Reexamining the Value of (CA)VAAs: The Effects on Political Knowledge, Interest, Voting intention, and Voter Turnout

Florian Verschuren

2086665

Master's thesis

Communication and Information Sciences/Business Communication and Digital Media

Department of Communication and Cognition

Tilburg University, Tilburg

Supervisor: Dr. C. Liebrecht

Second reader: C. Van der Lee

August 2022

Abstract

Tools that help the user to make an informed decision as to which party to vote for during political elections are being used by evermore citizens across democracies. The benefits of these voting advice applications (VAAs) on their users have been a topic of interest in a large body of research. To this end, studies have consistently reported positive results with respect to, for instance, the effects of VAAs on the user's political knowledge, their interest in politics, and turnout during elections. However, all VAA studies contain methodological problems due to the difficulty of conducting research toward the effects of VAAs in a real-life context, in which VAAs are already widely used because it is unethical to withhold a group of people from using a VAA to inform themselves about elections. Moreover, there is a shortage of quantitative data on how VAAs could be improved in the future. To these ends, the current study aims to overcome most limitations present in prior research by conducting a longitudinal study among 166 students making use of the unique real-life election context of the 2022 university council elections at Tilburg University. Specifically, this study compares the effects of using a VAA, no tool, and a newly developed VAA with an integrated chatbot function (a Conversational Agent Voting Advice Application, or CAVAA) at three different moments in time. By conducting several repeated measures ANOVAs, this study has found that voting advice applications can enhance one's factual and perceived political knowledge as well as their voting intention. Specifically, a VAA has been found to increase one's perceived political knowledge and voting intention as compared to no tool usage, whereas a CAVAA increases the users' perceived as well as factual knowledge about politics as compared to not using a tool. Moreover, this study found that mere exposure to political information can enhance one's perceived and factual political knowledge, interest in politics, voting intention, and voter turnout. Furthermore, differences in the stability of effects over time are found. Limitations, directions for future research, and implications are discussed.

Keywords: voting advice applications, CAVAA, political knowledge, political interest, chatbots, voting intention, voter turnout, perceived political knowledge, factual political knowledge

Reexamining the Value of (CA)VAAs: The Effects on Political Knowledge, Interest, Voting intention, and Voter Turnout

In a healthy democracy, most eligible voters cast a vote, have a good interest in politics, and possess a sufficient amount of political knowledge (Lijphart, 1997; Ou & Tyson, 2021). However, it appears that in most democracies, these requirements are not met. One manner in which these insufficient levels have been aimed to increase is by the creation of tailored Voting Advice Applications (VAAs). These VAAs are applications that present the user with statements about political issues, for which they have to indicate to what extent they agree or disagree. The contesting political parties have also indicated their opinions on each statement and the VAA is, therefore, able to compare the answers of the users to those of the political parties. Based on this comparison, the VAA is able to provide the users with tailored voting advice (De Graaf, 2010). In order to obtain information about which political party best matches their opinion, people all over the world have been using a VAA (Ladner & Pianzola, 2015). Many studies have indicated the potential benefits of using VAAs, such as a higher interest in politics (Krouwel et al., 2018; Ladner, 2012), an increase in political knowledge (Kamoen et al., 2015; Schultze, 2014), and an increase in voter turnout (Garzia, Trechsel & De Angelis, 2017; Gemenis & Rosema, 2014).

Nonetheless, despite these promising results, there are several important methodological problems that negatively impact the validity of the findings. These problems are due to the complexity of conducting (ecological) valid research on the effects of VAAs. It is complex, as the effects of VAAs can solely be measured by comparing the effects to a control group without a tool, in a real election context. However, this would mean that one part of the electorate has to be instructed not to use a VAA which is unethical. This leads to many researchers comparing people that did fill out a VAA to those that did not, which means that the participants are not randomly assigned, which is problematic as people of both groups are different from each other a priori (Fivaz & Nadig, 2010; Marschall & Schmidt, 2008). Moreover, there is also a need for insights into how these tools can actually be improved, as there is still room for improvement left. For instance, research by Kamoen and Holleman (2017) has shown that VAA users often experience difficulties in understanding all statements provided and most often do not make any effort to overcome these comprehension problems. Furthermore, it was found that VAA users rather make assumptions about the meaning than to actually search for the information on a search engine, which consequently could lead to invalid voting advices. Moreover, research has found that many people are likely to follow

their voting advice (Andreadis & Wall, 2014; Kleinnijenhuis et al., 2019). Hence, invalid voting advices could lead to a vote choice that does not best reflect the actual opinions of the user.

A promising way to overcome these comprehension problems without having to open up a search engine is by integrating a chatbot into the VAA (Kamoen & Liebrecht, 2022). Such a chatbot could be asked questions about definitions, the current situation with respect to the issue at hand, and the arguments in play. A VAA with an integrated chatbot function is better known as a Conversational Agent Voting Advice Application (CAVAA) (Kamoen & Liebrecht, 2022). One study on the benefits of using a CAVAA yielded promising results, as it was found that its use led to more factual political knowledge than a regular VAA (Kamoen & Liebrecht, 2022). However, this study has several important limitations that the current study aims to address such as the absence of a no tool condition and no pre-experimental measurements.

More research is needed towards the impact of VAAs on political knowledge, voter turnout, and political interest as well as insights into how VAAs could be improved as the current literature does not provide satisfying answers to any of these questions. The present research aims to overcome prior limitations with respect to VAA and CAVAA research in order to get a deeper understanding of the usefulness of both tools and consequently manners to improve the aspects discussed that lay at the basis of a healthy democracy. The present research fills several research gaps due to the application of a unique experimental design. For instance, this research is the first that allows the making of valid comparisons with respect to political interest, political knowledge, and voter turnout due to (CA)VAA usage, as it the first to conduct experimental research toward both CAVAAs and VAAs while effectively taking into account both prior differences between groups, and longitudinal effects. Moreover, this (CA)VAA experiment is one of the first that allows the researcher to randomly assign the participants to groups and create the voting advice applications themselves, and study their effects during a real election context. The latter is the 2022 university council election at Tilburg University, which is an interesting context as the participants did not have prior experiences with a (CA)VAA in these elections, as the present application is the first-ever, and because it allows verifying whether prior findings in different election contexts can be replicated. Furthermore, this election context makes it possible to specifically test the results of the (CA)VAA as the absence of already existing (CA)VAAs allows for the incorporation of a control group that does not fill out such a tool, which participants can be randomly assigned

to, without unethically withholding them from political information, as one of the voting advice applications was being spread to the public later on as well. Finally, this research is also the first that is able to obtain valid results on how VAAs can be improved by testing and comparing the effects of a CAVAA and a VAA on its users. Thus, by making use of a different, but real-life, election context and by taking the limitations and issues of other studies into account, the present research aims to answer the research question: "*What are the different effects of using either a VAA, CAVAA, or no tool at all, on the perceived and factual knowledge, the interest in politics, the voting intention and the voter turnout?*".

Literature Review

A high voter turnout is seen as a democratic ideal as it is important for democratic systems that the elected leaders represent the opinions and values of the public as a whole. More specifically this means that the more people cast a vote, the better the public opinion becomes known and the better it gets represented (Verba, 1996).

Nonetheless, besides a high voter turnout, it is perhaps equally important (if not more), that the votes are based on informed decisions (Tyson, 2016). Votes could only provide insight into the actual standpoints of the public if they are based on actual political knowledge. If one does not have sufficient political knowledge, they might vote for parties or politicians that do not actually reflect their opinion. This could become a problem if many voters cast such an uninformed vote as then many people might repent their choice and, consequently, oppose many governmental decisions, leading to potential turmoil. One's knowledge of politics can consist of several aspects such as one's knowledge of political terms and definitions (Kamoen & Liebrecht, 2022), knowledge about the current state of affairs with regard to a certain political issue (Delli Carpini & Keeter, 1996), and knowledge about the political parties and their stances (Schultze, 2014; Van de Pol, 2016).

Moreover, in order for people to actually seek political information and trust the government, they need to be interested in politics in the first place (Catterberg & Moreno, 2006; Lecheler & De Vreese, 2017; Strömback, Djerf-Pierre & Shehata, 2013). Political interest is a broad concept and consists of various matters such as interest in elections (Ladner, 2012), behavioral aspects such as political information seeking (e.g. by visiting party websites or watching debates), and whether one frequently holds political discussions (Ladner & Pianzola, 2015).

Thus, for democratic systems to work ideally it is important that everybody that is allowed to vote has sufficient political knowledge (Ou & Tyson, 2021; Tyson, 2016), enough political interest (Ou & Tyson, 2021), and that as many people cast a vote as possible (Lijphart, 1997; Rome, 2022; Verba, 1996).

However, despite the importance of these three aforementioned factors for democracies, research makes clear that people in general score rather low on each of these three factors. For instance, the political knowledge of the public has been rather low for quite some decades (Delli Carpini & Keeter, 1996; Pew research center, 2007). Furthermore, the interest of the public in politics seems to have been quite low as well over the years. British data reveal that during the period of 1991-2005 the British people were either in between being "not very" or "fairly" interested in politics (Prior, 2010). Slightly better, but still rather low, figures can be found for American citizens and the Suisse public (Prior, 2010). Moreover, also in the Netherlands, 49% of the people indicated having very low political interest during the period 2012-2016 (Schmeets, 2017). Finally, voter turnout has also been found to be continuously low across democracies. The amount of people voting has been ranging between 60-70% in the United States presidential elections from 1964-2020 (O'Neill, 2022), and with respect to the Dutch parliament elections, the voter turnout has been between 73 and 82 percent in the period of 1970-2021 (Kiesraad, 2021). Moreover, in Switzerland, more than half of those allowed to vote in the national elections remain absent from the ballots, which has been the case since 1975 (Ladner & Pianzola, 2010).

The low numbers of general political interest, political knowledge, and voter turnout have been, among other factors, one of the reasons for political scientists to develop Voting Advice Applications (VAAs). These applications have been around for a few decades and are being utilized by an increasing number of people and countries (Marschall, 2014).

The popularity of VAAs might be due to the positive effects that these tools are suggested to have according to prior research. For instance, research indicates that VAAs could be used to enhance one's interest in politics (Fivaz & Nadig, 2010; Garzia, Trechsel, & De Angelis, 2017; Krouwel et al., 2018; Ladner, 2012; Marschall & Schmidt, 2010).
In Marschall and Schmidt (2010) a sample of more than ten thousand users of the "*Wahl-O-Mat"* (a German VAA) filled out a questionnaire immediately after having used the tool. Several items about the perceived effects of the tool were being asked, such as whether one believed the VAA had motivated them to search for additional information regarding politics. It appeared that around 60% of the contestants indeed experienced a sense of motivation to

further inform themselves politically as a result of the "*Wahl-O-Mat*". Similar studies have been conducted with respect to other VAAs such as the Dutch "*StemWijzer*" (Boogers, 2006), and the Swiss "*Smartvote*" (Ladner, Felder & Fivaz, 2010), which provided similar results. The reason for the increase in political interest due to VAAs might, among others, simply be the fact that it exposes the user the political stimuli. For instance, Strömback and Shehata (2010) have found that even if one is passively consuming political information, this could still lead to a (temporal) increase in political interest. One's interest could be even more enhanced if actual attention is paid to the political stimuli. Therefore, VAAs can enhance one's political interest due to the political content that they contain.

With regard to political knowledge, a large body of VAA research has provided results indicating that there exists a positive relationship between the usage of such a tool and one's knowledge about politics (Fivaz & Nadig, 2010; Heinsohn et al., 2014; Kamoen et al., 2015; Schultze, 2014; Westle, Begemann, & Rütter, 2014). Kamoen et al. (2015) used VAA data obtained one month prior to the Dutch national elections of 2012 in which the respondents were asked whether their political knowledge had increased. One day after the election, the participants were asked to fill out a survey containing the same question. The results indicated that VAA use indeed led to more political knowledge among the participants.
Moreover, a third benefit of VAAs is that they could enhance the general voter turnout (Fivaz & Nadig, 2010; Garzia, De Angelis & Pianzola, 2014; Garzia, Trechsel & De Angelis, 2017; Gemenis & Rosema, 2014; Ladner & Pianzola, 2010; Marschall & Schultze, 2012). In the study by Garzia, De Angelis, and Pianzola (2014) eight datasets about users and nonusers of VAAs from four different countries were analyzed on a series of dependent variables such as voter turnout. Their results indicated that VAAs could indeed lead to a higher voter turnout, as in almost all the cases it was found that users of a VAA more often had cast a vote in the upcoming elections than people that did not use a voting advice application (nonusers). The reason that VAA usage has been found to increase voter turnout might be, among others, the fact that VAAs expose one to political stimuli. Exposure to political stimuli has namely been found to positively influence voter turnout (Finkel & Geer, 1998; Goldstein & Freedman, 2002).

Despite these promising results, it has to be addressed that most of the studies regarding the effects of VAAs on the discussed variables contain one or multiple important methodological limitations which might affect the validity of the results found. Moreover, there is also much unknown about how VAAs can be improved, for instance by transforming

them into a CAVAA. To these ends, the goal of the present research is to conduct VAA research towards the three aforementioned outcome variables, while taking into account the limitations of prior VAA research and investigating whether the results will alter in a more dynamic tool like a CAVAA.

The most important limitations and problems with respect to (CA)VAA research will separately be addressed in the following sections.

Absence of quantitative insights about manners to improve VAAs

The first issue with regard to VAA research is that there is limited quantitative data on how VAAs can actually be improved. This is in contrast with the amount of qualitative insights, which have been researched much more often (e.g. Baka, Figgou & Triga, 2012; Van Camp, Lefevere & Walgrave, 2014). It is important to also obtain quantitative data in order to obtain more objective and generalizable results on the effects of VAAs. Although the creation of VAAs has been an important step toward improving the quality of democratic systems all over the world, more (quantitative) insights into how these tools could further be improved are needed as these tools have been found to not lead to the desired results. In this regard, the study by Kamoen and Holleman (2017) is worth discussing.

Their research consists of two studies of which the first investigated to what extent the users of a VAA do actually understand the statements provided, and what actions they undertake if they do experience comprehension problems. During study one, data was obtained during a Dutch municipal election by asking users of a VAA to think out loud while filling out the tool. The results made clear that people often do not fully comprehend one or multiple statements in a voting advice application.

Kamoen and Holleman (2017) further specify the different natures of the comprehension problems. Accordingly, participants sometimes experienced semantic comprehension problems, in which the user did not know the literal meaning of one or several concepts in a statement, and sometimes the problems were due to a lack of knowledge about the current state of affairs with regard to a certain issue, which they refer to as a pragmatic comprehension problem. If users fail to understand the literal meaning of a word in a statement (e.g. semantic comprehension problem), it is problematic as, in the comprehension process, it is crucial that one is able to transform the literal meaning of the statement (in this

context) into a semantic representation in one's own mind (Graesser, Singer & Trabasso, 1994; Zwaan & Radvansky, 1998). If one does not understand what the words literally mean, they will not be able to form such a semantic representation and therefore will not fully understand what the statement means. On the other hand, if one does manage to form a semantic representation, one still needs to link it to their own related world knowledge in order to put the statement into context (e.g. pragmatic knowledge). It could occur that one does not know sufficiently about the context of the statement, which would mean that the semantic representation cannot be linked to one's related world knowledge, as this person does not possess it. Surprisingly, the data showed that when VAA users do experience comprehension problems, be it semantic or pragmatic, nearly all of the participants did not make any effort to look up the missing information (e.g. by opening up a search engine) and just proceeded to guess what either the correct meaning of a word or the actual status quo of an issue was. Based on their own inference they continued to select an answer in the VAA. The authors suggest that the reason behind this behavior might possibly be that VAA users tend to show satisficing behavior, which means that they provide just a sufficient amount of effort to give an answer that is somewhat plausible, just as when people fill out a survey (Krosnick, 1991).

The findings of Kamoen and Holleman (2017) are alarming as they indicate that VAA users often fill out VAAs without fully understanding certain aspects of the statements, which could potentially lead to invalid voting advices. The latter is problematic as users, especially those that have not made their voting decision yet (Andreadis & Wall, 2014; Kleinnijenhuis et al., 2019), are likely to follow their voting advice which, thus, possibly not reflects the party that best matches their actual opinions. Therefore, a lot of people might vote for parties that do not actually match their own interests best, potentially leading to an inaccurate governmental representation of the actual opinions and desires of the public.

Besides a lack of semantic and pragmatic knowledge, users of a VAA might also be in need to obtain other types of knowledge such as the stances of the political parties with regard to the statements (Schultze, 2014; Van de Pol, 2016), and the advantages and disadvantages of the possible decisions with respect to the issues presented in the statements. For instance, when one knows the stances of parties with regard to certain issues, this will help them in forming an informed picture of each party, helping them to critically assess their voting advice. Moreover, if one has a good understanding of the potential consequences of certain decisions, then this will help the person to actually weigh the impact of each decision. To

these ends, these additional types of knowledge might be important when it comes to forming an informed opinion about political matters, and subsequently, filling in well-reasoned answers in a VAA in order to receive a valid voting advice.

At this point, it is clear that VAAs are not yet perfect and could be especially improved in two aspects. The first one is the fact that people often experience comprehension problems while using the tools. The second one is related to the lack of effort of the users to actually overcome these comprehension problems. Future VAAs should thus be designed in such a way that they take into account the likelihood that the users are not very motivated and expected to show satisficing behavior (Krosnick, 1991). However, most recent VAAs often do have additional information buttons with which one could obtain information about, for instance, the stances of the political parties towards statements. Nonetheless, no research has been conducted toward whether these buttons could overcome the two main problems of VAAs. It seems likely, however, that the buttons lower the effort required to obtain extra information and that therefore the user's factual political knowledge increases. On the other hand, it could also be reasoned that merely the presence of static information buttons will not greatly increase one's motivation to actually obtain more information. It might be that a more dynamic manner of information provision could enhance one's information-seeking motivation and consequently factual political knowledge.

To this end, a recent body of VAA research has provided promising results, as several studies have aimed to overcome the existing comprehension problems with VAAs by integrating a chatbot into the tool (Aerts, 2022, Kamoen & Liebrecht, 2022; Kamoen, McCartan & Liebrecht, 2021). Chatbots have a high potential to be beneficial in VAAs as they can be programmed to provide quick answers to questions of users (which can be asked either via pre-programmed buttons or via an open text field). This, along with the fact they are integrated into the VAA, lowers the amount of effort needed to overcome comprehension problems, as VAA users would then not need to open up a separate search engine and look between loads of unnecessary information for the specific information they need. Moreover, research has indicated that when chatbots are designed in such a way that it actually feels like the users are having a conversation with a real entity, this could increase one's motivation to interact (Tsai, Liu & Chuan, 2021) which, in turn, might lead to one feeling more motivated to inform themselves via the tool.

The first research in which a voting advice application with an integrated chatbot function has been developed and tested is the study by Kamoen and Liebrecht (2022). In their

research, they named this new tool a Conversational Agent Voting Advice Application (CAVAA), which is how this tool will be addressed in the present study as well. Kamoen and Liebrecht (2022), conducted a between-subjects experiment in which Dutch participants were assigned to either fill out a regular VAA or the newly-developed CAVAA, which both mostly consisted of statements derived from the 2017 Dutch national elections. The CAVAA could be asked questions related to semantic and pragmatic information, information about the stances of the parties, and information about the advantages and disadvantages of each decision (agreeing or disagreeing). After filling out their respective tool, the participants answered several questions related to their perceived and factual political knowledge, their voting intention, and how they evaluated the tools. The results indicated that CAVAA users obtained a higher factual and perceived knowledge, and had more positive experiences with the tool they used, than the VAA users. Therefore, the researchers concluded that a CAVAA can be a successful tool, among others, for obtaining factual political knowledge without having to spend much cognitive effort. Moreover, in a second study by Kamoen and Liebrecht (2022), it was found that when the users of a CAVAA could request information via buttons, rather than merely via an open text field, the tool corresponds even better to the low elaboration mood of the users. This, as it was found that participants requested more information while using the CAVAA when there were buttons as opposed to when they could solely use an open text field. The results of both studies reported in Kamoen and Liebrecht (2022) seem promising as they indicate that the comprehension problems that arise with VAAs could be solved by integrating a chatbot into the application.

However, whether CAVAAs can actually overcome the problems that Kamoen and Holleman (2017) have addressed with regard to VAAs, remains unclear as the study by Kamoen and Liebrecht (2022) contains two important limitations. First, their study does not compare the results of both the VAA and the CAVAA with a no tool condition and neither makes use of pre-measurements, which would both have facilitated the comparisons of the results. Second, their study was conducted outside an actual election context and therefore has limited ecological validity. The present study aims to find valid experimental VAA and CAVAA results by taking into account the limitations of Kamoen and Liebrecht (2022).

Pre-tool comparison

Quite a number of VAA research merely tests whether a VAA has had an effect on the user by measuring either subjectively (e.g. Kamoen et al, 2015; Ladner, 2012) or objectively (e.g. Manavopoulos et al., 2018; Schultze, 2014) a series of dependent variables after using

the tool. Such post-tool measurements can be valuable as they allow for the comparison between multiple groups after a manipulation (e.g. VAA usage). However, this type of design does not allow for observing effects from within a person as the pre-tool scores of the individual with regard to the variables of interest are then unknown. Thus, post-tool measures fail to provide insights into how individuals have been affected by the experimental manipulation. The present study aims to obtain both within as well as between subjects insights into the effects of voting advice applications by applying both pre-and-post measurements.

No tool comparison

Another issue with regard to research on the effects of VAAs is that quite a number of studies do not compare the results with a control group that did not fill out any tool. This type of information would be valuable as it provides the possibility to gain insights into what actually are the effects of the voting advice application itself. Moreover, it appears to be that all research toward the effects of VAAs that actually did compare the results with a no tool condition did not randomly assign people to one of the groups, which caused the effects to be affected by confounding variables. This is the case, as multiple studies that rely on actual election data have made clear that users of VAAs and people who would normally not use a VAA (nonusers) are often different on a series of variables a priori. For instance, the strength of party identification and their political interest (Marschall & Schmidt, 2008), as well as, education, gender, and age (Fivaz & Nadig, 2010). Each of these factors has been found to influence the variables of interest in the present study. To illustrate, voter turnout is influenced by age, education, political interest, and degree of party identification, (Smets and Van Ham, 2013) political interest is influenced by age (Andersen et al., 2020; Gidengil et al., 2004), gender, and education (Furnham & Cheng, 2019), and factual political knowledge could alter as a result of one's gender (Frazer & Macdonald, 2003; Mondak & Anderson, 2004) and education (Grönlund & Milner, 2006). Therefore, most VAA studies fail to test for the effects of merely the tool, as they suffer from confounding variables due to a non-random assignation of the participants.

In order to actually test for the effects of a VAA on its users, experimental research with multiple measurements would be needed that randomly places a sample of people either into a condition in which they have to fill out a VAA or in a control condition without such a

tool. To this end, one could either conduct a field study or a laboratory study. The main advantage of a field study is that the results are likely to be ecological valid due to the real-life setting in which the experiment is conducted. On the other hand, a laboratory study normally does not suffer from confounding variables due to the controlled setting, which is a great advantage of this study design. However, there are several problems attached to both methodologies. With respect to a field study approach, there is a problem related to ethical concerns as it does not seem ethical to withhold people in the no tool condition to inform themselves about an election by forbidding them from using a VAA. However, if one would not forbid VAA usage to the participants in the no tool condition then there would exist an extreme risk of two-sided noncompliance as people are very likely to use such a tool in-between measurements due to the widespread availability of the readily existing VAA. Another problem is related to the lack of control for confounding variables, as, in a field study, multiple factors besides the usage of a voting advice application could influence the alternation of the studied variable. For instance, in a field study approach, it might be that merely people who are already politically interested and who would normally also use a VAA, decide to participate in the study, which would be a confounding variable due to the discussed differences between users and non-users. One could overcome these issues by conducting a laboratory study. However, the main disadvantage of this approach is that there is a limited ecological validity as the controlled circumstances might be very different from the natural environment in which the participants would normally, in this case, fill out a voting advice application. Moreover, as also people who would normally not use a voting advice application do participate, the results might be less generalizable due to the fact that the differences between users and non-users might impact the results found. To illustrate, the results would be more ecological valid if merely people who would normally use a voting advice application did participate, as they are the ones who would actually use such a tool in real life.

Thus, there are some challenges when it comes to conducting experiments toward the pure effects of VAAs. However, there is one study that managed to overcome nearly all of the aforementioned limitations in this section, which is the study by Garzia, Trechsel, and De Angelis (2017). Their study was conducted in a real electoral context and contained multiple measurements. Their research did not suffer from potential two-sided noncompliance, as there was no VAA during these particular elections except for the one they developed themselves for the experiment. Moreover, the results of the treatment group were compared to a no tool

group which both consisted of randomly selected people. Their study found that VAA users cast more votes during the election than nonusers (e.g. a difference of 10.7 percentage points). Therefore, there is strong experimental evidence that VAA usage could indeed increase voter turnout. Nonetheless, the approach of Garzia, Trechsel, and De Angelis (2017) does suffer from ethical limitations as, due to their experiment, one part of the public could easily inform themselves about the upcoming elections, whereas the other part did not have this option. More research is needed to test whether their results can be replicated while taking into account the obvious ethical issues. Moreover, the study of Garzia, Trechsel, and De Angelis (2017) merely looked at a part of the variables of interest in this study (e.g. voter turnout), therefore, for the other variables of interest, there is still a need for research that compares the VAA results with a no tool condition. The present study aims to satisfy these two needs in VAA research by conducting experimental research with a mix between a field and a laboratory study in which tool conditions will be compared to a no tool condition, making use of a newly developed CAVAA and a regular VAA. Moreover, this study will take place in an election context without pre-existing VAA, and does not suffer from the ethical limitations as one of the tools will be spread publicly after the pre-experimental measurement.

Factual and perceived political knowledge

In VAA research, political knowledge is one of the most investigated variables as the enhancement of political knowledge is one of the reasons why the tools were invented in the first place. However, this variable is most often being researched by making use of self-reports, in which participants are usually asked to indicate on a scale whether they, for instance, think that their political knowledge has increased due to the tool. An ever-existing problem with these types of measures is that one can simply not be sure if whatever the participant responds is indeed the case, as such answers depend on the person's own standards, which can be different for other people (Austin et al., 1998). Moreover, respondents of self-report measures tend to overreport their actual perceptions due to the tendency to report socially desirable answers (Karp & Brockington, 2005). This could lead to questions about the validity of the results found in a great amount of VAA studies with regard to political knowledge. If one measures political knowledge in such a subjective manner, then one can merely obtain insight into the perceived political knowledge of the participant. These perceived results might not directly translate into the factual results as research has provided

evidence with regard to discrepancies between perceived and factual knowledge (Hollander, 1995; Mondak, 1995; Park, 2001). For instance, Mondak (1995) showed that exposure to local newspapers did give the public the sensation of having gained political knowledge with regard to the U.S. elections, whereas their answers to 17 factual knowledge questions revealed that this was factually not the case.

In order to actually measure the factual political knowledge about specific and politically relevant issues for the upcoming election, rather than about political aspects in general, (which is a risk with self-report measures for political knowledge, as people might indicate a knowledge increase if they feel like they learned something more about anything politically related, thus also relatively unimportant aspects) the users of a VAA should be asked to answer several factual questions related to politics. As mentioned earlier, political knowledge can consist of semantic and pragmatic knowledge (Kamoen & Holleman, 2017), as well as knowledge about the positions of the parties. To the author's knowledge, it seems that all VAA research that did actually make use of a factual knowledge measure happens to have chosen to either measure the knowledge about semantic and pragmatic matters (e.g. Kamoen & Holleman, 2017; Kamoen & Liebrecht, 2022) or merely the knowledge about the political parties (e.g. Schultze, 2014; Van de Pol, 2016; Westle, Begemann & Rütter, 2015). The present research will differentiate between perceived and factual knowledge and will measure the latter by making use of a combination of the approaches used in prior research on the subject.

Focus on short-term effects

Another issue in most prior VAA research is related to the short-term focus of these studies. In typical research towards these tools, the participants are asked to fill out a short survey after having used a VAA, testing for any effects (e.g. Andreadis & Chadjipadelis, 2011; Garry et al., 2019; Kamoen & Liebrecht, 2022). Although this type of research could provide insight into the immediate effects of the tools, it does not make clear whether the effects found will last or decrease over time. It is important to obtain insight into the duration of the effects of VAAs as longitudinal effects are required to actually achieve the objectives crucial for an ideal democracy. For instance, the democratic ideal of an interested, knowledgeable, and participating public requires the people to continuously score high on these factors rather than solely after having filled out a VAA.

When taking a look at previous research, one might reason that political interest and political knowledge are likely to decrease over time (Bode, 2016; Chaffee & Kanihan, 1997; Hidi & Renniger, 2006). Concerning political interest, it is insightful to discuss the study by Hidi and Renniger (2006) in which they developed a four-phase model that explains how one's interest in a domain can develop over time and that it is dependent on, among others, affective, cognitive, and social factors. Accordingly, this would mean that if one, for instance, does not feel enough support from peers (e.g. peers not approving one's interest in a specific domain), their level of interest in the domain might decrease or even fall back completely to their initial level. The authors add that these types of changes in interest are most likely to appear when one has not yet developed a very strong interest in the specific domain. However, most experimental research does indicate strong degrees of interest stability over time when it comes to politics (Neundorf, Smets & García-Albacete, 2013; Prior, 2010). This is interesting because research toward the effects of VAAs on political interest, as mentioned before, actually did find enhancements in the levels of interest (e.g. Fivaz & Nadig, 2010; Garzia, Trechsel, & De Angelis, 2017; Marschall & Schmidt, 2010). This might indicate that VAAs have the ability to merely temporarily change one's interest in politics and that the user does not yet develop a high level of interest by using the tool, leading to a decrease or fallback to the user's original level of political interest after a certain period of time. As a consequence, it can be reasoned that VAAs possibly merely have the ability to trigger interest by creating short-term changes in one's affective and cognitive processing, which is what Hidi and Renniger (2006) see as the initial phase of interest development, which they call "Triggered Situational Interest". Therefore, it would be interesting to verify whether the positive effects found on political interest as a result of VAA (or CAVAA) usage would remain stable over time.

With regard to political knowledge, there seems to be an absence of longitudinal data as well. As discussed in the previous section, research should differentiate between perceived and factual knowledge, also when it comes to political knowledge. Therefore, in this paragraph, the effects of time on both knowledge variables will be discussed.
Factual political knowledge has nearly entirely been researched by asking VAA users to fill out several factual questions related to politics after having used the tool. To the author's knowledge, there has been one VAA study that factually measured political knowledge after VAA usage with some time in between the questionnaire and the tool usage, which is the study by Van de Pol (2016). In that research, the participants who answered the knowledge

questions had used the VAA for the upcoming election somewhere in the five weeks prior to the questionnaire. This implies that the data is likely obtained from people who had filled in the tool a long time before the survey as well as users that only recently received a voting advice. Therefore, their dataset contained both longitudinal and short-term data. Interestingly, the study by Van de Pol (2016) is the only study on the effects of VAAs on factual political knowledge that did not find a significant relationship (compared to, for instance, Schultze, 2014; Westle, Begemann & Rütter, 2015). Therefore, it could be suggested that the longitudinal data might have had a negative impact and that there would have been a positive effect if short-term and long-term data had been separated during the analysis. Would this suggestion be true, it would indicate that an increase in factual political knowledge obtained due to VAA usage is only of short-term duration. However, research suggests that one's level of factual political knowledge is dependent on the number of exposures with regard to political information (Bode, 2016; Chaffee & Kanihan, 1997). Therefore, whether a VAA user proceeds to contain a similar amount of factual political knowledge as was the case immediately after usage of the tool, might depend on the number of exposures to political information this person has after having used the VAA. More research is needed to test the stability of the effects of VAAs on factual political knowledge, which is one of the aims of the present study.

Perceived political knowledge has been researched much more often than factual political knowledge and has most often been found to be positively influenced by VAA usage (Fivaz & Nadig, 2010; Kamoen & Liebrecht, 2022). When it comes to longitudinal data in this regard, there are three prior studies in which such data can be found (e.g. Fivaz & Nadig, 2010; Kamoen et al, 2015; Ladner, 2012). All of these studies indicate that a significant amount of people perceive their political knowledge to be improved due to the use of the voting advice tool. However, in all three studies there happens to be, just as in the aforementioned study by Van de Pol (2016), a strong likelihood that the data is a mixture of both longitudinal and short-term data. For instance, in Ladner (2012) users of a Swiss VAA called "*Smartvote*" filled out a questionnaire after the elections were held. The participants could have filled in the VAA up to six weeks before the elections and it is therefore likely that the data consists of both longitudinal as well as short-term data. However, in this particular research, it might as well have been the case that all the data is longitudinal as it is not specified exactly how long after the elections the survey was held. By all means, it remains

unclear whether the positive results found in perceived political knowledge due to VAA usage are susceptible to change over time.

Nonetheless, it can be reasoned that one's perceived political knowledge decreases if one does not repeatedly encounter stimuli about politics (similar to factual political knowledge), as familiarity is an important heuristic when it comes to one's ability to judge their knowledge about a topic (Metcalfe, Schwartz & Joaquim, 1993). In other words, it appears to be that if one is familiar with a certain topic then they are more likely to perceive themselves as knowledgeable with regard to that specific topic. Familiarity, in turn, increases when one repeatedly encounters stimuli related to that topic (Koriat, 1995) and perceived knowledge is therefore likely to increase when people regularly encounter political information as well as likely to decrease when they do not encounter such stimuli. It can therefore be expected that VAAs temporarily enhance one's perception of their own political knowledge, as they are being exposed to political stimuli, but these effects might alter depending on the number of stimuli this person receives after the VAA usage.

One's voting intention has been found to be higher if one had recently been exposed to political stimuli (Durante & Gutierrez, 2012; Freedman, Franz & Goldstein, 2004). However, with respect to voting intention, it is important to have longitudinal data as well as research does indicate that people are likely to overreport their voting intention (Dolan, Metcalfe & Powdthavee, 2008; Rogers & Aida, 2011) which makes that the actual voter turnout is often lower than the number of people that indicate to go cast a vote during the elections. Therefore, VAA studies should not conclude that VAAs increase voter turnout based solely on a question that measures one's voting intention. Moreover, the results of studies in which participants have filled out a VAA and after the elections are asked to indicate whether they had cast a vote, should also be interpreted with caution. Research indicates, namely, that, besides voting intention, people oftentimes overreport voter turnout as well (Karp & Brockington, 2005; Traugott & Katosh, 1979), in order to transmit a socially desirable image of themselves (Andolina et al., 2003; Belli, Moore & VanHoewyk, 2006; Karp & Brockington, 2005). Research tends to suffer from this so-called social desirability bias if the questions are about possibly sensitive topics, such as voting. However, with regard to the longitudinal effect of VAAs on voting intention (e.g. the longitudinal effect is reflected in whether or not one casts a vote), there seems to be a strong body of research that indicates that VAAs could actually lead to more voter turnout even if there is some time between the tool usage and the election (Garzia, De Angelis & Pianzola, 2014; Garzia, Trechsel & De Angelis, 2017; Gemenis &

Rosema, 2014). For instance, in the study by Gemenis and Rosema (2014), it was found that in the Dutch parliamentary election of 2006, about 4.4% of the voter turnout was attributable to the usage of a VAA by making use of data containing information that the participants had used a VAA sometime before the election. They estimated the actual voter turnout as a result of the VAA by making use of real turnout data, rather than self-reports that are, as discussed, potentially influenced by the social-desirability bias. Thus, this particular variable seems to be researched with longitudinal data for most of the studies, and these studies do therefore not conclude the positive relation between VAAs and voting turnout on merely a voting intention measure.

The present study

The aim of the present study is to conduct research on the effects of VAAs and CAVAAs on three factors that are known to be important for democracies (e.g. voter turnout, political knowledge, and political interest) while taking into account the limitations and issues found in prior studies on the topic. To this end, this research will compare the effects of VAA, CAVAA (Kamoen & Liebrecht, 2022), and no tool usage on the perceived and factual political knowledge, voting intention and voter turnout, and political interest on three different measurements (e.g. pre, immediately post, and two weeks after usage) in a real election context. This election context is the 2022 university council election at Tilburg University. During these elections, students of the university have the opportunity to vote which 9 students will form the university council (UC) for the upcoming academic year.

Based on the finding that one's political interest can be enhanced merely by exposure to political stimuli (Strömback & Shehata, 2010), it is expected that the political interest will increase between T1 and T2 due to the exposure to political stimuli that is present in the questionnaire. Moreover, as due to the tools, the three conditions differ in the number of political stimuli that the participant is exposed to, it is expected that the participants in the CAVAA condition will obtain more political interest than those in the VAA and no tool condition and that the people in the VAA condition will experience a greater increase in political interest than those in the no tool condition. The latter is also expected as positive effects of VAAs on political interest were also found in prior studies (Fivaz & Nadig, 2010; Garzia, Trechsel, & De Angelis, 2017; Marschall & Schmidt, 2010).

However, due to the fact that the number of political stimuli that is present both online and offline is lower two weeks after the elections (e.g. T3) compared to right before the elections (e.g. T1 and T2), and because Neundorf, Smets, and García-Albacete (2013) and Prior (2010) found that one's interest in politics remains stable over time, it is expected that the levels of political interest will decrease between T2 and T3, regardless of condition. However, although the number of political stimuli available at T3 will probably be lower than at T1 it is expected that the levels of political interest at T3 will still be similar to the ones at T1.

With respect to the perceived political knowledge, a similar pattern is expected as for political interest. An increase is expected between T1 and T2 because the exposure to political stimuli will increase one's familiarity with political information (Koriat, 1995), and this familiarity is expected to increase the perceived political knowledge of this particular person (Metcalfe, Schwartz & Joaquim, 1993). However, due to the decrease in political stimuli around T3, the increased perceived knowledge levels are expected to fall back to a level similar to the one at T1. Based on the reasoning that the more political stimuli one encounters the more perceived knowledge they will acquire, it is expected that the people in the CAVAA condition will have more perceived knowledge than the other two conditions and that the VAA users also got higher perceptions of their political knowledge than those in the no tool condition. The latter hypothesis is also based on the findings by Fivaz and Nadig (2010), and Kamoen and Liebrecht (2022), which show a perceived knowledge increase between no tool and VAA users.

Concerning factual political knowledge, an increase is once more expected between T1 and T2 as a result of an increase in exposure to political stimuli, as previous work has found evidence for this positive relationship between the number of exposures to political stimuli and factual political knowledge (e.g. Bode, 2016; Chaffee & Kanihan, 1997; Schultze, 2014; Westle, Begemann & Rütter, 2015). Moreover, the factual knowledge at T3 is expected to revolve around the level found a T1, due to the decrease in available political stimuli. The latter can also be expected based on the study by Van de Pol (2016) which, at least partially, consists of longitudinal data and did not find an increase in factual political knowledge. When taking into account the conditions, it is also expected, based on the positive relationship found between VAA usage and factual political knowledge by Schultze (2014) and Westle, Begemann, and Rütter (2015), that the people in the tool conditions will have a higher factual political knowledge than those in the no tool conditions. Moreover, it is expected that the

people in the CAVAA condition will answer more knowledge questions correctly than those in the VAA condition due to the higher level of exposure to political stimuli in the CAVAA condition.

Finally, when it comes to voting intention and voter turnout, an increase in voting intention is expected between the first and second measurement, as the participants have been exposed to more political stimuli at T2, which can result in a higher voter intention according to the studies by Durante and Gutierrez (2012), Freedman, Franz and Goldstein (2004). Moreover, the voter turnout is also expected to be higher due to exposure to political stimuli as multiple studies have found strong results indicating that VAA usage (by which participants are exposed to political stimuli) could lead to a higher voter turnout compared to not using a tool (Garzia, De Angelis & Pianzola, 2014; Garzia, Trechsel & De Angelis, 2017; Gemenis & Rosema, 2014). Therefore, the people in the CAVAA condition are expected to show a higher voter intention at T2 and a higher voter turnout at T3, than the people in the VAA and no tool conditions.

Method

Design

This study has a one factor design with three conditions (VAA/CAVAA/no tool), which were tested on three different measurements (T1, T2, and T3). Each participant was randomly assigned to one of the three conditions and filled out a survey with measures related to the dependent variables at three different points in time. The first measurement moment (T1) took place shortly before the elections started. Participants were asked to fill out a questionnaire measuring their interest in politics, voting intention, perceived political knowledge, and factual political knowledge (T1). Immediately after filling out this questionnaire, the participants were randomly assigned to one of the three conditions. Participants in the tool conditions either filled out a VAA or a CAVAA that both provided them with several political statements about the 2022 Tilburg University council elections, as well as a voting advice based on their answers. Immediately afterward, they were presented a similar questionnaire with post measures on the dependent variables (T2). The participants of the no tool condition received an extended version of the T1 survey as it also contained the factual knowledge questions of T2. To measure the longitudinal effects of the manipulation, the participants were asked to fill out the last questionnaire (T3), which was identical in all

conditions, about two weeks after the elections. Upon completing this survey, the participant had finished the experiment. The ethics commission and the election board of the university had to give permission to execute the current study as there were concerns that not all students would receive similar information with regard to the elections if not all people got to fill out a voting advice application, this permission was granted by both bodies.

Participants

A total of 172 participants took part in the first two measurement moments (T1 and T2). All participants were students at Tilburg University and were, therefore, able to cast a vote during the university council elections. As they all were university students, the study exclusively consisted of highly-educated people. Several participants were removed from the analyses as they had not answered a great part of the questions and one participant was removed as they did not fill in the CAVAA in the corresponding condition. This led to a total of 166 participants of whom the data from measurements T1 and T2 were analyzed. Among those remaining participants, 49 were in the CAVAA condition, 56 took the VAA and 61 participants did not use any tool. This sample consisted of 104 (62.7%) females, 60 (36.1%) males, and 2 (1.2%) participants identified as something else. Moreover, the mean[1] age of the participants in the first two measurements was 21.98 years old (*SD* = 3.35), of which the youngest participant was 17 years old and the oldest 39 years.

For various reasons (e.g. participants not participating in T3, incorrectly using their unique three-digit number, not completing great parts of the T3 survey, or indicating to have used a voting advice application despite it not being part of their condition) participants dropped out of the study between T2 and T3. This has led to a remaining 104 participants for T3 of which 33 were in the CAVAA condition, 36 in the VAA condition, and 35 in the no tool condition. With regard to these participants the mean[1] age was 21.97 years old (*SD* = 3.04) and the range was 17-39. Furthermore, among these 104 participants, 65 (39.2%) were female, 37 (22.3%) male, and 2 (1.2%) identified themselves differently.

[1] Five people wished not to share their age. Therefore, the mean age and standard deviations for age are calculated using the data of the remaining 99 people at T1, T2, and T3.

Materials

Statements

In order to come up with the statements to be used in the VAA and CAVAA, information was gathered about the two parties that took part in the elections (e.g. Party SAM and Party Front) and about the issues that played a role at that time by investigating the minutes of the most recent university council meetings, by looking at several of the proposed initiatives of the two parties and by visiting the websites of both parties. However, after having thoroughly investigated all the sources, the conclusion had to be made that the two parties did not substantially differ in their stances with regard to nearly all issues that were discussed by the university council. Therefore, it was impossible to come up with statements that showed opposing stances and to ask the participants to indicate whether they are in favor or against a certain statement, which is the standard in regular voting advice applications. Consequently, another type of statement needed to be created and as the two parties did differ in what their overall focus was, it was decided to invent statements about matters that differed in whether or not their realization was one of the top priorities for each party. To this end, a list of topics that played at the level of the university council was produced. This list was shared with the election board as well as with representatives of both parties, as it was required to have consent from all these bodies in order to develop the applications. There have been several meetings with members of both parties. In these meetings, the list of topics was discussed as well as the stances of the political parties on these topics. In a final meeting, representatives of both parties shared their election programs which led to the creation of the final eighteen statements that have all been checked and approved by both parties. For fourteen of the statements, one of the parties considered the specific issue to be of top priority and the other party did not. In total there were seven statements for each party which one party considered to be of top priority as opposed to the other. This, in order to not favorize one party over the other. Moreover, besides these fourteen statements, there were two statements of which both parties perceived the topic to be of top priority (these were merely there to inform the user and did not count towards the voting advice), two statements about topics for which they actually did have different stances, and one practice statement in the beginning in order for the participant to understand how to use the chatbot (which was not used in the analyses). The number eighteen was chosen as the total number as there existed a risk that participants would lose motivation if there were too many statements and show

satisficing behavior (Krosnick, 1996), leading to potential inaccurate responses. This number is very similar to the number of statements that are normally used in VAAs, which lies around 20 – 30.

Answering options

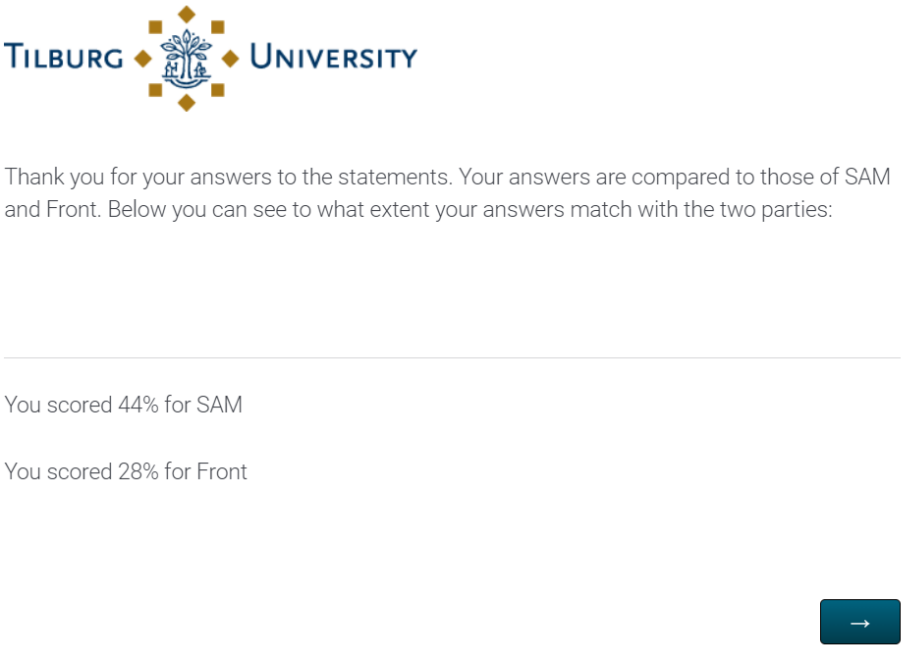
Due to the fact that both parties did not have very opposing stances, there also needed to be formed new questions and answering options. As the statements were about whether or not the specific issue was one of the top priorities for each party, the questions in the voting advice application were also changed accordingly. These questions could either be answered by clicking on one of the three answer buttons containing the options “Yes”, “No”, and “I don’t know”, or by typing one of the three options in the open text field. The “I don’t know” button initially contained a smiley with a neutral face, which apparently made the participants feel kind of incompetent if they were to choose that option, which is why the smiley was removed for the final version.

VAA

Every statement in the VAA was written down as a multiple choice question and no additional information could be requested. The final voting advice for each party was presented in percentages of similarity between the answers of the tool user and those of the parties (See figure 1). The percentages were made visible by adding a suitable JavaScript in Qualtrics. After having received the voting advice, the participants could go to the next page, which contained information about the rest of the questionnaire (e.g. T2).

Figure 1.

Presentation of the voting advice in the VAA



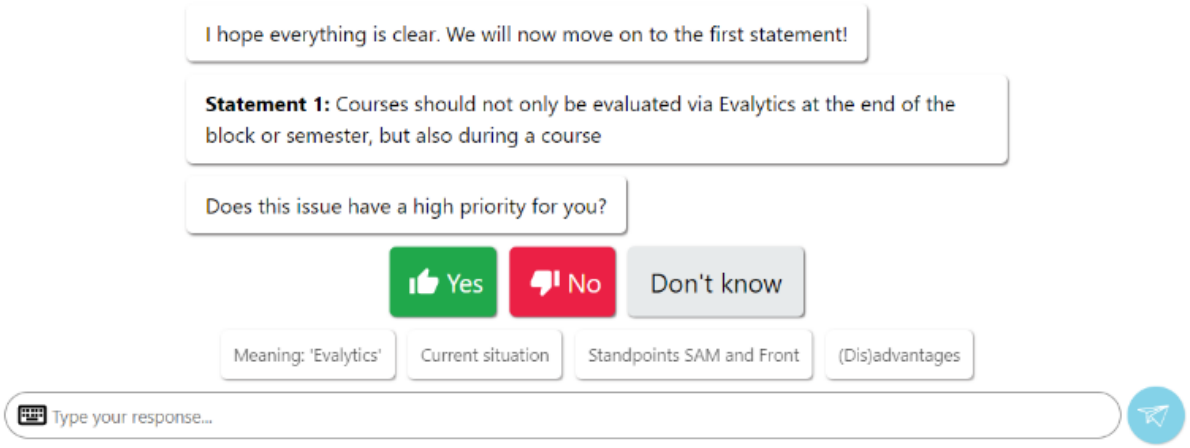
CAVAA

The CAVAA that was used in this research is developed by startup company Genius Voice. During the whole process, close contact was remained with Genius Voice, which made it possible to implement a lot of changes based on the feedback resulting from the multiple pre-tests that have been executed prior to launching the final tool.

In the CAVAA, the same 18 statements were presented but this time with the option to request additional information. The CAVAA consisted of a semi-structured chatbot, which means that the user had the possibility to either click on buttons to perform a certain action as well as the possibility to type in an utterance in order to perform the desired action (Kamoen, McCartan & Liebrecht, 2021). The tool did not contain many colors, and the chatbot did also not contain an (anthropomorphic) face, body, or pictogram as each of these factors would make the layout significantly different from the VAA. Instead, the CAVAA looked mostly grey and white and the messages from the chatbot just popped up on the screen as ordinary text balloons (see figure 2).

Figure 2

The look and feel of the CAVAA



When starting the CAVAA, the chatbot would welcome the user by letting a message pop up stating: "*Hi there, great that you want to fill out this voting advice application!"*. Then, additional messages followed in which the user was reminded that the chatbot would help them determine which party matched them best, and in which the user was asked whether they would like to start with the statements. Upon agreeing, the participant would be presented with information about the statements and that they could indicate whether they considered the statements to be one of their top priorities. The latter was followed by the explanation: "…*Meaning that you hope that the party representing you will bring the issue to the attention of the University Council, for example by proposing an initiative regarding the issue*". Then, the participant was informed about the two manners in which they could respond to the statements.

Furthermore, at any time, the chatbot provided the user with the three aforementioned answer options and an open text field at the bottom of the screen. Moreover, the users of the CAVAA were able to click on a total of four information buttons in order to obtain more knowledge about the topic of the statement. For instance, if one desired to receive information about the definition of a concept (semantic information), then this was possible by clicking on the button stating "*Meaning X*". Similarly if one needed to obtain information about the current state of affairs with regard to the central concept (pragmatic information) then the

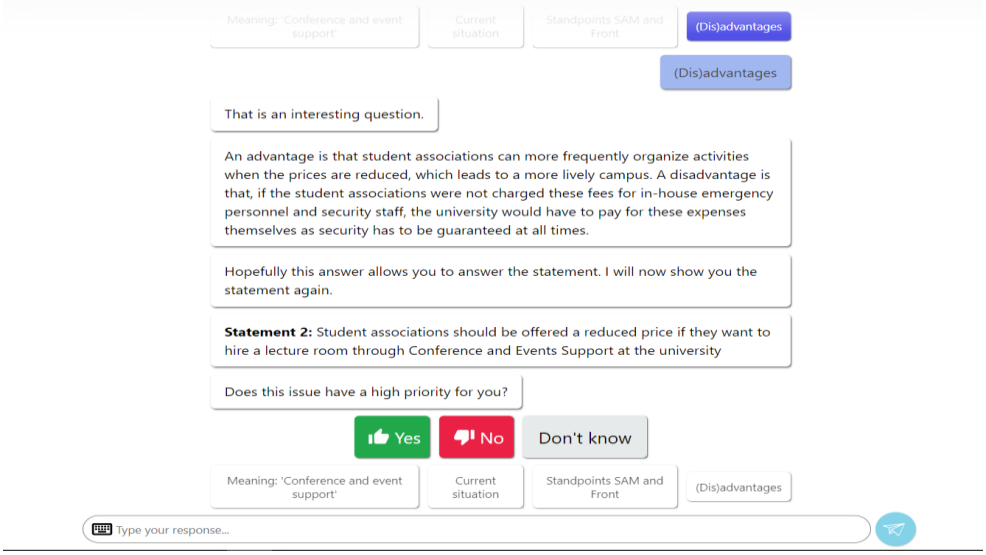
button with the text "*Current situation"* could be pressed. These two types of information were included as these are valuable for one's factual political knowledge and because these have been found to often lead to comprehension problems among VAA users (Kamoen & Holleman, 2017). Furthermore, a third button with the text "*Standpoints Sam and Front*", would make the chatbot show information about whether each party considered the realization of the idea presented in the statement to be one of their top priorities. This type of information could be valuable in order to get an insightful picture of what each party's main focus is on and what they stand for. Finally, the last information button provided insights into both the advantages and disadvantages of the implementation of the idea presented by the statements. This button simply stated "*(Dis)advantages*" and was included as this is valuable information, as it could help the user to make an informed decision with regard to the statement as they become able to weigh the arguments. When one clicked on, for instance, the "*Current situation*" button then there would immediately pop up multiple text balloons of which one contained the answer, stating for example: "*Currently, there is no digital Indoor Campus map for Tilburg University. Google maps can be used to navigate to specific buildings, but this app currently does not have information on the location of lecture rooms within buildings. There is signage available within buildings, as well as on campus.*"

Moreover, the reason that several text balloons pop up is that the chatbot was programmed in such a way that it also uttered conversational messages such as: "*That's a great question, thank you!",* "*I hope this answer has helped you to understand the issue. Let's return to the statement now*.", and "*Registered. Let's move on to the next statement*." because the aim was to let the user feel like they were talking to an actual human being rather than a robot (see figure 3). This is important in chatbot usage as these conversational elements make the user experience the tool to be easy to use and they lead to improved and increased interactions between humans and computers in general (Tsai, Liu & Chuan, 2021; Zadrozny et. al., 2000). Furthermore, when one typed in a question or an answer in the open text field, there existed the risk that the chatbot did not understand what was meant by the utterance. When such a situation occurred, there would pop up a text message in which the chatbot mentioned that it did not understand the question, sometimes preceded by an apology, for example: "*Sorry, I do not understand your question*". Then, the chatbot would try to let the user repeat their question by asking them, for instance, "*Would it be possible to rephrase your question?"*. Finally, if there happened to be another miscomprehension upon the rephrased question, then there would arise a text message aimed to inform the user about the

information buttons which stated, for instance, "*Perhaps you can use one of the buttons to get the information you need*?". A video in which the CAVAA is being used can be found in Appendix A

Figure 3

Example of conversational elements in the CAVAA



Pre-tests

Pre-tests were conducted for each condition with the purpose of verifying whether everything was clear, whether no information was missing, whether the chatbot in the CAVAA condition contained an adequate conversation, and whether the chatbot functioned properly at all. The procedure and the results of these pre-tests can be found in Appendix B.

Voting advice

The voting advice in each tool was calculated in the same manner. If the issue in a certain statement was one of the top priorities for party Front, then the participant would receive one point for party Front if they selected "yes" in the (CA)VAA. Similarly, if one had selected one of the other two buttons then this was counted as zero points for Front and one point for SAM. For the two statements that both parties considered to be of top priority,

answering “yes” resulted in a point for both parties. While filling out the (CA)VAA, the points for each party were added up and calculated into percentages, which indicated how many percent the user agreed with each party. This is the voting advice that the participants received after having answered the last statement

Instrumentation

The measurement of each dependent variable will be discussed in the present section. The political interest and perceived political knowledge questions were intentionally asked before the participants filled in the factual political knowledge questions, as this variable might have influenced the scores on the former two.

Interest in politics. The interest in politics was measured by first providing the participants the statement: “*I am interested in what issues are discussed in the university council*”, for which they had to indicate to what extent they agreed on a 7-point Likert scale ranging from (1) “completely disagree” to (7) “completely agree”.

Perceived political knowledge. Perceived political knowledge is measured by 2 items of which the first has been derived from Kamoen, McCartan and Liebrecht (2021) and was adapted to the current study. The first item was: “*I feel sufficiently informed about what party SAM and party Front stand for*” and the second: “*I feel informed about the political issues that are discussed in the university council”.* Both statements could be answered on a 7-point Likert scale ranging from (1) “completely disagree” to (7) “completely agree”. These two items appeared in all conditions at all three measurements.

Factual political knowledge. Factual political knowledge was measured by showing the participants seven statements about issues discussed by the university council. For all of these statements, they were asked to indicate whether they were true, false or whether they did not know it. All the answers to the factual knowledge questions could be found in the CAVAA condition, but not in the other two conditions. The questions were about semantic and pragmatic information (as was the case in Kamoen et al., 2015; Kamoen & Liebrecht, 2022) and about the stances of the parties (as was done by e.g. Schultze, 2014; Van de Pol, 2016; Westle, Begemann & Rütter, 2015). These studies measured knowledge via one of these two approaches, rather than testing them both. As the present study applies both

methods, this study combines the approaches used in the literature. See Appendix C for all factual knowledge questions.

Each correct answer provided a point during the analyses, which means that a total of seven points could be obtained during each measurement. The option "Do not know" is coded as an incorrect answer during the analyses. Moreover, the statements used were different at each measurement but were the same in each condition. Thus, there were a total of 21 factual knowledge questions, which were presented in sets of 7 during all three measurements.

Voting intention/voter turnout. Voting intention was measured before the elections (at T1 and T2) by one statement derived from Kamoen, McCartan, and Liebrecht (2021): "*I intend to cast a vote in the 2022 Tilburg University council elections".* As was the case in Kamoen, McCartan, and Liebrecht (2021) the participants were able to answer "*yes*", "*no*" or "*I don't know yet*" and the last two options were added together during the analyses. Moreover, voter turnout was measured after the elections (at T3) by providing the participant with the statement "*I have cast a vote in the 2022 Tilburg University council elections"* which they could merely answer with yes or no.

Demographics. The demographic questions were about the age of the participants (which could be typed in or not be indicated if preferred), their gender (with the options being: male, female, other), and how long they had been students at Tilburg University (multiple choice ranging from less than one year to more than five years), and the faculty that they were taking their study at (which could be selected). Moreover, all demographic questions contained a "prefer not to say" option.

Control questions. The control questions were about whether the participant had used the voting advice tool that was spread during the elections (e.g. the CAVAA), and if so, whether they did it outside of this study. Data of participants who indicated to have used the CAVAA outside of this study and who were not in the CAVAA condition was removed. These questions were solely asked during T3 and could be answered by clicking on either a "Yes", "No", or "I don't know" answer option.

Procedure

This study was conducted between April 5th, 2022 and May 7th, 2022. The first and second measurements took place shortly before the start of the elections. Initially, participants were recruited via the human-subject pool of Tilburg University, which is a large group of students from the School of Humanities and Digital Sciences (TSHD) faculty of the university. However, eventually, people from outside of this pool were recruited as well due to the fact that insufficient students enrolled themselves in the study. Participants were rewarded with a full "proefpersonen point". Bachelor and pre-master students of two faculties at Tilburg University need to obtain a certain amount of these proefpersonen points, which they merely are able to receive by taking part in studies. The students that desired such a point could indicate so via a corresponding system at Tilburg University. In this way, the participants could receive 1 full point (e.g. 0.5 points per survey).

It was aimed to let the participants that were recruited via the human-subject pool participate in the study in a controlled manner on the university campus. Once they arrived at the room, the participants were asked to take a seat in front of a computer. Then they were thanked for their participation, informed that they could leave the experiment at any moment without providing a reason, and were instructed that they would fill in a survey on the computer. Then, they were each handed a unique three-digit number that they were asked to either make a picture of or to write down in their phone. This number served as an identification to link all the data across the measurements. Then, the participants were randomly emailed a link to one of the three conditions. Upon clicking on the link, the participants would open up the Qualtrics questionnaire on the computers and fill it out (T1). All questionnaires started with a welcome message followed by an informed consent form which contained information about the data collection, the ethical approval, the fact that the participants could leave at any moment without having to provide a reason, and the reward that they would receive after completing the two parts of this study. After having given consent, the participants could go to the next page in which they were asked to fill in the unique three-digit number that had been given by the researcher.

They were then directed to the first question of the survey after having given consent. First, the participant's political interest was measured followed by questions that tested for the user's perceived political knowledge. Then, the factual political knowledge questions were presented to the participant, as well as questions related to their voting intention. Lastly, some control and demographic questions were asked. Another question about the sureness of the

user which party to vote for popped up if one had indicated to vote during the council elections. This question, however, was not used in the analyses.

When finished, the participants in the CAVAA conditions were presented with a page that contained a link that led to the CAVAA and an instruction about what this voting advice application is and that they should return to the survey page once they have received their voting advice. The participants in the VAA condition were shown a page that informed them about the VAA that was programmed inside Qualtrics, which would provide them a voting advice, and were asked to fill that one out. Both conditions then received another set of questions (T2).

It should be noted that the no tool condition had some differences in comparison with the CAVAA and VAA questionnaires, such as some alternations in the consent form, no direct post-measurement, and twice as many factual political knowledge questions at T1. The latter was the case as they would not fill in a direct post-measure survey (as this would presumably lead to the same answers as at T1, as no manipulation had taken place) and it would be more practical during the analyses if all participants had the same amount of factual political knowledge questions in total. These extra questions were therefore counted as T2 for the no tool condition during the analyses (although there not actually being three different measures for this condition) and were the same as the actual factual knowledge questions that the people in the VAA and CAVAA conditions received at T2. All participants were instructed that they could leave the room as soon as they had finished the questionnaire.

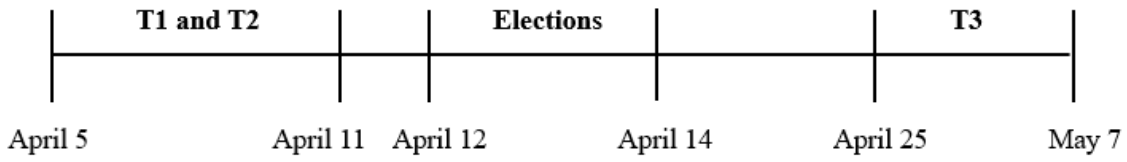
The second part (e.g. T3) of the experiment took place approximately two weeks afterward. The participants once again came to the computer room and were asked to fill out the last questionnaire (T3). The questionnaire at T3 was the same for all three conditions. After having given consent the participants were asked to once again fill in their unique three digit number in order for the data to be linked between the different measurements. When having finished all the questions, they were directed to another screen in which they were debriefed. They had now completed the experiment and could be given their second half of the proefpersonen point.

As mentioned before, it was initially planned to let all participants fill out the survey in a controlled setting on campus via the human-subject pool. Nonetheless, as an insufficient number of students applied via the human-subject pool and because the elections were almost about to start, it was decided to also recruit participants outside of the human-subject pool,

who were allowed to do the survey online and therefore did not have to come to the computer room. This was done by approaching random groups of students that were present at the university campus and asking them whether they were students at Tilburg University and whether they would want to participate in the study. Upon agreement, they were each given a unique three-digit number, asked to take a picture of it, and were being sent the survey to their email addresses. Two weeks afterward, the survey for the second part of the study was sent to those people via email as well with the instruction to fill that one out too. See figure 4 for an overview of the time period in which this study has been conducted and when every event took place.

Figure 4

Overview of the time period of all the events that were part of this study



Results

For each dependent variable[2], a two-way repeated-measures ANOVA was performed, with time (T1/T2/T3)[3] as within subjects-factor and condition (no tool/VAA/CAVAA) as between-subjects factor. In case a significant effect was observed, pairwise comparisons with

[2] See Appendix D for the correlations between political interest, perceived political knowledge, and voting intention

[3] As the no tool condition did not have a measurement at T2, the values of T1 for political interest, perceived knowledge, and voter turnout were copied and also used as the values for T2 in the no tool condition during the analyses of these variables. The values in the no tool condition are presumed to be the same in T2 as in T1 since there was not any manipulation for this condition and because T2 almost immediately followed T1. However, the no tool condition did have a T2 measurement for the factual knowledge questions, but these questions were asked at T1 and were the same as the T2 factual knowledge questions in the tool conditions.

Bonferroni correction were applied, unless the assumption of sphericity was not met, in that case, a Greenhouse-Geisser correction was applied.

Political interest

The two-way repeated measures ANOVA revealed that there was a significant main effect of time on perceived political interest, sphericity assumed $F(2, 202) = 4.18$, $p = .017$, $\eta^2 = .040$. Pairwise comparisons using a Bonferroni correction show that perceived political interest was significantly higher on T2 ($M = 4.18$, $SD = 1.58$) compared to T1 ($M = 3.82$, $SD = 1.57$, $p = .011$), but not compared to T3 ($M = 3.94$, $SD = 1.61$, $p = .174$). Also, no significant differences with respect to perceived political interest were found between T1 and T3 ($p = 1.000$). The test did not reveal a significant main effect of condition on perceived political interest $F(2, 101) = 1.04$, $p = .359$, $\eta^2 = .020$. Moreover, there was no significant interaction effect between time and condition with respect to perceived political interest, sphericity assumed $F(4, 202) = 1.52$, $p = .197$, $\eta^2 = .029$. See table 1 for the means and standard deviations. All in all, this suggests that interest is merely increased during the elections as compared to before and irrespective of whether one had used a CAVAA, VAA, or no tool at all.

Table 1

The number of participants (N), means (M), and standard deviations (SD) with respect to the perceived political interest in the three conditions at each time. The M and SD scores are based on a seven-point Likert scale ranging from (1)"Completely disagree" to (7) "Completely agree".

Condition	Time	N	M	SD
No tool	1	35	3.83	0.27
	2	36	3.83	0.26
	3	33	3.77	0.27
VAA	1	35	3.72	0.26
	2	36	4.06	0.26
	3	33	3.89	0.27
CAVAA	1	35	3.91	0.28
	2	36	4.70	0.27
	3	33	4.18	0.28

Perceived Knowledge

The results of a two-way repeated measures ANOVA indicated that there was a significant main effect of time on perceived political knowledge, Greenhouse-Geisser, $F(1.80, 181.49) = 45.50$, $p < .001$, $\eta^2 = .311$. Pairwise comparisons using a Bonferroni correction show that the perceived knowledge of the participants was significantly higher directly after the experimental manipulation (T2) ($M = 3.68$, $SD = 1.43$, $p < .001$), and two weeks after (T3) ($M = 3.57$, $SD = 1.47$, $p < .001$), compared to right before the experimental manipulation T1 ($M = 2.36$, $SD = 1.21$). However, there was no significant difference between T2 and T3 ($p = 1.000$).

Moreover, the test made clear that there was a significant main effect of condition on perceived political knowledge, $F(2,101) = 12.65$, $p < .001$, $\eta^2 = .200$, such that the scores in the no tool condition ($M = 2.65$, $SD = 1.21$) were significantly lower than those of the VAA condition ($M = 3.29$, $SD = 1.23$, $p = .007$) and the CAVAA condition ($M = 3.70$, $SD = 1.29$, $p < .001$). The VAA and CAVAA conditions did not differ significantly ($p = .168$). See table 2 for the means and standard deviations.

Finally, there was a significant interaction effect between time and condition, Greenhouse-Geisser, $F(3.59, 181.49) = 9.63$, $p < .001$, $\eta^2 = .160$. With respect to the no tool condition, there was no significant difference between T1 ($M = 2.40$, $SD = 1.03$) and T2 ($M = 2.40$, $SD = 1.03$), but there was an increase in perceived political knowledge at T3 ($M = 3.14$, $SD = 1.57$) compared to T1 ($p = .049$) and T2 ($p = .015$). However, when it comes to the VAA condition, it appeared that there was an increase in perceived knowledge at T2 ($M = 4.17$, $SD = 1.10$) compared to T1 ($M = 2.18$, $SD = 1.16$, $p < .001$) and this level decreased between T2 and T3 ($M = 3.53$, $SD = 1.43$, $p = .042$). However, the perceived knowledge was still higher at T3 compared to T1 ($p < .001$). Similarly, in the CAVAA condition, there is an increase from T2 ($M = 4.50$, $SD = 1.17$) compared to T1 ($M = 2.52$, $SD = 1.44$, $p < .001$) and from T1 to T3 ($M = 4.08$, $SD = 1.25$, $p < .001$). However, in the CAVAA condition, the participants' perception of their knowledge did remain the same between T2 and T3 as there was no significant difference ($p = .343$).

Table 2

The number of participants (N), means (M), and standard deviations (SD) with respect to the perceived political knowledge in the three conditions at each measurement. The M and SD scores are based on a seven-point Likert scale ranging from (1)"Completely disagree" and (7) "Completely agree".

Condition	Time	N	M	SD
No tool	1	35	2.40	1.03
	2	36	2.40	1.03
	3	33	3.14	1.57
VAA	1	35	2.18	1.16
	2	36	4.17	1.10
	3	33	3.53	1.43
CAVAA	1	35	2.52	1.44
	2	36	4.50	1.17
	3	33	4.08	1.25

Factual Knowledge

The two-way repeated measures ANOVA revealed that there was a significant main effect of time on factual political knowledge, sphericity assumed, $F(2, 202) = 32.63$, $p < .001$, $\eta^2 = .244$. Pairwise comparisons using a Bonferroni correction show that there was a significant increase directly after the experimental manipulation (T2) ($M = 2.29$, $SD = 1.07$) as well as two weeks afterward (T3) ($M = 1.76$, $SD = 1.40$) compared to right before the experimental manipulation (T1) ($M = 1.04$, $SD = 1.16$) ($p < .001$, for both comparisons). Moreover, the factual knowledge of the participants decreased between T2 and T3 ($p = .003$).

Furthermore, the test made clear that there was a significant main effect of condition with respect to factual political knowledge, $F(2, 101) = 4.49$, $p = .014$, $\eta^2 = .082$, such that the factual knowledge in the CAVAA condition ($M = 2.02$, $SD = 1.23$) was significantly higher than in the no tool condition ($M = 1.49$, $SD = 1.11$, $p = .016$). The levels in the VAA condition ($M = 1.60$, $SD = 1.19$) were not significantly different from the no tool condition ($p = 1.00$) and CAVAA condition ($p = .080$). Finally, there was no significant interaction effect between time and condition for factual political knowledge, sphericity assumed, $F(4, 202) = 1.64$, $p = .166$, $\eta^2 = .031$. See table 3 for the means and standard deviations.

All in all these results imply that factual knowledge is highest during the elections (in all conditions) and that across measurement moments the factual political knowledge is highest in the CAVAA condition.

Table 3

The number of participants (N), means (M), and standard deviations (SD) with respect to the factual political knowledge in the three conditions at each measurement. The M and SD scores are based on the number of correct answers to a total of 7 questions.

Condition	Time	N	M	SD
No tool	1	35	0.83	0.95
	2	36	1.86	0.97
	3	33	1.77	1.42
VAA	1	35	1.03	1.11
	2	36	2.39	1.23
	3	33	1.39	1.25
CAVAA	1	35	1.27	1.38
	2	36	2.64	0.82
	3	33	2.15	1.48

Voting intention and Voter turnout

The two-way repeated measures ANOVA made clear that there was a significant main effect of time on voting intention, Greenhouse-Geisser, $F(1.44, 145.30) = 36.69$, $p < .001$, $\eta^2 = .266$. Pairwise comparisons using a Bonferroni correction show that the voting intention was significantly higher at T2 ($M = .34$, $SD = .47$) than at T1 ($M = .26$, $SD = .44$, $p = .046$). Moreover, the actual voter turnout (e.g. T3) was significantly higher ($M = .673$, $SD = .47$) than the voting intention at T1 ($p < .001$) and at T2 ($p < .001$).

Moreover, there was a significant main effect of condition on voting intention $F(2, 101) = 4.35$, $p < .015$, $\eta^2 = .079$. The voting intention in the VAA condition ($M = .55$, $SD = 0.49$) was significantly higher than the voting intention in the no tool condition ($M = .31$, $SD = .40$, $p = 0.13$). However, there were no significant differences between the CAVAA

condition (M = .40, SD = 0.46) and the VAA (p = .242) and no tool (p = .814) conditions. Finally, there was no significant interaction effect between time and condition, Greenhouse-Geisser, $F(2.88, 145.30) = 2.37$, p = .075, η^2 = .045. See table 4 for the means and standard deviations.

Table 4

The number of participants (N), means (M), and standard deviations (SD) with respect to the voting intention in the three conditions at each measurement. The M and SD scores are based on whether one had the intention to vote (T1 and T2) or actually voted (T3) which could be answered as "No" or "I don't know" (0) or as "Yes"

Condition	Time	N	M	SD
No tool	1	35	0.14	.36
	2	36	0.14	.36
	3	33	0.66	.48
VAA	1	35	0.44	.50
	2	36	0.50	.51
	3	33	0.69	.47
CAVAA	1	35	0.18	.39
	2	36	0.36	.49
	3	33	0.67	.48

Note. The data of T3 represents the actual voter turnout.

Conclusion

The goal of the present study, conducted in the unique real-life setting of university council elections, was to obtain insights into what the effects are of VAAs and CAVAAs, as compared to not using a tool, on its user with respect to three variables that have been discussed to be important for democracies: Political interest, political knowledge, and voter turnout. With respect to political knowledge, a differentiation was made between perceived and factual knowledge. Moreover, concerning voter turnout, this variable was measured as

one's intention to vote during the first two measurements (e.g. T1 and T2) and as voter turnout during the last measurement (e.g. T3)

This study has found that one's political interest, perceived and factual political knowledge, and voting intention are higher directly after exposure to university council politics-related stimuli (T2) compared to right before (T1). The increased levels of political interest and perceived political knowledge remain stable over time, as two weeks afterward the levels have neither decreased nor increased (T3). However, one's factual knowledge is solely temporarily enhanced due to exposure to the political stimuli political as it decreases two weeks afterward. With respect to voting intention and voter turnout, the voter turnout (T3) was higher than the reported voting intention directly after exposure (T2)

Moreover, this research shows that usage of voting advice applications has benefits over not using a tool at all. For one, this study demonstrates that the participants that used a VAA have more perceived knowledge and a higher voting intention than those who did not use a tool. Moreover, users of a CAVAA obtain more factual knowledge compared to people that did not use a tool at all. However, there is never a difference between the users of the CAVAA and those who used a VAA.

Besides the several main effects of the tool type, there was also an interaction between time and condition for one of the dependent variables.

To this end, the experiment made clear that using a tool (be it a VAA or a CAVAA) leads to more perceived political knowledge directly after usage compared to right before usage. However, this increased level of perceived knowledge is only stable over time in the CAVAA condition as the participants who used the VAA experienced a decrease in perceived knowledge in the two weeks after usage, which was not the case for the CAVAA users. However, despite the decrease, at T3 the perceived knowledge of the VAA users is still higher than before tool usage. Finally, it can be concluded that people who either used a VAA or a CAVAA obtain a higher perceived political knowledge than those who did not fill out a tool, but the effects remain more stable over time for the CAVAA users compared to the VAA users.

Discussion

First, the effects of time will be discussed, followed by the effects of voting advice applications

The effect of time

As was expected, the levels of all variables measured increased between T1 and T2. The longitudinal results, however, were not always as expected.

For instance, the finding that one's political interest can actually be increased for a longer period of time is in contradiction with what was hypothesized. Based on the works of Prior (2010) and Neuendorf, Smets, and García-Albacete (2013), that found political interest to be quite stable over time, it was reasoned that political interest is a stable construct and that one's interest would therefore solely be enhanced for a short duration as a result of exposure to political stimuli. It might be that there exists a difference in interest stability between national and regional elections, which Prior (2010) and Neuendorf, Smets, and García-Albacete (2013) conducted their research upon, and university council elections. This difference might be related to the perceived importance of the elections, as one could imagine that people would find national elections more important than the council elections of their university. Further research is needed to find out if and why there exists a difference in the stability of interest in university council and national and regional elections.

Furthermore, the found stability of the perceived knowledge is surprising as it was reasoned that repeated exposure to political stimuli would increase the familiarity with the political information (Koriat, 1995) which would also increase the perceived knowledge (Metcalfe, Schwartz & Joaquim, 1993). It was thus expected that the participant's perceptions about their political knowledge would decrease as the participants did not receive any exposure to UC political information from the experiment in the two weeks after T2. Therefore, it might have been the case that the participants repeatedly encountered political stimuli about the university council elections outside of this experiment in the time after T2. This suggestion is quite likely as the elections started right after T2, and both parties, and the university put a lot of effort into promoting the elections and the parties at that time. However, the participants did not encounter so many stimuli between T2 and T3 that their perceptions of their knowledge actually further increased, but rather just as much that they did not experience a decrease compared to T2.

In contrast with the political interest and perceived political knowledge, one's factual knowledge is, as expected, solely temporarily enhanced due to exposure to the political stimuli, as it significantly decreases two weeks afterward. Despite there being a decrease, the level of factual knowledge had not yet fallen back to one's level before exposure to the political stimuli. There, thus, is a discrepancy between factual knowledge and perceived knowledge, which is in support of the findings by Park (2001) and Mondak (1995). However, based on the reasoning that both perceived and factual knowledge are dependent on exposure to political stimuli, one would expect that the factual political knowledge would remain stable over time, as the perceived political knowledge did so as well. The discrepancy found might be due to differences in the type of political information that is being spread at T3 compared to the period before the elections (e.g. around T1 and T2). For instance, at T3 most political information available was about the results of the elections, rather than the standpoints of the two parties, as was the case in the period around T1 and T2. Information about the results of the elections does not add as much factual knowledge as information about the stances of the political parties, and it could therefore be that the content of this stimuli contributed less to the factual knowledge of the participants, resulting in a decrease between T2 and T3. However, due to the fact that information about the results of the elections still is political stimuli, albeit not that insightful, it might have been that this information did keep the enhanced level of perceived political knowledge at T3 on a similar level as at T2.

With respect to voting intention and voter turnout, the voter turnout (T3) was higher than the reported voting intention directly after (T2) and slightly before exposure (T1). This may have been caused by an underreporting of voting intention. This finding is in contrast with prior studies on voting intention, as these found that people tend to overreport their actual intentions (Dolan, Metcalfe & Powdthavee, 2008; Rogers & Aida, 2011). Moreover, the tendency to give socially desirable answers appears to not have occurred, which is in contrast to what had been expected based on the studies by (Andolina et al., 2003; Belli, Moore & VanHoewyk, 2006; Karp & Brockington, 2005). A potential explanation for this finding might be that the participants did not perceive the university council elections to be that much of a sensitive topic that they would lie about their intentions in order to save their face. Nonetheless, it is unclear why the actual voter turnout appeared to be so much higher than the voting intention. On the one hand, it could be that the participants were more attentive to the political stimuli after T2 (e.g. a few days before and during the elections) as their interest in politics had also increased, resulting in a higher voting intention after the first

two measures of the experiment. On the other hand, it is also possible that people did experience a sense to give socially desirable answers at T3, which would indicate that the participants did overreport their actual turnout, which often happens as well according to Karp and Brockington (2005) and Traugott and Katosh (1979). However, what this sudden change in sensitivity to give socially desirable answers would have caused is unclear and is an interesting question to study in future research.

Besides demonstrating a development over time, this study also shows the effects of the usage of voting advice applications, which will be discussed in the following section.

The effect of voting advice applications

The hypothesized higher levels in the tool conditions compared to the no tool condition, were only partially confirmed as will be discussed in the following paragraphs.

When it comes to political interest, the complete absence of differences between the three conditions is surprising as was expected that the CAVAA users would be exposed to more political stimuli and therefore acquire the most political interest, followed by respectively the VAA and the no tool conditions. Moreover, the expectation that VAA users would have a higher interest than no tool users was neither confirmed. As differences over time have been found between T1 and T2, due to the exposure to political stimuli, and as there were no differences found between the conditions, despite there being more political stimuli in the tool conditions compared to the no tool condition, it might be that the specific type of political stimuli played a moderating role and that the three conditions did not differ in the specific type of political stimuli required to enhance one's interest in politics. Future research should be conducted on the potential influence of different types of political stimuli on political interest.

Moreover, regarding perceived knowledge, this research shows that usage of voting advice applications has benefits over not using a tool at all. This study demonstrates that the participants that used a VAA or a CAVAA have more perceived political knowledge than people who did not use a tool. The difference in perceived knowledge might be due to the fact that people encountered more political stimuli when they were in one of the tool conditions compared to people that were in the no tool condition in which political stimuli could solely be obtained via the questionnaire. Again, this would indicate that, as the participants were exposed to more political stimuli, they became more familiar with the political information and therefore experienced more perceived knowledge as compared to the people in the no tool

condition. This finding is in line with the findings of the studies by Fivaz and Nadig (2010), Kamoen and Liebrecht (2022), Kamoen et al. (2015), and Ladner (2012), in which it was also found that the usage of a voting advice application could enhance one's perceived political knowledge. The present research, however, is the first to find valid results as the increase can be compared to both a no tool condition and a pre-measurement. Interestingly, in contrast to what was hypothesized, it does not appear to matter which voting advice application is used when it comes to one's perceived knowledge as a CAVAA does not outperform a VAA in this regard. Moreover, the interaction effect made clear that in both conditions there is an increase from T1 to T2 and that these levels slightly decrease in the two weeks between T2 and T3 but still remain higher than the levels before tool usage. Therefore, usage of a voting advice application is recommended when it comes to increasing one's perceived political knowledge.

However, the increased levels of perceived knowledge at T2 are only stable over time in the CAVAA condition as the participants who used the VAA experienced a decrease in perceived knowledge in the two weeks after usage, which was not the case for the CAVAA users. However, despite the decrease, at T3, the perceived knowledge of the VAA users is still higher than it was before tool usage. The reason why the enhanced perceived political knowledge is stable in the two weeks after T2 for the people that used a CAVAA and not for those who filled out a VAA might partly have to do with the fact that people in the CAVAA condition could have had the feeling that a part of the provided political stimuli was presented to them in a conversation, rather than in a static manner via a VAA. It might be that the conversational elements in the CAVAA, have led to a stronger and livelier presentation of the political information to the users, resulting in a stronger perception of having actually remembered the political information. To illustrate, a CAVAA participant might better remember the moment in which they were exposed to the political stimuli (e.g. the conversation with the chatbot) than the participants in the VAA condition. However, in that case, the participant merely has a better memory of how the stimuli were presented, rather than of the actual stimuli, as the participants in the CAVAA condition do not have a better factual political knowledge than the VAA users, as will be discussed in the following paragraph.

When it comes to factual political knowledge, there is a different pattern observable when it comes to the effects of the usage of a voting advice application compared to what was the case concerning perceived political knowledge. To this end, merely CAVAA usage led to more factual knowledge compared to not using a tool, which indicates that not every voting

advice application is beneficial when it comes to increasing one's factual political knowledge. CAVAA usage leads to more exposure to political information and therefore also to more knowledge compared to not using a tool, which is in line with the hypothesis and the findings by Kamoen and Liebrecht (2022). VAA usage, however, did not lead to more factual knowledge than no tool usage which was not expected. This finding might be the result of the VAA condition not exposing the user to the specific political stimuli required to answer the factual knowledge questions correctly and therefore retaining the user's factual knowledge on a similar level as when no tool had been used. Therefore, CAVAA usage is preferred over VAA and no tool usage when a factual political knowledge increase is desired. This finding once more demonstrates that it is important to consider both forms of knowledge when conducting research on the effects of (CA)VAAs.

The effects found of the voting advice applications on the intention to vote and the actual voting behavior were also partially expected as well as unexpected. The increase in voting intention and voter turnout in the VAA condition compared to the no tool condition was expected as the users of the VAA have had more exposure to political stimuli than those in the no tool group, and previous research had already demonstrated a positive relationship between exposure to political stimuli and one's voting intention and behavior (Durante & Gutierrez, 2012; Freedman, Franz & Goldstein, 2004; Garzia, De Angelis & Pianzola, 2014; Garzia, Trechsel & De Angelis, 2017; Gemenis & Rosema, 2014). Nonetheless, the fact that the people in the CAVAA condition did not obtain a higher intention to vote and did not more often cast a vote compared to the no tool users, is in contradiction with these aforementioned studies and with what was expected, as CAVAA users were expected to show an even higher increase in both matters, compared to the no tool users, than the people in the VAA condition. One possible explanation for the fact that CAVAA usage did not lead to more voting intention and a higher voter turnout than not using a tool might be that the CAVAA provided the user with so much information that they became aware of their own lack of political knowledge. The latter might have caused a feeling of inhibition to vote as the participants felt like their vote would be rather uninformed, leading them to not have the intention to vote and to remain absent from the ballots. The VAA, as a matter of fact, did not provide as many political stimuli as the CAVAA but still more than the no tool condition, which may have prevented the users from becoming aware of their lack of knowledge and experiencing the discussed inhibition. This phenomenon would be an interesting direction for future research as it would reveal why VAAs can have benefits over CAVAAs.

Beforehand it was expected that the users of the voting advice application would outperform the no tool users in every aspect and that CAVAA usage would always lead to higher scores than the usage of a VAA. Now it is clear that users of voting advice applications can indeed outperform no tool users when it comes to increasing political interest, perceived and factual political knowledge, voting intention, and voter turnout but some goals can merely be reached by CAVAA usage, and some by VAA usage, indicating that CAVAAs are not superior to VAAs in every aspect.

Limitations and directions for future research

Due to practical constraints, this paper cannot provide a comprehensive insight into whether or not the participants in the CAVAA condition actually used the options to ask for additional information. This data would have been insightful as it could explain the absence of differences found between the VAA and the CAVAA with respect to the dependent variables analyzed in this study. Future research could implement such a measure by simply asking the CAVAA users about their information-seeking behavior after they have filled out the tool. However, one might expect that such an approach would be susceptible to the social desirability bias, as such a question might be perceived to be sensitive to participants. Therefore, an alternative might be to make use of software that tracks the actions of the users while filling out the tool, which would count the number of clicks on information buttons and the number of questions asked via the open text field.

Moreover, the reader should bear in mind that this study merely considers the context of university council elections and that it is questionable to what extent this setting is generalizable to national and regional elections, in which voting advice applications are mostly used. For instance, the university council elections solely consisted of two parties of which the stances did not differ to a great extent, as opposed to the other two types of elections which usually contain numerous parties with very different opinions. Moreover, this particular election context was focused on a very specific type of people which is different from national and regional elections that consist of voters with a great variety in demographic characteristics. Due to these differences, the generalizability of the results obtained in this study is limited.

Furthermore, this study lacks data on whether, and if so, how many political stimuli the participants have seen between T2 and T3. This is important to know, as all variables

discussed in this study are dependent on the amount of information related to this particular political context that has been seen by the participants. Therefore, it is not entirely clear whether the longitudinal effects are solely the result of the manipulation or also from other factors, such as the increase in likelihood to be exposed to political stimuli in the period slightly after T2. Future research should therefore take into account the degree to which participants have been exposed to political information related to the context of interest. This can either be done by asking questions about the number of stimuli seen in the time after the last measurement at T3 or in the form of asking participants to keep a diary in which they report their daily exposure.

Another limitation of this study is that the results cannot optimally be compared to the results of prior research due to differences in the CAVAAs used. It might have been that the users of the CAVAA in this research were less likely to ask for additional information compared to those in the study of Kamoen and Liebrecht (2022), as the CAVAA in the present study might have unintentionally nudged the participants to the answering buttons rather than the information buttons. This is the case as the answer buttons were placed above the information buttons and because two of the answering buttons were more vivid and bright than the information buttons. This might have influenced the obtained results. Future research should make sure that all buttons consist of the same color and that the information buttons are placed above the answering buttons.

Implications

The effects of VAAs have been studied by many researchers in the past, in contrast to the newly developed CAVAAs. However, this study is one of the first to find valid results when it comes to the effects of both tools. It has done so by applying a methodological design that consists of repeated measures (one pre-experimental and two post-experimental), different types of measurements of the same variable (e.g. perceived and factual knowledge), two different tools, random assignation of participants to groups, and the inclusion of a no tool condition. A first important implication is that this study has demonstrated in an experimental valid manner that usage of voting advice applications does have benefits over not using such a tool. Moreover, this study demonstrates that it is valuable to consider both a no tool condition, as well as to apply repeated measures in the study design. The former allowed for the verification of the actual impact of both tools and the latter made clear that not

all positive findings remain stable over time, which is vital to meeting the requirements of an ideal democracy. Moreover, based on the differences found between perceived and factual knowledge, it is advised to take into account both factual as well as perceived measures for most variables in (CA)VAA studies, as this study once more found that perceived and factual measures do not always show similar patterns. Furthermore, this study makes clear that much is still unknown about the workings and usefulness of voting advice applications and that further research should be directed toward these themes. Finally, for both (CA)VAA developers, as well as for democratic governments worldwide, the findings of this study are valuable to consider as they show that voting advice tools can affect different aspects that are on the basis of an ideal democracy and because they strongly suggest that the design and wording of the tools could matter to a great extent.

References

Aerts, C.M.E., (2022). *Young voters and CAVAAs* [Master's thesis, Tilburg University]. http://arno.uvt.nl/show.cgi?fid=157941

Andersen, K., Ohme, J., Bjarnøe, C., Bordacconi, M. J., Albæk, E., & De Vreese, C. (2020). *Generational Gaps in Political Media Use and Civic Engagement*. Routledge. https://doi.org/10.4324/9781003111498

Andolina, M., Keeter, S., Zukin, C., & Jenkins, K. (2003, June). *A guide to the index of civic and political engagement*. The Center for Information and Research on Civic Learning and Engagement. https://www.researchgate.net/publication/267399505_A_guide_to_the_index_of_civic_and_political_engagement

Andreadis, I., & Chadjipadelis, T. (2011). Voting advice applications and their impact on elections. *61st Political Studies Association Annual Conference "Transforming Politics: New Synergies", 19-21 April 2011. London, UK.*

Andreadis, I., & Wall, M. (2014). The impact of voting advice applications on vote choice. In: D. Garzia & S. Marschall (Eds.), *Matching Voters With Parties and Candidates, Voting Advice Applications in Comparative Perspective*, 115-128. ECPR Press.

Austin, E. J., Deary, I. J., Gibson, G. J., McGregor, M. J., & Dent, J. B. (1998). Individual response spread in self-report scales: Personality correlations and consequences. *Personality and Individual Differences*, *24*(3), 421-438. https://doi.org/10.1016/S0191-8869(97)00175-X

Baka, A., Figgou, L., & Triga, V. (2012). 'Neither agree, nor disagree': a critical analysis of the middle answer category in Voting Advice Applications. *International Journal of Electronic Governance*, *5*(3-4), 244-263.

Belli, R. F., Moore, S. E., & VanHoewyk, J. (2006). An experimental comparison of question forms used to reduce vote overreporting. *Electoral Studies*, *25*(4), 751-759. https://doi.org/10.1016/j.electstud.2006.01.001

Bode, L. (2016). Political news in the news feed: Learning politics from social media. *Mass Communication and Society*, *19*(1), 24-48.

Boogers, M. (2006). *Enquête bezoekers* Stemwijzer. Universiteit van Tilburg. https://docplayer.nl/51714049-Enquete-bezoekers-stemwijzer.html

Catterberg, G., & Moreno, A. (2006). The individual bases of political trust: Trends in new and established democracies. *International Journal of Public Opinion Research*, *18*(1), 31-48.

Chaffee, S. H., & Kanihan, S. F. (1997). Learning about politics from the mass media. *Political Communication*, *14*(4), 421-430. https://doi.org/10.1080/105846097199218

De Graaf, J. (2010). The irresistible rise of Stemwijzer. In: Cedroni, L., Garzia, D. (Eds.), *Voting Advice Applications in Europe: The State of The Art*, 35-46. ScriptaWeb.

Delli Carpini, M. X., & Keeter, S. (1996). *What Americans know about politics and why it matters*. New Haven, CT: Yale University Press.

Dolan, P., Metcalfe, R., & Powdthavee, N. (2008). Electing happiness: does happiness affect voting and do elections affect happiness?. *Discussion Papers in Economics*, *17*. Department of Economics and Related Studies, University of York.

Durante, R., & Gutiérrez, E.M. (2014). Political Advertising and Voting Intention: Evidence from Exogenous Variation in Ads Viewership. *Sciences Po Publications*.

Goldstein, K., & Freedman, P. (2002). Campaign advertising and voter turnout: New evidence for a stimulation effect. *Journal of Politics*, *64*(3), 721-740.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371-395.

Finkel, S. E., & Geer, J. G. (1998). A spot check: Casting doubt on the demobilizing effect of attack advertising. *American Journal of Political Science*, 42(2), 573-595.

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.

Frazer, E., & Macdonald, K. (2003). Sex differences in political knowledge in Britain. *Political Studies*, *51*(1), 67-83.

Freedman, P., Franz, M., & Goldstein, K. (2004). Campaign advertising and democratic citizenship. *American Journal of Political Science*, *48*(4), 723-741.

Furnham, A., & Cheng, H. (2019). Personality traits and socio-demographic variables as predictors of political interest and voting behavior in a British cohort. *Journal of Individual Differences*, *40*(2), 118-125.

Garry, J., Tilley, J., Matthews, N., Mendez, F., & Wheatley, J. (2019). Does receiving advice from Voter Advice Applications (VAAs) affect public opinion in deeply divided societies? Evidence from a field experiment in Northern Ireland. *Party Politics*, *25*(6), 854-861.

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 With Parties and Candidates, Voting Advice Applications in Comparative Perspective*, 105-114. 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.

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

Gidengil, E., Blais, A., Nevitte, N., Nadeau, R. (2004). *Citizens*. Vancouver: UBC Press.

Grönlund, K., & Milner, H. (2006). The determinants of political knowledge in comparative perspective. *Scandinavian Political Studies*, *29*(4), 386-406.

Heinsohn, T., Israel, J., Marschall, S., & Schultze, M. (2016). Online-Wahlhilfen in Wahlkämpfen. *Zeitschrift für Politikwissenschaft*, *26*(3), 253-277.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*(2), 111-127.

Kamoen, N., & Holleman, B. (2017). I don't get it. Response difficulties in answering political attitude statements in Voting Advice Applications. *Survey Research Methods*, *11*(2), 125-140. https://doi.org/10.18148/srm/2017.v11i2.6728

Kamoen, N., Holleman, B., Krouwel, A., Van de Pol, J., & De Vreese, C. (2015). The effect of voting advice applications on political knowledge and vote choice. *Irish Political Studies*, *30*(4), 595-618.

Kamoen, N., & Liebrecht, C. (2022). I Need a CAVAA: How Conversational Agent Voting Advice Applications (CAVAAs) Affect Users' Political Knowledge and Tool Experience. *Frontiers in Artificial Intelligence, 5*. https://doi.org/10.3389/frai.2022.835505

Kamoen, N., McCartan, T., & Liebrecht, C. (2021). Conversational agent voting advice applications: A comparison between a structured, semi-structured, and non-structured chatbot design for communicating with voters about political issues. In: *International Workshop on Chatbot Research and Design* (pp. 160-175). Springer, Cham.

Karp, J. A., & Brockington, D. (2005). Social desirability and response validity: a comparative analysis of overreporting voter turnout in five countries. *The Journal of Politics*, *67*(3), 825-840.

Kiesraad (2021). Verkiezingsuitslagen. Kiesraad. https://www.verkiezingsuitslagen.nl/verkiezingen/index/1

Kleinnijenhuis, J., van de Pol, J., van Hoof, A. M., & Krouwel, A. P. (2019). Genuine effects of vote advice applications on party choice: Filtering out factors that affect both the advice obtained and the vote. *Party Politics*, *25*(3), 291-302.

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General*, *124*(3), 311-333. https://doi.org/10.1037/0096-3445.124.3.311

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*(3), 213-236.

Krouwel, A., Holleman, B., Kamoen, N., Van de Pol, J., De Vreese, C. (2018). Effecten van online stemhulpen. In: Peters, K. (Ed.), *Griffiers Jaarboek*, 19-28. Groningen: iWink.

Ladner, A. (2012). Voting advice applications–impact on voting decisions in the 2011 Swiss national elections. Paper presented 62nd Political Studies Association Annual International Conference, 2012. *Belfast, Ireland*.

Ladner, A., Felder, G., & Fivaz, J. (2010). More than toys? A first assessment of voting advice applications in Switzerland. In: Cedroni, L., Garzia, D. (Eds.), *Voting Advice Applications in Europe: The State of The Art Voting Advice Applications in Europe. The state of the art*, 91-123. ScriptaWeb.

Ladner, A., & Pianzola, J. (2010). Do voting advice applications have an effect on electoral participation and voter turnout? Evidence from the 2007 Swiss Federal Elections. In: *International Conference on Electronic Participation* (pp. 211-224). Springer, Berlin, Heidelberg.

Ladner, A., & Pianzola, J. (2015). Voting advice applications. In: *Encyclopedia of Information Science and Technology, Third Edition* (pp. 6427-6436). IGI Global.

Lecheler, S., & de Vreese, C. H. (2017). News media, knowledge, and political interest: Evidence of a dual role from a field experiment. *Journal of communication*, *67*(4), 545-564.

Lijphart, A. (1997). Unequal participation: Democracy's unresolved dilemma presidential address, American Political Science Association, 1996. *American political science review*, *91*(1), 1-14.

Manavopoulos, V., Triga, V., Marschall, S., & Wurthmann, L. C. (2018). The impact of VAAs on (non-Voting) aspects of political participation: Insights from panel data collected during the 2017 German Federal Elections Campaign. *Statistics, Politics and Policy*, *9*(2), 105-134.

Marschall, S. (2014). Profiling users. In: D. Garzia & S. Marschall (Eds.), *Matching Voters With Parties and Candidates, Voting Advice Applications in Comparative Perspective*, 93-104. ECPR Press.

Marschall, S., & Schmidt, C. K. (2008). Preaching to the converted or making a difference? Mobilizing effects of an internet application at the German general election 2005. In: Farell, D. and Schmitt-Beck, R. (Eds.), *Non-party Actors in Electoral Politics* (pp. 259-278). Nomos Verlagsgesellschaft mbH & Co. KG.

Marschall, S., & Schmidt, C. K. (2010). The impact of voting indicators: The case of the German Wahl-O-Mat. In: Cedroni, L. and Garzia, D. (Eds.), *Voting Advice Applications in Europe: The State of the Art* (pp. 65-90). Napoli: ScriptaWeb

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.

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(4), 851-861. https://doi.org/10.1037//0278-7393.19.4.851

Mondak, J. J. (1995). Newspapers and political awareness. *American Journal of Political Science*, *39*(2), 513-527. https://doi.org/10.2307/2111623

Mondak, J. J., & Anderson, M. R. (2004). The knowledge gap: A reexamination of gender-based differences in political knowledge. *The Journal of Politics*, *66*(2), 492-512.

Neundorf, A., Smets, K., & García-Albacete, G. M. (2013). Homemade citizens: The development of political interest during adolescence and young adulthood. *Acta Politica*, *48*(1), 92-116.

O'Neill, A. (2022, June 21st). *Voter turnout in U.S. presidential elections by age 1964-2020* [Chart]. Statista. https://www.statista.com/statistics/1096299/voter-turnout-presidential-elections-by-age-historical/#statisticContainer.

Ou, K., & Tyson, S. A. (2021). The concreteness of social knowledge and the quality of democratic choice. *Political Science Research and Methods*, 1-18. https://doi.org/10.1017/psrm.2021.65

Park, C. Y. (2001). News media exposure and self-perceived knowledge: The illusion of knowing. *International Journal of Public Opinion Research*, *13*(4), 419-425. https://doi.org/10.1093/ijpor/13.4.419

Pew Research Center. (2007). What Americans know: 1989 – 2007. Public knowledge of current affairs little changed by news and information revolutions. Retrieved from https://www.pewresearch.org/politics/2007/04/15/public-knowledge-of-current-affairs-little-changed-by-news-and-information-revolutions/

Prior, M. (2010). You've either got it or you don't? The stability of political interest over the life cycle. *The Journal of Politics*, *72*(3), 747-766.

Rogers, T., & Aida, M. (2011). Why Bother Asking? The Limited Value of Self-Reported Vote Intention. *John F. Kennedy School of Government Faculty Research Working Paper Series*.

Rome, S. H. (2022). Why Voting Matters. In: *Promote the Vote* (pp. 31-49). Springer, Cham.

Schmeets, H. (2017). Politieke betrokkenheid in Nederland. *Statistische Trends*, (december).

Schultze, M. (2014). Effects of voting advice applications (VAAs) on political knowledge about party positions. *Policy & Internet*, *6*(1), 46-68.

Smets, K., & Van Ham, C. (2013). The embarrassment of riches? A meta-analysis of individual-level research on voter turnout. *Electoral Studies*, *32*(2), 344-359.

Strömbäck, J., Djerf-Pierre, M., & Shehata, A. (2013). The dynamics of political interest and news media consumption: A longitudinal perspective. *International Journal of Public Opinion Research*, *25*(4), 414-435.

Strömbäck, J., & Shehata, A. (2010). Media malaise or a virtuous circle? Exploring the causal relationships between news media exposure, political news attention and political interest. *European Journal of Political Research*, *49*(5), 575-597.

Traugott, M. W., & Katosh, J. P. (1979). Response validity in surveys of voting behavior. *Public Opinion Quarterly*, *43*(3), 359-377.

Tsai, W. H. S., Liu, Y., & Chuan, C. H. (2021). How chatbots' social presence communication enhances consumer engagement: the mediating role of parasocial interaction and dialogue. *Journal of Research in Interactive Marketing*, *15*(3), 460-482.

Tyson, S. A. (2016). Information acquisition, strategic voting, and improving the quality of democratic choice. *The Journal of Politics*, *78*(4), 1016-1031.

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: D. Garzia, & S. Marschall (Eds.), *Matching Voters with Parties and Candidates: Voting Advice Applications in Comparative Perspective* (pp. 11-31). ECPR Press.

Van de Pol, J. (2016). *Voting wiser: The effect of Voting Advice Applications on Political Understanding*. Universiteit van Amsterdam.

Verba, S. (1996). The citizen as respondent: sample surveys and American democracy presidential address, American Political Science Association, 1995. *American Political Science Review*, *90*(1), 1-7.

Westle, B., Begemann, C., & Rütter, A. (2014). The "Wahl-O-Mat" in the course of the German Federal Election 2013–Effects of a German VAA on users' election-relevant political knowledge. *Zeitschrift für Politikwissenschaft*, *24*(4), 389-426.

Zadrozny, W., Budzikowska, M., Chai, J., Kambhatla, N., Levesque, S., & Nicolov, N. (2000). Natural language dialogue for personalized interaction. *Communications of the ACM*, *43*(8), 116-120.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162-185.

Appendix

Appendix A – Video of the CAVAA in action

Link to the video in which the CAVAA is used:

https://surfdrive.surf.nl/files/index.php/s/fTSFbP1KhYQVTr0?path=%2FVideo%27s%20CAVAA#/

Appendix B – pre-test procedure and additional results

Around one and a half weeks before the start of the elections, all three conditions were pre-tested several times. During these tests, the researcher would let other people of the target group use the materials and ask the participants to think out loud at every unique element of that specific part of the material. They were instructed to have a critical attitude and to not be afraid to give negative feedback. Then, their feedback was written down by the researcher. In this manner, the no tool condition was tested two times, the VAA condition four times and the CAVAA condition was pre-tested thirteen times. The feedback provided some useful results and especially for the chatbot function there was quite some room left for improvement. As the feedback for the no tool condition mostly consisted of grammatical aspects, solely pre-test results of the VAA and CAVAA conditions will be discussed below.

VAA

There was not very much feedback on the VAA during the pre-tests other than several grammatical errors and the fact that it was unclear that the participants should answer whether or not the issue had a high priority for them rather than whether or not they agreed with the statement. The latter had been improved by also adding the question "*Does this issue have a high priority for you?*" right underneath the statement. See Appendix B for the procedure and more information about the pre-tests.

CAVAA

The pre-tests have given useful insights with respect to the CAVAA. The most important point of feedback was the fact that multiple participants indicated that they, as was the case with the VAA, kept forgetting that they had to answer whether or not the issue had a high priority for them instead of indicating whether or not they agreed with the statement. This was the case as the question "*Does this issue have a high priority for you?*" was merely asked a few times and when it was, it was done before the statement popped up. The participants, thus, read the statement just before answering it, which instinctively made them indicate whether or not they agreed with it. Therefore, for the final version, this question was asked always directly after the statement and placed right above the answer buttons. Furthermore, it was tested to locate the chatbot inside the Qualtrics survey in order to prevent participants from having to switch between tab pages and consequently running the risk of closing one of the pages by accident. However, there appeared to be several disadvantages with this approach, one of them being the limited visibility of the chatbot due to a small

frame. Therefore, it had been decided to let the participants use the CAVAA on another webpage than the survey, similar to the research by Kamoen and Liebrecht (2022)

Moreover, quite some text that was present in the pre-test has been removed for the final version, as the test participants indicated that the amount of text distracted them and that many sentences were often perceived as unnecessary. Furthermore, the pre-tests made clear that there were still some mistakes remaining in the content of the bot, such as that for certain statements, the information buttons provided incorrect information as well as texts that contained several grammatical errors. Finally, the pre-test had made clear that the phrase "*it is top priority for party x that…",* which was visible after clicking on the button "*Standpoints SAM and Front",* was confusing as the participants interpreted this as if that specific issue was the only top priority of that specific party. This became confusing later on in the voting advice application as by then it had been mentioned several times that a certain issue was of top priority for that specific party. The wording of this phrase has therefore been altered to "*For party x it is one of their top priorities that …*", as this does imply that there are multiple top priorities for one party.

Appendix C – The factual political knowledge questions

Please look at the following statements and indicate whether they are true or not. If you do not know the answer, please choose the option "I don't know":

T1

- Amnesty International's #Let'sTalkAboutYes manifesto is an initiative that aims to improve the democratic systems within universities all over the world.
- An Edubadge is a digital badge that can be shared online, e.g., on LinkedIn. The badge is an accredited proof of skills mastered (such as computer skills).
- Currently, it is not possible at Tilburg University to make an exam on distance when this person is on exchange.
- Currently, Tilburg University has an online health checker, which is an app that allows students to monitor their own health.
- When it comes to increasing the amount of student working spots, party Front is particularly in favor of creating more outdoor working spots.
- Party Front is against hybrid teaching.
- The introduction of course bundles is one of the top priorities for party SAM.

T2

- Soft skills are (among others) attributes and personality traits that help students interact with others and succeed in the workplace.
- The Grade Point Average (GPA) is the mean grade of the grades obtained by all students during a specific exam of a specific course.
- Currently, there is an Indoor Campus Map for Tilburg University.
- Currently, the honors program is not available for everyone as students often need a certain average grade to be able to participate.
- For party SAM it is one of their top priorities that courses will be evaluated via Evalytics not only at the end of the block or semester but also during the course.
- For party SAM it is one of their top priorities that student associations should be offered a reduced price if they want to hire a lecture room through Conference and Events Support at the university.
- For Party Front it is one of their top priorities that the university is going to rely more frequently on student assistants in education.

T3

- The Career Portal is an online environment in which only job offers at Tilburg University are listed.
- Blended learning consists of face-to-face learning in the classroom as well as online learning and the content of both types of learning is exactly the same.
- Currently, Tilburg University does not make use of Edubadges.
- Currently, Tilburg University has peer-to-peer support groups.
- For party Front, it is one of their top priorities that student associations will be offered a reduced price if they want to hire a lecture room through Conference and Events Support at the university.
- For party SAM, it is one of their top priorities that the pilot for free menstrual products at the university will be expanded.
- For party SAM, it is one of their top priorities that in each Master program, it will be possible for students to follow a course bundle outside their program that will be labelled as an additional expertise on the diploma.

Appendix D - The correlations between political interest, perceived political knowledge, and voting intention

Correlations between perceived political knowledge, political interest, voting intention, and voter turnout (e.g. T3), during every measurement.

	Perceived pol. know	Political interest	Voter intention/turnout
T1			
Perceived pol. know	-	.11	-.13
Political interest	-	-	-.02
Voter intention/turnout	-	-	-
T2			
Perceived pol. know	-	.39*	-.08
Political interest	-	-	-.24*
Voter intention/turnout	-	-	-
T3			
Perceived pol. know	-	.60*	-.20*
Political interest	-	-	-.28*
Voter intention/turnout	-	-	-

*Note. Significant interactions contain a **