR Graduate Student Conference, and the 2014 Annual Convention of the Swiss Political Science Association.

Notes

[1]A second problem is that the left-right self-placement (which is measured post-treatment) might be affected by post-treatment bias: VAA users might have placed themselves differently in the absence of VAA usage.

[2]Formal sensitivity analysis constitutes another useful strategy to probe the plausibility of selection on observables (see e.g. Rosenbaum, 2002).

[3]*Smartvote* first appeared before the 2003 election, but the 2003 election study does not include an item on VAA usage.

[4]Turnout in past elections and referendums constitutes a strong proxy for almost any concept linked to turnout and VAA use. Controlling for past turnout also helps to compensate for possible deficiencies resulting from the necessary adjustments to the measurement of political interest and political knowledge (see below).

[5]We consider VAA use unlikely to affect party identification, at least in the short run.

[6]Polychoric correlations range from .14 (signature collection) to .49 (active party member).

[7]Argovia, the two Appenzells, Basel-City, Basel-Country, Glarus, Grisons, Jura, Neuchâtel, Obwalden, Schaffhausen, Schwyz, Solothurn, Uri, Valais, and Zug.

[8]The same argument holds if self-reported VAA use were affected by social desirability.

[9]Refer to Appendix E for the regression output and covariate balance statistics.

[10]It should be noted that the model does not include 18–21 year olds due to the accounting for turnout in the 2003 federal election.

[11]None of the second differences of the VAA mobilization effect at different levels of education even approaches statistical significance (see Appendix H).

[12]Several of the second differences are statistically significant, including the differences between voters

aged 45 and voters aged 60 above and (at the 10% level) between voters aged 30 and voters aged 75 and above (see Appendix H).

Funding

Both authors gratefully acknowledge financial support through the e-Democracy project (2012-000413) funded by the Swiss cantons of Argovia (main contributor), Basel-City, Geneva, Grisons, and Schaffhausen, as well as by the Swiss Federal Chancellery. Germann in addition gratefully acknowledges funding through the Swiss National Science Foundation (grant no. 162220).

References

Ansolabehere, S., & Hersh, E. (2012). Validation: What big data reveal about survey misreporting and the real electorate. *Political Analysis*, *20*(4), 437–459. doi: 10.1093/pan/mps023

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, *46*(3), 399-424. doi: 10.1080/00273171.2011.568786

Bedolla, L. G., & Michelson, M. R. (2012). *Mobilizing inclusion: Transforming the electorate through get-out-the-vote campaigns.* New Haven, CT: Yale University Press.

Caliendo, M., & Kopeining, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*(1), 31-72. doi: 10.1111/j.1467-6419.2007.00527.x

Dehejia, R. H., & Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, *94*(448), 1053–1062. doi: 10.2307/2669919

Dinas, E. (2012). The formation of voting habits. *Journal of Elections, Public Opinion and Parties*, *22*(3–4), 431-456. doi: 10.1080/17457289.2012.718280

Dinas, E., Trechsel, A. H., & Vassil, K. (2014). A look into the mirror: Preferences, representation and electoral participation. *Electoral Studies*, *36*, 290-297. doi: 10.1016/j.electstud.2014.04.011

Enyedi, Z. (2016). The influence of voting advice applications on preferences, loyalties and turnout: An experimental study. *Political Studies*, *64*(4), 1000-1015. doi: 10.1111/1467-9248.12213

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. doi: 10.2202/1944-2866.1025

Freedman, D. A. (2006). Statistical models for causation: What inferential leverage do they provide? *Evaluation Review*, *30*(6), 961-713. doi: 10.1177/0193841X06293771

Garzia, D., de Angelis, A., & Pianzola, J. (2014). The impact of voting advice applications on electoral participation. In D. Garzia & S. Marschall (Eds.), *Matching voters to parties and candidates* (p. 105-114). Colchester: ECPR Press.

Garzia, D., Trechsel, A. H., & De Angelis, A. (2017). Voting advice applications and electoral participation: A multi-method study. *Political Communication*, *34*(3), 424–443. doi: 10.1080/10584609.2016.1267053

Gemenis, K. (2013). Estimating parties' positions through voting advice applications: Some methodological considerations. *Acta Politica*, *48*(3), 268-295. doi: 10.1057/ap.2012.36

Gemenis, K. (2015). An iterative expert survey approach for estimating parties' policy positions. *Quality & Quantity*, *49*(6), 2291–2306. doi: 10.1007/s11135-014-0109-5

Gemenis, K., & Rosema, M. (2014). Voting advice applications and electoral turnout. *Electoral Studies*, *36*, 281-289. doi: 10.1080/10584609.2016.1267053

Gemenis, K., & van Ham, C. (2014). Comparing methods for estimating parties' positions in voting advice applications. In D. Garzia & S. Marschall (Eds.), *Matching voters to parties and candidates* (p. 33-47). Colchester: ECPR Press.

Gerber, A. S., Green, D. P., & Larimer, C. W. (2008). Social pressure and voter turnout:

Evidence from a large-scale field experiment. *American Political Science Review*, *102*(1), 33–48. doi: 10.1017/S000305540808009X

Germann, M., & Mendez, F. (2016). Dynamic scale validation reloaded: Assessing the psychometric properties of latent measures of ideology in VAA spatial maps. *Quality & Quantity*, *50*(3), 981-1007. doi: 10.1007/s11135-015-0186-0

Germann, M., Mendez, F., Wheatley, J., & Serdült, U. (2015). Spatial maps in voting advice applications: The case for dynamic scale validation. *Acta Politica*, *50*(2), 214-238. doi: 10.1057/ap.2014.3

Germann, M., & Serdült, U. (2017). Internet voting and turnout: Evidence from Switzerland. *Electoral Studies*, *47*, 1–12. doi: 10.1016/j.electstud.2017.03.001

Gil de Zúñiga, H., Jung, N., & Valenzuela, S. (2012). Social media use for news and individuals' social capital, civic engagement and political participation. *Journal of Computer-Mediated Communication*, *17*(3), 319–336. doi: 10.1111/j.1083-6101.2012.01574.x

Green, D. P., & Gerber, A. S. (2008). *Get out the vote: How to increase voter turnout* (Second ed.). Washington, DC: Brookings Institution.

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, *20*(1), 25-46. doi: 10.1093/pan/mpr025

Hainmueller, J., & Xu, Y. (2013). ebalance: A Stata package for entropy balancing. *Journal of Statistical Software*, *54*(7), 1-18. doi: 10.18637/jss.v054.i07

Ho, D. E., Imai, K., King, G., & Stuart, E. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, *15*(3), 199-236. doi: 10.1093/pan/mpl013

Krouwel, A., Vitiello, T., & Wall, M. (2014). Voting advice applications as campaign actors: Mapping VAA's interactions with parties, media and voters. In D. Garzia & S. Marschall (Eds.), *Matching voters with parties and candidates* (p. 67-78). Colchester: ECPR Press.

Kruikemeier, S., van Noort, G., Vliegenthart, R., & de Vreese, C. (2014). Unravel-

ing the effects of active and passive forms of political Internet use: Does it affect citizens' political involvement? *New Media & Society*, *16*(6), 903-920. doi: 10.1177/1461444813495163

Laakso, M., & Taagepera, R. (1979). "Effective" number of parties: A measure with application to West Europe. *Comparative Political Studies*, *12*(1), 3-27. doi: 10 .1177/001041407901200101

Louwerse, T., & Rosema, M. (2014). The design effects of voting advice applications: Comparing methods of calculating results. *Acta Politica*, *49*(3), 286-312. doi: 10.1057/ap.2013.30

Mahéo, V.-A. (2016). The impact of voting advice applications on electoral preferences: A field experiment in the 2014 Quebec election. *Policy & Internet*, *8*(4), 391–411. doi: 10.1002/poi3.138

Mahéo, V.-A. (2017). Information campaigns and (under)privileged citizens: An experiment on the differential effects of a voting advice application. *Political Communication*, *34*(4), 511–529. doi: 10.1080/10584609.2017.1282560

Marschall, S. (2014). Profiling users. In D. Garzia & S. Marschall (Eds.), *Matching voters to parties and candidates* (p. 93-104). Colchester: ECPR Press.

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 D. Farrell & R. Schmitt-Beck (Eds.), *Non-party actors in electoral politics* (p. 259-278). Baden-Baden: Nomos.

Marschall, S., & Schultze, M. (2012). Voting advice applications and their effect on voter turnout: The case of the German Wahl-O-Mat. *International Journal of Electronic Governance*, *5*(3–4), 349-366. doi: 10.1504/IJEG.2012.051314

Mendez, F. (2012). Matching voters with political parties and candidates: An empirical test of four algorithms. *International Journal of Electronic Governance*, *5*(3–4), 264-278. doi: 10.1504/IJEG.2012.051316

Murray, M. P. (2006). Avoiding invalid instruments and coping with weak instruments. *Journal of Economic Perspectives*, *20*(4), 111-132. doi: 10.1257/jep.20.4.111

Pianzola, J., Trechsel, A. H., Gchwerdt, G., Vassil, K., & Alvarez, R. M. (2012). *The effect of voting advice applications (VAAs) on political preferences: Evidence from a randomized field experiment.* Paper presented at the 2012 APSA Annual Meeting, New Orleans, LA.

Rosenbaum, P. R. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society Series A*, *147*(5), 656-666. doi: 10.2307/2981697

Rosenbaum, P. R. (2002). *Observational studies* (Second ed.). New York, NY: Springer.

Sciarini, P., & Goldberg, A. (2016). Turnout bias in postelection surveys: Political involvement, survey participation, and vote overreporting. *Journal of Survey Statistics and Methodology*, *4*(1), 110–137. doi: 10.1093/jssam/smv039

Sekhon, S. S. (2009). Opiates for the matches: Matching methods for causal inference. *Annual Review of Political Science*, *12*, 487-508. doi: 10.1146/annurev.polisci.11.060606.135444

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin.

Smets, K., & van Ham, C. (2013). The embarrassment of riches? A meta-analysis of individual-level research on voter turnout. *Electoral Studies*, *32*, 344-359. doi: 10.1016/j.electstud.2012.12.006

Stolzenberg, R. M., & Relles, D. A. (1990). Theory testing in a world of constrained research design: The significance of Heckman's censored sampling bias correction for nonexperimental research. *Sociological Methods & Research*, *18*(4), 395-415. doi: 10.1177/0049124190018004001

Stuart, E. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, *25*(1), 1-21. doi: 10.1214/09-STS313

The Economist. (2015). *The Big Mac index.* Retrieved from `http://www.economist.com/content/big-mac-index`.

Tolbert, C. J., & McNeal, R. S. (2003). Unraveling the effects of the Internet

on political participation? *Political Research Quarterly*, *56*(2), 175–185. doi: 10.1177/106591290305600206

Vassil, K. (2012). *Voting smarter? The impact of voting advice applications on political behavior* (Unpublished doctoral dissertation). European University Institute.

Walgrave, S., Nuytemans, M., & Pepermans, K. (2009). Voting advice applications and the effect of statement selection. *West European Politics*, *32*(6), 1161-1180. doi: 10.1080/01402380903230637

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 Politica*, *43*(1), 50-70. doi: 10.1057/palgrave.ap.5500209