

I don't get it. Response difficulties in answering political attitude statements in Voting Advice Applications.

Naomi Kamoen
Utrecht University
and Tilburg University
The Netherlands

Bregje Holleman
Utrecht University
The Netherlands

Voting Advice Applications are online tools that provide users with a voting advice based on their answers to a set of political attitude questions. This study investigated to what extent VAA users understand the questions that lead to the voting advice, and what search and response behaviour they expose in case of comprehension difficulties. Two studies were conducted to investigate these issues: a cognitive interviewing study among 60 VAA users during the Dutch municipal elections in the city of Utrecht, and a statistical analysis of all answers provided by 357,858 users who accessed one of the 34 municipal VAAs during these same elections. Results of the two studies show a coherent picture: difficult concepts (e.g., tax names or municipal jargon), geographical locations (e.g., reference to a specific street), and vague quantifying terms (e.g., "more") all complicate the question. In case of comprehension difficulties, Study 1 shows that VAA users make little effort to solve their problems, for example by looking up difficult terms on the Internet. Instead, they draw inferences about what the question might mean and proceed to answer nonetheless. These are often neutral or no opinion answers, which seems to suggest that the meanings of those options are confounded. In Study 2, however, we found that the choice for either a neutral or no opinion response is not accidental: semantic meaning problems often result in no opinion answers, whereas pragmatic problems are related to neutral responses. We discuss the implications of these findings for survey theory and practice.

Keywords: voting advice applications; political attitudes; comprehension process; big data

1 Introduction

Political knowledge and feelings of political competence are essential for political participation (Delli Carpini & Keeter, 1996). Voting Advice Applications (VAAs) aim to contribute to that. These VAAs are online survey tools targeted at increasing voters' understanding of their own position in the political landscape and those of the parties running in the elections. In a VAA, users respond to a set of attitude statements about political issues, such as "Wearing a burka should be forbidden." (*Kieskompas*, Dutch national elections 2012). Based on a comparison between the users' answers and the parties' issue positions, the VAA subsequently provides a voting advice (De Graaf, 2010). By supplying political information in this cost-efficient way, VAA developers aim to increase voters' political knowledge, as well as their interest in political matters (Garzia, 2010).

VAA surveys have become omnipresent elements in today's elections worldwide. In the Netherlands, for example, where they find their origin, up to half of the electorate

filled out a VAA during the 2012 national elections. Also in other European countries VAA usage is consistently on the rise (see Marschall, 2014), and recently, VAAs have been introduced outside of Europe during national elections in the USA, Israel and Egypt.

Although VAA surveys are a hallmark of contemporary elections, they are also a relatively young phenomenon. The first wave of VAA studies has focused primarily on their political impact, showing that VAAs increase people's political knowledge (e.g. Kamoen, Holleman, Krouwel, Van de Pol, & De Vreese, 2015), elevate the turnout at the ballots (e.g. Garzia, De Angelis, & Pianzola, 2014), and affect the content of the vote cast (e.g. Andreadis & Wall, 2014; Wall, Krouwel, & Vitiello, 2014). Unaddressed in these studies, however, is how well VAA users understand the political statements that should enlighten them. The current research focuses on this topic, by investigating which question characteristics cause comprehension problems, and what kind of search and response behaviour users expose when they experience such difficulties.

While novel within VAA research, the topic of question comprehension has already received considerable attention within the domain of survey methodology. Based on practical experience and decennia of scientific research, a clear list

Contact information: Naomi Kamoen, Tilburg University, PO Box 90153, 5000 LE Tilburg (email: n.kamoen@uvt.nl)

of guidelines has been formulated describing which question characteristics facilitate or complicate comprehension (e.g. Dillman, Smith, & Christian, 2009; Fowler Jr & Cosenza, 2008; Krosnick & Presser, 2010). An unanswered question, however, is how this knowledge generalizes to the specific context of VAA surveys, as VAA users are assumed to be highly motivated to answer the questions in search of a valid voting advice (Holleman, Kamoen, & De Vreese, 2013). Therefore, it is relevant to investigate how survey advice generalizes to this specific language in use context. In turn, an investigation of VAA surveys may also bring to light new question characteristics that lead to comprehension difficulties in the broader domain of political attitude surveys.

In order to examine question comprehension and response behaviour in VAA surveys, two studies have been conducted. The first is a cognitive interviewing study (Willis, 2005), in which VAA users ($N = 60$) are asked to think aloud while responding to VAA statements. Survey researchers often apply cognitive interviewing to pre-test individual surveys among small samples of respondents, rather than to theorize about comprehension problems (Schwarz, 2007). By applying cognitive interviewing on a relatively large scale, we aim to theorize about what kind of question characteristics cause comprehension problems in a specific political attitude survey context. Study 2, in turn, investigates to what extent the question characteristics leading to response problems in Study 1, indeed result in specific response behaviour (i.e., neutral and no opinion responding) in real-life and large-scale VAA datasets.

By applying this novel combination of research methodologies, the current research provides a broad insight into question characteristics that complicate the comprehension of political attitude statements.

2 A cognitive perspective on question comprehension

The “Tourangeau model” (Tourangeau & Rasinski, 1988; Tourangeau, Rips, & Rasinski, 2000) is often used as a starting point for theorizing about comprehension problems. According to this model, answering questions is a cognitive process involving four steps. During the first step, a respondent should comprehend the question and determine what kind of opinion or attitude is asked for. Second, the respondent has to retrieve attitudinal information from long-term memory. Some respondents are able to retrieve a summary evaluation of their beliefs directly (e.g., *Smoking is disgusting and therefore it should be forbidden in playgrounds*). Others, however, will need a third step of weighting and scaling individual beliefs (*I like to smoke when I am outside with my child versus smoking is bad for children’s health*) in order to reach a judgment. In the fourth and last stage, the respondent fits the judgment made to the response options. Respondents may adapt their answers during this process, for example for reasons of social desirability (see Groves et al.,

2009; Tourangeau et al., 2000).

When assessing the link between question characteristics and comprehension difficulties, we should zoom into the process of question comprehension. Following theories from discourse studies (Graesser, Singer, & Trabasso, 1994; Zwaan & Radvansky, 1998), respondents first construct a semantic representation of the literal meaning of the question. This is what Rips (1995) refers to as the representation of the question. Next, respondents enrich this representation with their world-knowledge, resulting in a pragmatic representation about the question.

Survey handbooks generally focus on how question characteristics can facilitate building a semantic representation of the question. Among the most well-known recommendations (e.g., see Dillman et al., 2009; Fowler Jr & Cosenza, 2008; Krosnick & Presser, 2010), are the following: use simple and familiar words rather than technical terms and jargon; avoid negations; avoid double-barreled questions; ask questions about concrete issues rather than about vague underlying values and norms; avoid vague quantifying terms such as *often* and *more*.

The reason to avoid such questions is that they leave room for different interpretations. For example, what does a no-answer to a double-barreled question like *To decrease the number of traffic jams, the roads should be broadened* mean? That traffic jams are not a problem, or that broadening roads is not a good way to decrease traffic jams? This uncertainty may not only be problematic for individual respondents, but also for the survey researcher, as it will lead to a non-uniform question interpretation. This, in turn, complicates summarizing responses across respondents.

Like regular attitude surveys, VAAs also require question formulations that give room to only one interpretation, but this is for a different reason. VAA developers do not aim to describe the attitudes of the population of VAA users, but rather, to provide a valid voting advice to each individual user. In calculating this voting advice, they draw up on a Downsian model of voting (Downs, 1957; for limitations of this perspective and other models of voting see Mendez, 2012), as the voting advice is based on the number of matches between the user’s answers, and the parties’ issue positions (Krouwel, Vitiello, & Wall, 2012). Hence, when users have problems constructing a semantic representation of the question, or when they construct an interpretation that differs from the interpretation of political parties, the answers are an invalid basis for the voting advice.

In light of the above, it is surprising that VAA developers regularly violate survey guidelines. A corpus research by Van Camp, Lefevere, and Walgrave (2014) shows that one in every five VAA questions is double-barreled and that about the same proportion of statements includes a vague quantifying term. The study by Van Camp et al. (2014) did not analyze the occurrence of negations or political jargon, but there

is initial evidence that VAA questions regularly include these elements too (Van den Ven, 2014). There are various reasons why VAA developers violate survey guidelines in constructing their questions. For one, VAA statements are the product of a deliberative process involving not only the VAA developers, but also all parties running in a certain election. This leads to difficulties in balancing the needs of these different stakeholders and of recipient design. For example, sometimes wording choices are made simply because one of the stakeholders prefers a certain wording. In addition, VAA developers often add reasons and conditions to VAA questions because they consider this to be necessary for differentiating between political parties. For example, many political parties will agree with the statement that “Money should be spent on environmental issues”, but less parties will agree when a conditional clause is added “even if this means cutting expenses on culture”. Seen from a survey methodological perspective, i.e. from a respondent’s perspective, such choices are expected to lead to comprehension difficulties. Overall, we may therefore expect that VAA users frequently experience comprehension problems.

In addition to an investigation of the semantic meaning problems, the current research will also explore to what extent pragmatic meaning problems arise. Once a VAA user has formed a semantic representation-of the question, this representation is enriched with world knowledge related to the concepts in the question. For example, when responding to the statement *A new hospital should be built in Amsterdam*, VAA users may be perfectly well able to construct a semantic representation-of the question, because they comprehend all individual words and understand how they relate to one another. Nevertheless, there may be a comprehension problem, because users are unable to connect this question to their world knowledge: why should or shouldn’t there be a new hospital? In the current research, we will explore how frequently VAA users experience such a lack of knowledge about the issue at stake, and whether certain question characteristics trigger pragmatic meaning problems.

3 What do people do when they do not understand a question?

After a respondent has constructed a representation-of and representation-about the question, attitudinal information has to be retrieved from memory and a judgment has to be formed (Tourangeau et al., 2000). Comprehension problems during the first stage of question answering often disperse to these later stages. For example, if a respondent does not know the term “dog tax”, the processes of attitude retrieval and judgment formation are also disrupted, as it is obviously quite difficult, if not impossible, to retrieve information from memory and form a judgment about an unknown concept (Fowler Jr & Cosenza, 2008). In such cases, a clarification can improve retrieval, and therefore, the accuracy

of respondents’ answers (Schober & Conrad, 1997; Schober, Conrad, & Fricker, 2004).

In online surveys, there are various ways to provide respondents with additional information. For example, internal or external links may provide users with definitions of relevant concepts (Galesic, Tourangeau, Couper, & Conrad, 2008). Moreover, respondents can consult external information sources themselves, by opening Google or another search engine to look up the meaning of a difficult term.

Even though various options are often available, survey respondents rarely use these to solve their comprehension problems. For example, an eye-tracking study by Galesic et al. (2008) showed that two-thirds of the respondents in an online survey did not look at the definitions that were accessible when “rolling over” a concept. This study also showed that the more effort necessary for accessing the definition, the less likely respondents were to inspect the available information. This limited effort matches Krosnick’s idea (1991) that survey respondents often expose satisficing behaviour, which means they spend just enough effort to provide a plausible answer that satisfies the survey researcher. Hence, rather than looking up what an unfamiliar term means, survey respondents make an assumption about its meaning, or they will just provide a neutral or don’t know answer as an easy-way out.

As VAAs are online surveys, VAA users always have the opportunity to search the web for information. In addition, some VAAs also offer information options within the tool: *Stemwijzer*, one of the leading VAA brands, for example, provides information about the parties’ stances towards each of the political issues in the VAA. While respondents in regular attitude surveys rarely make use of such options, we may expect VAA users to consult this information quite frequently when they experience a comprehension problem, because VAA users have a personal gain in providing well-considered answers: obtaining a valid voting advice.

4 Reporting opinions on difficult questions

After respondents have retrieved information and formed a judgment about the attitude object in the question, they will have to translate this opinion to a given set of response options. Comparable to regular attitude surveys, VAAs generally use agree-disagree scales supplemented with a non-response option.

Studies in survey research have shown that these agree-disagree scales are often misused in practice. For example, several studies demonstrated that if respondents answer questions on fictitious issues about which they cannot possibly have an opinion, about 30% of the respondents picks a substantive answer rather than giving a non-substantive response (Bishop, Oldendick, & Tuchfarber, 1983, 1986). In addition, while the neutral middle response option should be chosen solely to express feelings of ambivalence, research shows

that this option often functions as a “hidden don’t know” (e.g. Nadler, Weston, & Voyles, 2015; Sturgis, Roberts, & Smith, 2014).

In a regular attitude survey context, the misuse of the agree-disagree scale primarily has consequences for the survey researcher, as it leads to biased conclusions about population means and correlations between measurements (Sturgis et al., 2014). In a VAA survey, however, there are also direct consequences for the user, as misuse results in a biased voting advice. This is because in the calculation of the voting advice, non-response answers are excluded whereas neutral answers are taken into account (Krouwel et al., 2012). Hence, when a “no opinion” answer is provided, the voting advice is calculated based on, for example, 29 rather than 30 questions. Neutral answers, on the other hand, are taken into account: for these answers a match is calculated between the VAA user and each of the political parties. This implies that if VAA users pick the middle response category to communicate that the question is not fully understood, the voting advice is biased because the overlap between parties that gave a middle response is unintendedly high.

5 Research question and hypotheses

Two studies will be conducted to investigate how VAA users answer political attitude questions, and what search and response behaviour they expose in case of comprehension problems. Based on the literature, we expect VAA users to encounter semantic meaning problems related to political jargon, negation, double-barreledness, and vague quantifying terms. In addition, we will explore how frequently VAA users experience pragmatic meaning problems. As VAA users answer the VAA statements for the benefit of receiving a voting advice, we expect them to consult information services within and outside the tool in case of comprehension difficulties. Finally, we will explore what answers they provide in such cases.

6 Method Study 1

6.1 Design

Study 1 is a cognitive interviewing study. Cognitive interviewing is an observational method in which respondents verbalize their thoughts while filling out a survey (Campanelli, 2008). Within survey methodology, this method is primarily used to pre-test individual surveys among small samples of 10-12 respondents (Collins 2002, cf. Campanelli, 2008). In the current research, we applied cognitive interviewing on a larger scale: 60 participants verbalized their thoughts while filling out one of the two leading VAAs during the Dutch municipal elections of 2014 in the city of Utrecht ($N=20$ for *Stemwijzer* and $N=40$ for *Kieskompas*). These verbalizations were subsequently scored for comprehension problems. By applying cognitive interviewing on

a large scale, we aim to theorize about the types of questions that lead to comprehension problems, rather than pinpointing specific questions that cause comprehension problems in the context of an individual VAA. Within writing research and cognitive psychology, cognitive interviewing is frequently used for this purpose of theory building (e.g. Flower & Hayes, 1981; Van Weijen, Van den Bergh, Rijlaarsdam, & Sanders, 2009).

6.2 Participants

Purposive sampling was used to select our sample: we specifically recruited participants who were residents of the municipality of Utrecht, and who were already planning to fill out a VAA for the municipal elections. The average age of participants was 28.7 ($SD=8.21$). A total of 55 participants was enrolled in or had completed a Bachelor or Master program, while the remaining 5 participants were enrolled in or had completed a medium vocational education. As VAA users are usually rather young and highly educated (Marschall, 2014; Van de Pol, Holleman, Kamoen, Krouwel, & De Vreese, 2014), our sample roughly resembles the typical VAA user in terms of age and educational level. Our sample consisted of 55% females, which is slightly more than usually found in VAAs.

6.3 Procedure

Each session took place individually and in a quiet environment. After the respondent had been thanked for participating, the experimenter asked the participant to fill out the VAA for the municipal elections while verbalizing his/her thoughts. Prior to starting the vote test, participants practiced with this procedure. In case the participant fell silent during this session, the experimenter encouraged the participant to keep verbalizing (e.g., “What are you thinking right now?” cf. Boren & Ramey, 2000).

6.4 Codings

Two types of verbalizations were scored as indicating comprehension problems.¹ First, these are verbalizations explicitly pointing to a comprehension problem, such as “I don’t know what the OZB [a type of tax on housing] is”. Second, these are cases in which a participant’s line of reasoning did not match with the answer provided. For example, one of the questions in *Kieskompas* read: “There should be no widening of the A27 highway near Amelisweerd” and a participant verbalized “Yes, yes, yes. The accessibility of the city is always important, but not very important, so I’ll pick a regular ‘agree’”.

¹As we did not count facial expressions as signs of comprehension problems, the current study is likely to give an underestimation of the actual number of comprehension problems in VAAs.

When a comprehension problem was observed, we also scored the type of comprehension problem. Based on survey handbooks (e.g. Dillman et al., 2009; Krosnick & Presser, 2010), we distinguish between *unknown terms/jargon*, *double-barreledness*, and *negation*. As we view vague quantifying terms to be a specific kind of vagueness, we decided to score these comprehension problems as one type: *vagueness*.

In addition to these categories taken from the survey literature, we included several categories that were derived inductively. As VAA users often made a remark indicating they did not know a location mentioned in the question (“I don’t know where Polder Rijnenburg [a part of Utrecht] is”), we distinguished a category of *unknown locations*. Furthermore, we included a category with *additional remarks about the question wording*. This residual category includes all critical remarks about the question wording that do not fit into another category, for example, the question is “too strong”. Finally, we coded *pragmatic meaning problems*. This category includes all comprehension problems arising because the respondent does not understand the semantic meaning of the question, but has too little information about the context of the issue to answer the question. The coding scheme can be found in Appendix B.

For coding the verbalizations ($N = 1786$)², the following procedure was used. The first author developed a coding scheme, and coded a subset ($N = 200$) of all verbalizations. Subsequently, the second author independently coded the same subset of verbalizations, using the coding scheme constructed by the first author. The Kappa’s (Cohen, 1977, 1988) were then compared, both for the agreement about whether there is a comprehension problem to begin with, and for the type of problem. These comparisons revealed that the intercoder agreement about whether there is a comprehension problem was very high, whereas there was quite some disagreement about the type of problem. These differences were discussed, and based on these discussions, the original coding scheme was further specified.

Subsequently, the first author coded all verbalizations based on the revised coding scheme. To check the intercoder agreement, a third coder (research assistant) received a training in coding verbalizations and independently coded a random selection of 50% of the cases in which users experienced a comprehension problem ($N = 180$). The choice to code only instances where the first coder signalled a comprehension problem was made, because in the first round of coding there was much agreement about which utterances indicated a problem.

For the second round, we compared the codings of the research assistant to the codings of the first author. Only in 1 instance, the third coder thought that an utterance that had been selected by the first coder lacked a comprehension problem, which indicates that our coding strategy worked well

Table 1

Kappa and Kappa/Kappa max for each problem category (based on $N = 180$)

	Kappa	$\frac{\text{Kappa}}{\text{Kappa}_{\text{Max}}}$
Unknown Concept/Jargon	0.84	0.94
Unknown Location	0.96	0.98
Vagueness	0.79	0.86
Negation	0.56	0.58
Other remarks about quest. wording	0.53	0.66
Pragmatic meaning problem	0.65	0.96

No inter-rater reliability could be calculated for the category of double-barreledness, as no utterances related to this category appeared in the test set.

and the coders agree about which utterances indicate a comprehension problem. Table 1 shows the Kappa divided by the maximal Kappa for each problem category. As can be read, it was substantial to moderate for the categories *other remarks about the question wording* and *negation*. For all other categories, it was excellent judging by the standards of Landis and Koch (1977).

6.5 Results

Comprehension problems in VAAs. In 361 out of the 1786 valid respondent and item combinations (20.2%), there was a clear sign of a comprehension problem. Each respondent encountered between 0 and 14 comprehension problems across the 30 questions ($M = 6.01$; $SD = 2.98$). The respondent who encountered most comprehension problems ($N_{\text{problems}} = 14$) was a regular participant in terms of demographic characteristics: a 32-year-old male who had finished a higher vocational education. As for the item perspective, each question showed between 0 and 29 problems across all participants ($M = 6.01$; $SD = 5.49$). The item for which most comprehension problems were observed ($N_{\text{problems}} = 29$), was about whether there should be new houses built in the polder Rijnenburg. Many problems arose for this items, because “polder Rijnenburg” was an unknown location. Overall, this exploratory analysis reveals that both respondent and item characteristics are important for the number of comprehension problems.

Table 2 shows the most common sources of comprehension difficulties across all respondents and items. As can be read, the most common semantic meaning problem is a lack of knowledge about a concept in the question. Concepts that often lead to comprehension problems are often names of taxes (e.g., dog tax, tax on waste, and OZB [a form of tax

² Each participant answered 30 questions, resulting in 1800 verbalizations. Of these, 14 verbalizations had to be discarded due to the quality of the recording.

Table 2
The types of comprehension problems VAA users experience

	Times scored as a problem	Percentage relative to the total number of problems (in %)
<i>Pragmatic meaning problem</i>		
Too little information about the topic	134	35.7
<i>Semantic meaning problem</i>		
Unknown Concept/ Jargon	82	21.9
Unknown Location	66	17.6
Vagueness	38	10.1
Negation	10	2.7
Double-barreledness	6	1.6
Other remarks about quest. wording	35	9.3
<i>Others</i>		
Unknown comprehension problem ^a	4	1.1
Total	375	

In 14 instances, there was more than one problem for one and the same question. That is why the numbers in Table 1 add up to 375 rather than 361.

^aIn some instances, users clearly experienced a comprehension problem, but from their wordings, it was not clear what the exact cause for the comprehension problem was. For example, one respondent said: "I don't understand this statement ... (reads statement again)". These verbalizations were scored as "other/unknown comprehension problems".

on houses]), and political jargon (e.g., welfare committee, liveability budget and welfare work).

Semantic comprehension problems also arose when a geographical location was mentioned in the question. Several questions included a reference to a specific location within the municipality of Utrecht, such as "There should be new houses built in the Polder Rijnenburg [a specific area in Utrecht]" or "All prostitution should be abolished from the Hardebollenstraat [a specific street in Utrecht]". Many participants experienced difficulties answering such questions: "Hmhmhm. I don't know that area".

Furthermore, about 10% of the comprehension difficulties relates to a concept in the question being vague: "social facilities ... like what? Depends on the specific type of social facilities ... that is too vague a concept. I'll pick neutral.". Of the comprehension problems in this category, only one related to the usage of a vague quantifying term, and all other utterances related to other terms or concepts being vague (e.g., hard approach, social facilities, decide for themselves).

While *political jargon*, *geographical locations*, and *vagueness* often lead to comprehension problems, the other question characteristics mentioned in the survey literature received very little remarks: only six comments were made about the question being double-barreled, and ten about negations. Interestingly, problems related to the usage of

negation often remained unnoticed for the VAA users themselves; the majority of the problems in this category was users picking an answer that did not match their line of reasoning (see *Codings*).

In addition to semantic meaning problems, a relatively large share of the total number of comprehension problems related to VAA users' inability to relate the question to relevant world knowledge. The most common pragmatic problem was a lack of information about the status quo of a political issue. Such problems were often triggered by a vague quantifying term in the question (e.g., *decrease*, *raise* and *extra*). For example, in *Kieskompas* participants were asked to indicate whether a form of taxes on housing should be increased. This frequently led to verbalizations like: "Well, I don't know how high that tax currently is". Hence, while vague quantifying terms rarely lead to semantic meaning problems, they do trigger a certain state of affairs (the current status of the height of a tax) and the users' lack of knowledge thereof.

What do people do when they do not understand a statement?. *Stemwijzer*, one of the VAAs of interest in Study 1, offers a clickable menu called "Election Board", which provides information about the parties' issue positions towards the attitude statements. For the 600 respondent and items combinations related to *Stemwijzer*, the Election board option was accessed only once. In addition to that, across all

of the 1800 trials related to both *Stemwijzer* and *Kieskompas*, the Internet was consulted 26 times for finding information. These 26 instances were strongly clustered within specific users, as one person consulted the internet for 9 questions, another respondent used it 3 times, and five more respondents each consulted the internet 2 times.

Instead of consulting additional resources, we observed that, in case of comprehension problems, VAA users often proceeded by making inferences about what a concept in the question might mean. This can be illustrated by various readings of the term “welfare work”: “Welfare work ... is that voluntary work? It probably is”, “Welfare work ... is that some form of care?”, and “What is welfare work ... is that good for one’s well-being?”. Overall, these findings suggest that rather than looking up information, VAA users make assumptions about what a term or a question might mean and provide an answer.

What kinds of answers do respondents provide?. Table 3 shows the answers VAA respondent provided in cases they did and did not experience comprehension problems. Analysis of these answers indicates that whereas substantive answers are chosen disproportionately more often in case VAA users do not experience comprehension problems, neutral and no opinion answers are provided more often when VAA users do experience comprehension problems ($\chi^2 = 392.23$; $df = 2$; $p < 0.001$). This seems to indicate that both the middle and the no opinion category are used to express comprehension difficulties and may therefore be used as a proxy to signal such problems.

While VAA users with comprehension problems choose neutral and no opinion answers most of the times, still in a little over 40% of the cases a substantive and directional answer is provided. Hence, participants quite frequently provide substantive answers while they do not fully understand the question. These substantive answers are sometimes the result of respondents making inferences about what a concept might mean: “I don’t know where that area [polder Rijnenburg] is, but “the polder” sounds as if there are often floods there ... , so it doesn’t sound as if they should build houses there.” Substantive answer also seem to be a result of satisficing behaviour on another occasion: “I don’t know the area ... whatever. Agree.”.

Conclusion Study 1. In a cognitive interviewing study, we analyzed the type of comprehension problems VAA users encounter while filling out a VAA. Results show that users encounter comprehension problems for, on average, 1 in every 5 questions. About two-thirds of these relate to the semantic question meaning, covering difficulties with *political jargon*, *tax names*, or *geographical locations*. In addition, one-third of the problems relate to the pragmatic representation-about the question. Such problems are often triggered by vague quantifying terms. Hence, while in the survey literature it is often mentioned that vague quantifiers

are difficult simply because they refer to vague quantities, we can add that in a political attitude survey context these terms are difficult because they trigger a lack knowledge about the current state of affairs. In case of comprehension problems, VAA users often assume a certain question meaning, rather than looking for information on the web. Next, they supply a neutral or no opinion answer. Hence, the response behaviour VAA users expose in case of comprehension difficulties appears to be comparable to what is observed in regular attitude surveys (compare the limited effort in looking up unknown concepts reported in Galesic et al., 2008 or the response behaviour in Nadler et al., 2015, and Sturgis et al., 2014). Consequently, we can say that neutral and no opinion answering can function as proxies for detecting comprehension problems.

To examine if these patterns generalize to an externally and ecologically valid setting, we present a second study in which we predicted response patterns (neutral answering and no opinion answering) from a set of question characteristics in a dataset comprising the answers of all users accessing 1 of the 34 *Kieskompas* VAAs during the Dutch 2014 Municipality Elections. If question characteristics like *tax names* or *political jargon* indeed lead to comprehension problems, we expect these characteristics to be highly correlated with neutral and no opinion answering.

7 Method Study 2

7.1 Design

During the Municipal Elections of 2014, VAA developer *Kieskompas* developed a VAA in 34 municipalities in the Netherlands. Because of our collaboration with *Kieskompas*, we had access to an anonymized version of these data. In each municipal VAA, between 2,200 and 31,647 VAA users provided an answer to 30 political statements on a five point agree-disagree scale added with a no opinion option, resulting in a dataset with 1,020 statements related to 357,858 respondents. We coded all statements for several question characteristics and used these to predict the proportion of no opinion and neutral answers.

7.2 Codings

All statements were coded for *tax names* (e.g., dog tax), *municipal jargon* (e.g., welfare committee), *locations* (e.g., a street within the municipality), *explicit negations* (e.g., not) and *vague quantifying terms* (e.g., extra). This was done because these question characteristics were shown to be related to comprehension difficulties in Study 1.

While coding for these characteristics, we observed that the category of *locations* should be refined further. This is because VAA statements sometimes refer to a *specific location within the municipality* (e.g., reference to a specific

Table 3
The answers provided in case of comprehension problems versus no comprehension problems

	No sign of comprehension difficulties		Comprehension difficulties		Total	
	n	%	n	%	n	%
Directional answer	1259	88.4	159	44.0	1418	79.4
Neutral answer	142	10.0	120	33.2	262	14.7
No opinion answer	24	1.7	82	22.7	106	5.9
Total	1425	100.0	361	100.0	1786	100.0

In 1425 out of the 1786 cases (79.8%), the respondent experienced no notable comprehension problems.

street or area), to *the municipality as a whole* (e.g., Haarlem), to a more *abstract description of a location* (e.g., on the south-side of the city), or to a *proper name* rather than location name (e.g., soccer club FC Dordrecht). We coded these characteristics separately, as there might be differences in the extent to which they lead to comprehension difficulties.

Finally, we scored the VAA statements for two types of double-barreledness: questions that introduce a *reason for a policy change* (“In order to decrease the number of traffic jams, the roads should be broadened”) and *conditional clauses* (“Money should be invested in cultural heritage, even if this means raising taxes”). These types of double-barreled questions are particularly common in VAAs (Van Camp et al., 2014). Overall, this means that we coded the VAA statements for 10 question characteristics. We refer to Appendix C for the full coding scheme.

We applied the same procedure as in Study 1 for coding the statements. The first author constructed a coding scheme, and coded a subset of the data ($N = 200$). Next, the second author coded the same subset of statements and the differences between the coders were assessed in order to refine the coding scheme. Subsequently, the first author coded all statements based on this revised coding scheme. A third coder (a research assistant) then received training in coding the statements and independently coded a random set of 200 statements (about 20% of the data). The codings of the first author were compared to the codings of the research assistant (see Table 4). These analyses show that the value of Kappa is excellent (Landis & Koch, 1977) for most categories (between 0.85 and 0.95), and substantial for *municipal jargon* (Kappa = 0.70) and *abstract location descriptions* (Kappa = 0.73).

7.3 Statistical Analysis

To analyze the data, we predicted the proportion of no opinion answers (Model 1) and the proportion of neutral answers (Model 2) from the 10 question characteristics. Hence,

Table 4
Kappa and Kappa/Kappa max for each problem category (based on $N = 200$)

	Kappa	$\frac{\text{Kappa}}{\text{Kappa}_{\text{max}}}$
Tax names	0.95	0.96
Municipal jargon	0.70	0.86
Location	0.89	0.94
Abstract location description	0.73	0.80
Name of the municipality	0.91	0.95
Proper name	0.85	0.90
Negation	0.92	0.94
Status quo trigger	0.90	0.96
Reason for policy change	0.90	0.97
“Even if” construction	0.94	0.96

the dependent variable in the analysis is the chance that a random VAA user provides a no opinion answer (Model 1) or neutral answer (Model 2) to a random question. As these models have a similar structure, we will discuss the models for the proportion of no opinion answers as an example. In this model, the Logit of the proportion of no opinion answers was estimated across all municipalities, respondents, and questions. This score was allowed to vary between municipalities (because in some municipalities more no opinion answers are provided than in another municipality), respondents (because some respondents provide more no opinion answers than other respondents), items (because some items elicit more no opinion answers than others) and due to the interaction between respondent and item (all remaining variance). To account for this structure of the data, the municipal variance is modelled at the highest level, and the respondent and item variance are estimated at the same lower level, within municipalities. This means that a cross-classified model is in operation (Quené & Van den Bergh,

2004, 2008). We refer to Appendix A for a formalization.³

7.4 Results

Question characteristics and no opinion answers. Table 5 shows that VAA users are hesitant to choose a no opinion answer: when a question does not contain one of the characteristics we have coded for, a no opinion answer is chosen in about 2% of the cases. Nevertheless, some question characteristics are correlated with a significantly higher percentage of no opinion answers. These are mentioning a *tax name* ($z = 2.67$; $p < 0.01$), *proper name* ($z = 3.75$; $p < 0.001$), *municipal jargon* ($z = 2.42$; $p = 0.01$), and *geographical locations* ($z = 11.75$; $p < 0.001$). Of these four characteristics, especially *geographical locations* have a large effect. This shows from the effect size in Table 5, as well from applying a more intuitive standard of comparison: the mentioning of a specific geographical location, such as a street name, more than doubles the percentages of no opinion responses for a random VAA question. The effects of *municipal jargon*, *proper names* and *tax names* are also substantial, but slightly smaller in terms of both percentages (20–45% increase in no opinion answers) and effects size (effects are small- or medium-sized according to Cohen, 1977, 1988).

Question characteristics and neutral answers. Similar to what we have seen for the proportion of no opinion answers, the mentioning of a *geographical location* has a large impact on the response distribution: when a VAA question includes a location description, the proportion of neutral answers increases with about 5% ($z = 6.52$; $p < 0.001$). Table 6 also shows that mentioning a *municipality name* (e.g., the phrase “in Utrecht” in the VAA of Utrecht) significantly increases the proportion of neutral responses ($z = 2.42$; $p = 0.01$). This finding partially converges with the results for the proportion of no opinion answers (Table 5), as these results already indicated a trend for an effect of this question characteristic ($z = 1.81$; $p = 0.07$). Comparable to what we have seen for the proportion of no opinion answers, the mentioning of an *abstract location description*, such as “in the city centre”, does not affect the proportion of neutral answers ($z = 0.02$; $p = 0.98$).

While the various types of location descriptions largely have the same impact on both neutral and no opinion responding, the impact of *municipal jargon*, *proper names* and *tax names* is only found for the proportion of no opinion responses, and not for the proportion of neutral responses. By contrast, the introduction of quantifying terms that trigger status quo inferences only affect the proportion of neutral responses ($z = 6.18$; $p < 0.001$), rather than the proportion of no opinion answers.

Conclusion Study 2. Study 2 shows that, in a large dataset, statistically significant and meaningful relations can be observed between question characteristics and two response strategies related to comprehension difficulties: neu-

tral and no opinion responding. If we consider neutral and no opinion responding to be indicative of comprehension problems, mentioning a specific geographical location is the most problematic question feature. This characteristic doubles the percentages of non-substantive answers, and, on top of that, boosts neutral responses with about 5%. Results also show that the coding of specific types of location descriptions was worthwhile, whereas concrete locations lead to comprehension problems, abstract location descriptions are not associated with a higher percentage of neutral or no opinion answers. Finally, the results show that *municipal jargon*, *proper names*, and *tax names* affect the proportion of no opinion responses solely, whereas *vague quantifying terms* only affect the proportion of neutral responses. We will elaborate on these findings in the overall conclusion and discussion.

8 Overall Conclusion and Discussion

Study 1 and Study 2 provide converging evidence for the conclusion that *municipal jargon*, *tax names*, *geographical locations*, and *vague quantifiers* complicate the comprehension of political attitude questions in VAA surveys. In addition, and contrary to expectations based on the survey literature (e.g. Dillman et al., 2009; Krosnick & Presser, 2010), they lack support for the conclusion that the use of *negation* and *double-barreledness* lead to comprehension problems. Akin to regular attitude surveys (e.g. Nadler et al., 2015; Sturgis et al., 2014), Study 1 shows that both middle response answers and the non-response answers are chosen in case VAA users experience comprehension difficulties. Hence, these answers can function as proxies to detect comprehension problems. Based on Study 2, this conclusion can be extended, as the choice for either a neutral or a no opinion response seem to be not-random: municipal jargon and

³We re-ran all analyses several times to check if our findings are robust against different ways of cleaning the data. In these analyses, we have taken all cases into account (1), we have excluded straightliners ($N = 16,393$) providing the same answer to all items (2), we have excluded fast responders ($N = 13,542$) who took less than 2 minutes to answer all statements (3), we have excluded slow responders ($N = 20,222$) who took longer than 30 minutes to answer all statements (4), and we have applied a combination of these cleaning methods (5; excluding 37,712 cases). The results reported in the Results Section are based on all available cases. All effects reported here showed to be robust against different cleaning methods, as these effects occurred in all of the samples. Some samples did show additional effects. In sample that excluded fast responders, as well as in the sample that excluded both fast responders, straightliners and slow responders, we observed an additional effect for the mentioning of the municipality name for the proportion of no opinion answers. Moreover, when excluding fast responders, we observed an effect for “even if” constructions and the proportion of no opinion answers. Overall, however, these additional analyses show that our findings are robust against different cleaning methods.

Table 5
Parameter estimates of the proportion of substantive answers relative to the proportion of no opinion answers.

	Estimate Coef.	Std. Error	Estimate in %	Cohen's d_{items}
<i>Fixed Parameters</i>				
Constant	3.898	0.06	98.0	-
Name of tax	-0.29	0.11*	97.4	0.58
Municipal jargon	-0.19	0.08*	97.6	0.38
Location	-0.94	0.08**	95.1	1.86
Other Location	-0.17	0.09	-	-
Proper Name	-0.38	0.10**	97.1	0.76
Municipality Name	-0.14	0.08	-	-
Negation	0.16	0.13	-	-
Status Quo trigger	0.09	0.07	-	-
Reason	0.08	0.08	-	-
Also if	0.12	0.10	-	-
<i>Variances</i>				
Municipality	0.04	0.02	-	-
Item	0.25	0.03	-	-
Respondent	3.48	0.07	-	-

Analyses are based on 10,281,201 respondent and item combinations; 454,539 cases were missing as some respondents stopped filling out the VAA at a certain point in time.

* $p < 0.05$ ** $p < 0.001$

Table 6
Parameter estimates of the proportion of neutral answers relative to the proportion of directional answers

	Estimate Coef.	Std. Err.	Estimate in %	Cohen's d_{items}
<i>Fixed Parameters</i>				
Constant	-1.515	0.02	18.0	-
Tax name	-0.09	0.05	0.00	-
Municipal jargon	0.03	0.04	-	-
Location	-0.29	0.04**	22.7	0.97
Other Location	0.001	0.04	-	-
Proper Name	-0.07	0.05	-	-
Municipality Name	0.10	0.04*	19.5	0.33
Negation	-0.08	0.05	-	-
Status Quo trigger	-0.22	0.04**	21.5	0.73
Reason	0.03	0.04	-	-
Also if	0.09	0.05	-	-
<i>Variances</i>				
Municipality	0.00	0.00	-	-
Item	0.09	0.08	-	-
Respondent	0.00	0.00	-	-

For this analysis, no opinion responses are coded as missing. The analysis is based on 9,995,773 cases.

* $p < 0.05$ ** $p < 0.001$

tax names increase neutral responses solely, whereas vague quantifying terms only boost no opinion answers.

The two studies reported here are likely to give an underestimation of the number of comprehension problems that arise in answering political attitude questions. This is because both methodologies only provide an insight into comprehension problems that VAA respondents are aware of. This may explain why we did not observe many comprehension problems related to the use of *negation* and *double-barreledness*. Study 1 indeed indicates that VAA users are often unaware of negation problems: most of the negation difficulties that we observed, came to light because respondents chose an answer that did not match their line of reasoning (e.g., a “wrong” agreeing answering for a question including a negation). Such problems also went unnoticed in Study 2, because in that study we only analyzed the answer respondents gave (missing “wrong” agreeing answers). For a slightly different reason, problems related to *double-barreledness* may also have gone unnoticed, as the complexity of these questions may only show when comparing interpretations across different respondents. For example, in response to the question “To decrease the number of traffic jams the roads should be broadened” VAA user 1 may just evaluate part A (should the number of traffic jams be decreased?), user 2 part B (should the roads be broadened?), and user 3 the combination of A and B (is A a solution for B?). If this is the case, VAA user 1, 2, and 3 all believe to understand the question quite well, whereas across users, differences arise in the interpretation. In a future study, this explanation can be tested further. Such a study will provide valuable insights for survey methodology on whether or not the classic advice to avoid double-barreledness and negatives should be reconsidered.

A second finding that deserves further discussion, is how to explain that VAA respondents do not “just” mix up the meaning of neutral and no opinion, as has been assumed in the survey literature (e.g. Nadler et al., 2015; Sturgis et al., 2014), but rather, that some comprehension problems are more strongly related to neutral and others to no opinion responding. In our view, this is because different linguistic features affect different parts of the comprehension process. In Study 1, we observed that comprehension problems with the usage of *tax names* and *political jargon* are related to the semantic representation-of the question, whereas *vague quantifiers* merely affect the pragmatic representation-about the question. In case of semantic meaning problems, it is difficult, if not impossible to judge the attitude object in the question: for example, how can respondents answer a question about a specific tax on housing [OZB], if they do not know what kind of tax this is? On the other hand, if a respondent does understand the meaning of all words in the question and “just” lacks background knowledge about the issue at stake (e.g., the exact height of the tax), some beliefs

related to the attitude object can be retrieved, and hence, a vague opinion can be formed. We think that such vague opinions more often result in neutral responses, whereas a lack of semantic knowledge rather results in no opinion responding, explaining why different question characteristics lead to differential response behaviour. The current research is, to the best of our knowledge, the first to show an apparent relation between the type of comprehension problem and the type of response behaviour. Therefore, for building theory within survey methodology, it is crucial that this relation is tested further in a future study. This can be done by asking respondents to motivate their answer choices for a set of questions containing these different types of linguistic elements.

Moreover, it is important to reflect on why, contrary to expectations, VAA respondents expose rather sloppy response behaviour: Study 1 shows that VAA users rarely look up additional information on the web and oftentimes “just” provide a neutral or no opinion answer, and Study 2 confirms that these response patterns generalize to real-life datasets. In our view, these findings simply indicate that VAA surveys are more like attitude surveys than one would guess at a first sight, and than is hypothesized in some articles on VAAs (e.g. Holleman et al., 2013). This implies that rather than providing the best possible answer for each question, VAA users consider the VAA to be a quick-and-dirty tool to obtain just some insight in politics, without spending too much effort. In a future study in which the same attitude statements are answered in two different settings, this explanation can be tested further. In addition, it would be relevant to use in-depth interviews to get an insight into why VAA respondents use a suboptimal response strategy for obtaining a valid voting advice.

Finally, a point of discussion concerns the external validity of our findings. As the two studies reported here show converging evidence for comprehension problems associated with several linguistic factors, we can be quite sure that these actually complicate the answering of questions in Municipal VAAs. As the same kind of people have been found to visit both local and national VAAs, and as the questions in national VAAs are about similar topics, we think that our results generalize to VAAs in general. One exception may be the effect of locations, as national VAAs do not often ask about locations. Moreover, our results are relevant for the broader context of political attitude surveys and surveys on policy issues, as such surveys frequently ask about such topics as raising taxes. Therefore, we expect *political jargon* and *vague quantifying terms* to lead to semantic and pragmatic comprehension problems and stereotypical response behaviour in these contexts too.

Overall, the current research has obtained valuable insights into how question comprehension is affected by linguistic characteristics of political attitude questions, and into the relation between comprehension difficulties and response

behaviour. It has also made a methodological contribution by showing how cognitive interviewing and the statistical analysis of large datasets can be combined. Finally, this study has practical implications for the design of political attitude surveys: *tax names, municipal jargon, geographical locations* and *vague quantifying terms* all have to be avoided because they lead to comprehension difficulties and differential response behaviour. In the specific context of a Voting Advice Application this will inevitably affect the validity of the voting advice negatively.

Acknowledgements

The data corresponding to Study 1 have been uploaded as supplementary file to this article. The answers to the VAA statements analyzed in Study 2 are copyrighted by Kieskompas and cannot be provided. The codings of the 1,020 VAA statements corresponding to Study 2 are provided as online supplementary files. The files are also uploaded to Dataverse (see Kamoen & Holleman, 2017): <https://dataverse.nl/dataset.xhtml?persistentId=hdl:10411/O8EVWE>.

References

- 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* (pp. 115–128). Colchester: ECPR Press.
- Bishop, G. F., Oldendick, R. W., & Tuchfarber, A. J. (1983). Effects of filter questions in public opinion surveys. *Public Opinion Quarterly*, *47*, 528–546.
- Bishop, G. F., Tuchfarber, A. J., & Oldendick, R. W. (1986). Opinions on fictitious issues: the pressure to answer survey questions. *Public Opinion Quarterly*, *50*, 240–250.
- Boren, T. & Ramey, J. (2000). Thinking aloud: reconciling theory and practice. *IEEE Transactions on Professional Communication*, *43*, 261–278.
- Campanelli, P. (2008). Testing survey questions. In E. D. De Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 176–200). New York: Lawrence Erlbaum Associates.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum Associates.
- De Graaf, J. (2010). The irresistible rise of stemwijzer. In L. Cedroni & D. Garzia (Eds.), *Voting advice applications in Europe: the state of the art* (pp. 35–46). Napoli: Scriptaweb.
- Delli Carpini, M. X. & Keeter, S. (1996). *What Americans know about politics and why it matters*. New Haven: Yale University Press.
- Dillman, D., Smith, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed mode surveys. The Tailored Design Method*. Hoboken: Wiley.
- Downs, A. (1957). *An economic theory of democracy*. New York: Harper and Row.
- Flower, L. & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, *32*, 365–387.
- Fowler Jr, F. & Cosenza, C. (2008). Writing effective questions. In E. De Leeuw, J. Hox, & D. Dillman (Eds.), *International handbook of survey methodology* (pp. 136–160). New York: Lawrence Erlbaum Associates.
- Galesic, M., Tourangeau, R., Couper, M. P., & Conrad, F. G. (2008). Eye-tracking data new insights on response order effects and other cognitive shortcuts in survey responding. *Public Opinion Quarterly*, *72*, 892–913.
- Garzia, D. (2010). The effects of VAAs on users' voting behaviour: an overview. In L. Cedroni & D. Garzia (Eds.), *Voting advice applications in europe: the state of the art* (pp. 13–33). Napoli: Scriptaweb.
- Garzia, D., De Angelis, A., & Pianzola, J. (2014). The impact of voting advice applications on electoral participation. In E. Garzia & S. Marschall (Eds.), *Matching voters with parties and candidates. voting advice applications in comparative perspective* (pp. 105–114). Essex: ECPR Press.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *10*, 371–395.
- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken: Wiley.
- Holleman, B. C., Kamoen, N., & De Vreese, C. H. (2013). Stemadvies via Internet: antwoorden, attitudes en stemintenties. *Tijdschrift voor Taalbeheersing*, *35*, 25–46.
- Kamoen, N. & Holleman, B. C. (2017). Think aloud data and codings of VAA statements corresponding to the article “I don’t get it. Response difficulties in answering political attitude statements in voting advice applications.”. online. Retrieved from <https://dataverse.nl/dataset.xhtml?persistentId=hdl:10411/O8EVWE>
- Kamoen, N., Holleman, B. C., Krouwel, A. P. M., Van de Pol, J., & De Vreese, C. H. (2015). The effect of voting advice applications on political knowledge and vote choice. *Irish Political Studies*, *30*, 595–618.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*, 213–236.
- Krosnick, J. A. & Presser, S. (2010). Questionnaire design. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of survey research* (2nd ed., pp. 263–314). Bingley: Emerald Group.

- Krouwel, A. P. M., Vitiello, T., & Wall, M. (2012). The practicalities issuing vote advice: a new methodology for profiling and matching. *International Journal of Electronic Governance*, 5, 223–243.
- Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Marschall, S. (2014). Profiling users. In D. Garzia & S. Marschall (Eds.), *Matching voters with parties and candidates: voting advice applications in comparative perspective* (pp. 93–106). Colchester: ECPR Press.
- Mendez, F. (2012). Matching voters with political parties and candidates: an empirical test of four algorithms. *International journal of electronic governance*, 5, 264–278.
- Nadler, J. T., Weston, R., & Voyles, E. C. (2015). Stuck in the middle: the use and interpretation of mid-points in items on questionnaires. *Journal of General Psychology*, 142, 71–89.
- Quené, H. & Van den Bergh, H. (2004). On multilevel modeling of data from repeated measures designs: a tutorial. *Speech Communication*, 43, 103–121.
- Quené, H. & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, 59, 413–442.
- Rips, L. J. (1995). The current status of research on concept combination. *Mind and Language*, 10, 72–104.
- Schober, M. F. & Conrad, F. G. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly*, 61, 576–603.
- Schober, M. F., Conrad, F. G., & Fricker, S. S. (2004). Misunderstanding standardized language in research interviews. *Applied Cognitive Psychology*, 17, 22–41.
- Schwarz, N. (2007). Cognitive aspects of survey methodology. *Applied Cognitive Psychology*, 21, 277–287.
- Sturgis, P., Roberts, C., & Smith, P. (2014). Middle alternatives revisited: how the neither/nor response acts as a “face-saving” way of saying “I don’t know”. *Sociological Methods and Research*, 43, 15–38.
- Tourangeau, R. & Rasinski, K. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, 103, 299–314.
- Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.
- Van Camp, K., Lefevre, J., & Walgrave, S. (2014). The content and formulation of statements in voting advice applications. In D. Garzia & S. Marschall (Eds.), *Matching voters with parties and candidates. Voting advice applications in comparative perspective* (pp. 11–32). Essex: ECPR Press.
- Van de Pol, J., Holleman, B. C., Kamoen, N., Krouwel, A. P. M., & De Vreese, C. H. (2014). Beyond young, higher educated males: a typology of VAA users. *Journal of Information Technology and Politics*, 11, 397–411.
- Van den Ven, D. (2014). *Formuleringseffecten in stemhulpen: aanwezig of niet afwezig? Experimenteel onderzoek naar de effecten van valence framing en attitudesterkte op de antwoorden die mensen geven op stellingen in stemhulpen*. MA Thesis: Tilburg University.
- Van Weijen, D., Van den Bergh, H., Rijlaarsdam, G., & Sanders, T. (2009). L1 use during L2 writing: an empirical study of a complex phenomenon. *Journal of Second Language Writing*, 18, 235–250.
- Wall, M., Krouwel, A. P. M., & Vitiello, T. (2014). Do voters follow the recommendations of voter advice application websites? A study of the effect of kieskompas.nl on its users’ vote choices in the 2010 Dutch legislative elections. *Party Politics*, 20, 416–428.
- Willis, G. B. (2005). *Cognitive interviewing. A tool for improving questionnaire design*. Thousand Oaks: Sage Publications.
- Zwaan, R. A. & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162–185.

Appendix A
Multi-level models used

In Equation 1, the model used for analyzing the proportion of no opinion answers is formalized. In this model, $Y_{i(jk)}$ indicates whether in municipality i ($i = 1, 2, \dots, 34$) individual j ($j = 1, 2, \dots, 31, 647$) provides a no opinion answer for question k ($k = 1, 2, \dots, 30$). In the model, the mean proportion of no opinion answers is estimated in Logits. This average proportion of no opinion answers (CONS β_1) is allowed to vary between municipalities ($u_{i(00)CONS}$), and between persons ($v_{0(j0)CONS}$), and items ($w_{0(0k)CONS}$) within municipalities. The between-person and between-item variance are estimated at the same time, which means that a cross-classified model is in operation (Quené & Van den Bergh, 2004, 2008).⁴

In addition, to estimating one average proportion of no-opinion answers, 10 deviations (β_2 – β_{11}) are estimated, each one indicating how the proportion of no opinion answers changes if the question contains the prescribed characteristic. These deviations are estimated by creating dummy variables, which can be turned on if the observation matches the prescribed type. Hence, DEVIATION_D_TAX indicates how much questions in which a tax name has been mentioned deviates from the average Logit proportion of no opinion answers. All residuals are normally distributed with an expected value of zero.

$$\begin{aligned} \text{Logit}(Y_{i(jk)}) = & \\ & \text{CONS}_{i(jk)}(\beta_1 + u_{i(00)CONS} + v_{0(j0)CONS} + w_{0(0k)CONS}) + \\ & \text{DEVIATION_D_TAX}(\beta_2) + \\ & \text{DEVIATION_D_JARGON}(\beta_3) + \\ & \text{DEVIATION_D_LOCATION}(\beta_4) + \\ & \text{DEVIATION_D_ABSTRACT_LOCATION}(\beta_5) + \\ & \text{DEVIATION_D_NAME_MUNICIPALITY}(\beta_6) + \\ & \text{DEVIATION_D_PROPERNAME}(\beta_7) + \\ & \text{DEVIATION_D_NEGATION}(\beta_8) + \\ & \text{DEVIATION_D_REASON}(\beta_9) + \\ & \text{DEVIATION_D_EVENIF}(\beta_{10}) + \\ & \text{DEVIATION_D_STATUSQUOTRIGGER}(\beta_{11}) \quad . \end{aligned}$$

⁴Please note that the model also implies that there is also variance due to the interaction between respondent and item. However, because the DV in the model is binomial this variance is not estimated, as it is fixed if the mean proportions are known. The interaction variance can be approximated applying the formula $p(1-p)$, in which p represents the estimated proportion of no-opinion answers.

Appendix B
Coding Scheme for Verbalizations

The different comprehension problems are displayed in Table A1. If the verbalization pointed to a comprehension problem, but could not be classified into these categories, it was scored as *other/unknown comprehension problem*.

Table B1

Description and examples of the characteristics coded in Study 1.

Characteristic	Description	Example
Unknown Concept	The respondent lacks knowledge of the semantic meaning of a concept.	"Budget X [e.g., liveability budget] . . . never heard of that budget".
Location	The respondent does not know where a certain location is.	"I don't know where location X [e.g., Amelisweerd] is."
Vagueness	A term in the question is vague/ too broad.	What do they mean by "punishing harder"? That is a broad concept.
Negation	The respondent thinks the question is confusing due to the usage of an explicit negation (not; in Dutch "niet" or "geen"), or the respondent picks an answer that doesn't match his line of reasoning due to the usage of negation.	"Yes, yes, yes. The accessibility of the city is important, but not very important, so I'll pick a regular "agree" (to a question about not widening a specific road)"
Double-barreledness	The respondent either explicitly mentions that the question is difficult because it is double-barreled, or this shows from his line of reasoning.	"[reads part of the question]: No subsidies or loans . . . loans yes, but subsidies no . . . euhm . . . I guess will pick disagree then"
Other remarks about the language use	The respondent makes a remark about the question wording, e.g. because the question wording is too strong, or contradictory. Please note that for some remarks in this category it is unclear whether they comprise "just" feedback on the question wording (e.g. I think must is too strong), or also a comprehension problem (e.g. "that is too strong and therefore I cannot give an opinion"). To avoid confusion, we scored all remarks about the question wording as "other remarks about the question wording".	"That is too black and white"
Pragmatic meaning problems	The respondent has (too) little information about at least one facet of the question. Please note that semantic meaning problems related to a lack of knowledge of a term often presuppose a pragmatic meaning problem (e.g. "Liveability budget . . . never heard of that budget"). In these cases, we only scored the semantic meaning problem.	"Budget X [e.g., liveability budget] . . . I don't know how much is spent on that budget"

Appendix C
Coding Scheme for VAA Statements in Study 2

Table A2 shows the categories the VAA statements were scored for in Study 2. We have attempted to construct definitions that can be applied easily, such as: Does the statement contain a negation (*not* or *none*), or not? Does the statement contain the name of the municipality or not? This choice was made because these types of codings can be made reliably.

Table C1
Description and examples of the characteristics coded in Study 2.

Characteristic	Description	Example
Tax name	Name of a tax.	OZB, dog tax, including references to “local payments” and “local taxes”.
Other political jargon	Terms commonly used in politics (list available through first author).	Referendum, Ombudsman, decibelnorm.
Locations	Reference to a specific location outside or within the municipality; start with a capital letter. E.g. this includes names of streets, polders, neighbourhoods, and squares.	Rijnenburg, Amelisweerd
Abstract description of a location	Abstract description of a location. If the abstract description of a location precedes a location description, it is scored as abstract description of a location.	In the city centre, In the suburbs
Proper name	Reference to a Proper name, starts with a Capital Letter.	In the CKC-theatre
Name of Municipality	Name of the municipality	Haarlem (in the VAA for the Municipality of Haarlem)
Negation	Explicit negation	Not, None
Status quo trigger	Sentences containing the words extra, increase, decrease, more, and less, or inflections thereof.	The budget for culture should be increased
Reason for policy change	Reason for policy change; these sentences include the words “om”, “omdat” or “zodat” (because).	To decrease the number of traffic jams, the roads should be broadened
“Even if” construction	Construction with “ook als” of “ook al” (even if).	More money should be spent on culture even if this means raising taxes