

Measuring Congruence Between Voters and Parties in Online Surveys: Does Question Wording Matter?

Bruinsma, Bastiaan

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:

Bruinsma, B. (2023). Measuring Congruence Between Voters and Parties in Online Surveys: Does Question Wording Matter? *Methods, data, analyses : a journal for quantitative methods and survey methodology (mda)*, 17(1), 71-92.
<https://doi.org/10.12758/mda.2022.12>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:
<https://creativecommons.org/licenses/by/4.0>

Measuring Congruence Between Voters and Parties in Online Surveys: Does Question Wording Matter?

Bastiaan Bruinsma

Chalmers University of Technology

Abstract

Congruence on policies between political parties and voters is a frequently assumed requirement for democracy. To be able to study this, we should be able to calculate accurate and precise measures of policy congruence in political systems. This could then tell us more about the political system we study, and the “distances” that exist between parties and voters on either issues or broader ideological dimensions. Here, I draw on experimental data from a Voting Advice Application to show that the wording of the issues can influence the degree of congruence one measures. Yet, this comes with the complication that this influence depends on the type of issue, the characteristics of the voters themselves, and the party the congruence is calculated with. These findings should serve as a warning for those who aim to measure congruence that even minor changes in question-wording can (but do not have to) cause relatively large changes in congruence, especially when many parties are involved and the differences between the congruences are small.

Keywords: congruence, wording effects, voting advice applications



© The Author(s) 2023. This is an Open Access article distributed under the terms of the Creative Commons Attribution 3.0 License. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Congruence between voters and parties is one of the cornerstones of democratic representation (e.g., Huber & Powell, 1994; Katz, 1997; Powell, 2004). The idea is that the higher the level of congruence, the better-represented we can say citizens are. As such, we can use the degree of congruence to say something about the degree of democracy in a country (Diamond & Morlino, 2005). While congruence is usually high on broad ideological dimensions - such as economic left-right - it is often lower for individual policy issues (Dolný & Baboš, 2015; Dalton, 2017). This is problematic, as while dimensions are still important, there is a significant number of voters who vote on issues, instead of broad ideological dimensions and partisanship (Dalton, 2017). So, it is no surprise that the number of studies that focus on issue congruence is increasing (e.g., Belchior & Freire, 2013; Romeijn, 2020; Rosset & Stecker, 2019; Costello et al., 2020).

To measure issue congruence, one needs valid and unbiased positions of both voters and parties on these issues, as well as a mechanism to compare them. To find the position of a party, there are various methods, each of which has its advantages and drawbacks (Mair, 2001; Benoit & Laver, 2006). Often, they are dependent on experts, though they may also use the opinion of the party itself. For the position of the voter, the only option seems to be to ask the voter themselves, most often in the form of a survey. Here, one presents them with a series of issues and ask them to state their agreement or disagreement with them. While straightforward, this requires a certain degree of response quality on part of the voter to be useful. In other words, we need the voter to provide a response that is as close to their “true” opinion as possible.

Yet, such a high degree of response quality is by no means guaranteed. At its best, the response quality of the average voter is often only sub-optimal (Blasius & Thiessen, 2012, p.3). This is especially so if the voter perceives the formulation of the issue to be difficult, which can happen when the wording of the issue is long and complicated, or when the issue uses negative wording. Given the increased relevance of issue congruence, the question is then how relevant response quality is when measuring congruence. To put it in other words: to what extent does the “difficulty” of the formulation of the issue influence its measure?

Of the various ways in which an item can be difficult, one of the simplest forms is whether an issue is positive or negative. This is also known as the *polarity* of an issue. Negative issues, as well as their positive counterparts, occur in almost all surveys. Survey designers include them to: a) minimize response styles such as acquiescence bias, and b) allow for the inclusion of negatively worded issues from previous surveys. Yet, it is well known that negative issues come with certain prob-

Direct correspondence to

Bastiaan Bruinsma, Department of Computer Science and Engineering, Chalmers
University of Technology, Göteborg, Sweden
E-mail: sebastianus.bruinsma@chalmers.se

lems. For example, respondents might miss the negation in the sentence, or become confused about how to map their response to the response options. As a result, they spend longer on reading the item (Kamoen et al., 2011) and mapping their response to the correct option (Chessa & Holleman, 2007). This, in turn, leads to an increase in non-response and respondents giving the opposite response to the one they intended. Holleman et al. (2016) show the latter in an experiment where they changed some issues from positive to negative, such as changing *forbid* to *allow*. They found that the responses were different than they should be (i.e., their direct opposite), though the effect they found was rather small (on a scale of 1 to 5, only 1 out of 20 respondents showed a scale point difference) (Holleman et al., 2016, p. 9).

Yet, many small differences in response might have large effects when it comes to the congruence between the voter and the party. Besides, it is unclear if the effect of changing the polarity is the same for all issues, parties, and voters. For example, certain issues might be more susceptible to such effects while others might not show any difference at all. Also, if a party associates itself with a certain formulation of an issue it might suffer if the formulation of the issue is the opposite. Finally, voters with different levels of political sophistication might show differences in the way polarity influences them. For example, voters with low political sophistication could find it more difficult to handle negatively worded issues than voters with a high level of political sophistication.

In this paper, I will thus assess the following: *does changing the polarity of an issue affect the degree of congruence between a voter and a party?* For this, I will run an experiment in which I present voters with both the positive and negative versions of the same issues. In addition, I will also take the effects of issue salience and respondent characteristics into account. The experiment itself is carried out in the context of a Voting Advice Application (VAA) - an online tool voters can use before the elections to calculate their congruence with certain parties. Not only is congruence central to the idea of the VAA, but the VAA community itself has also been actively involved in studying how and when variations in congruence occur (Louwerse & Rosema, 2014).

From here on, the structure of this paper is as follows. First, I will introduce the increased relevance of issue congruence, which leads to a discussion on how we should measure it. Based on the congruence literature, I will identify several factors that might influence the effect of the question wording such as issue salience and the political sophistication of the respondent. Then, I will describe the research design and the set-up of the experiment as well as the measures used. Finally, I will attempt to answer the main research question as well as discuss several implications of the findings.

Relevance of Issue Congruence

We can define congruence as the distance between a citizen and their representative (Dalton, 1985). This representative could be either a single candidate or a party. The idea of congruence itself was for a long time equated with ideological congruence, which referred to the congruence between voters and parties on an ideological dimension. If on this dimension the distance between the voter and those that represent them neared or equaled zero, they were said to be congruent.

While simple, there are several problems with the ideological congruence approach. The first is that the positions of voters and parties often have different meanings. Voters are, by definition, a larger group than parties, resulting in their positions varying more than those of the parties. Second, issues not included in the dimension might be relevant for either the voter or the party. It could thus be that a voter and a party appear to be congruent, even though they disagree on issues that are fundamental to both. Besides, there is no reason to suppose that congruence on a dimension also implies congruence on a certain issue (Thomassen, 2012). And while this might not be a problem when the issue is not salient to either the party or the voter, it becomes a problem when it is (Giger & Lefkofridi, 2014). As a result of this, Dalton (1985, 2017) shows that congruence not only differs between issues but between parties as well. Finally, for ideological congruence to work, one needs to position both the party and the voter on the same metric. Yet, due to data limitations, this is often not possible, and parties and voters are often positioned on different metrics that scholars assume to be similar.

Partly because of the first two issues, the focus on issue congruence has increased over the last years. To measure it, one can use either of three approaches, dependent on how one measures the positions of the parties and the voters (Powell, 2009). These are a) a voter identification and voter perception approach, b) a party vote and party manifesto approach, and c) a voter survey and expert survey approach. In the first, voters' position both themselves and any parties on a certain issue. While this tackles the problem of not using the same metric, it assumes that voters have a good knowledge of both their own and of the parties' position. This goes against the experience that voters are rarely that well informed on these matters. Also, a voter's perception of the position of a party is dependent on their own position as well. In the second approach, one takes voters in part out of the equation by looking at their votes during the elections to estimate their position. For the position of the party one then takes the electoral manifesto the party released for the same elections. While this circumvents the problems with the voters, one's vote is rarely indicative of one's true position - especially if choices are limited. Also, the use of party manifestos to position parties is not uncontroversial as well (Dinas & Gemenis, 2010). The third option - combining a voter survey and expert survey approach seems to tackle most problems and is, therefore, the approach used

by most issue-congruence scholars (e.g., Giger & Lefkofridi, 2014; Costello et al., 2020). The reason it works is that voters are often better aware of their positions on single issues than on a whole dimension and expert surveys are often viewed as the “gold standard” for party positioning and are flexible to implement.

Effects of Question Wording

The remaining problem with measuring issue congruence are then the issues themselves. Not only does the actual content of an issue matter, but also its formulation (e.g., Hippler & Schwarz, 1986). For example, Schuldt et al. (2011) found Republicans less likely to endorse the existence of “global warming” than the existence of “climate change”. This is because different terms draw the readers’ focus to different aspects of the phenomenon. One might accept the “warming” while agreeing on “change” is something very different. The same occurs when we alter the polarity of the question. That is, agreeing with something does not mean that one disagrees with the opposite. For example, agreeing that soft drugs should *be forbidden* does not mean that one disagrees that they should *be legalized*.

A negation such as *forbidden* is known as an implicit negative. That is, the negation is in the word itself. Its alternative is the explicit negative, where one places the word “not” in front of the positive verb (Clark, 1976). Thus, it would be *legalize* versus *not legalize*. Of the two, the implicit version is often the most popular option. This is because one can only use an implicit negative in a context where it makes sense. For example, one does not talk of *keeping someone from* suicide unless one supposes that someone intended to commit suicide in the first place (Horn, 1989, p.523). Also, the alternative, the explicit negatives, take longer to understand. This is as one first must reconstruct the positive version, and then make it negative. As an example, Kaup et al. (2006) asked respondents to imagine a non-open door. Most respondents took twice as long in imaging this as they would in imaging an open door, as they first had to open the door and then “non-open” it.

The effect of a change in the polarity of an item can then differ on three levels: the respondent, the party, and the issue. As such, we should focus on each of them to see how they behave when we change the polarity of an item. To begin with, we set out to see whether question polarity has any influence on the response behavior of the respondents at all. Given that it seems reasonable to assume that respondents will respond differently when an item has a different formulation, I hold that:

H1 Question polarity affects the responses respondents give to items.

Turning first to the respondents, it is likely that the characteristics of the respondent condition the effect of changing the polarity of the question. On considering public policy issues, political sophistication is one of the main relevant characteristics.

The concept itself is multi-faceted. While it relates to political knowledge, it also includes the facility to acquire new information, the ability to translate personal values into opinions and behavior, and one's motivation to do so (Luskin, 1990; Highton, 2009). Thus, respondents with low levels of political sophistication are more likely to have problems with negative questions, as the negative version of a question is more complicated (Holleman et al., 2017). Besides, respondents with a low level of political sophistication will pay less attention to these questions (Bassili & Krosnick, 2000), again increasing the degree of misunderstanding. It is thus no surprise that Holleman et al. (2016) found that lower levels of political sophistication led to larger effects of the polarity. It seems thus reasonable that the opposite is also the case. Therefore:

H2 The higher the sophistication of the respondent, the lower is the effect of the polarity.

As for the issues, one characteristic that might influence the responses is the salience of an issue. Most often, one measures this by asking after the "most important issue" during a certain period. The more important the issue to people, the higher its salience. As a result, voters are better informed on those issues, as they tend to be so when the issue is important to them (Giger & Lefkofridi, 2014). This makes the issue less difficult for the voter, leading them to earlier spot a change in the difference of wording. Thus:

H3 The higher the salience of the issue, the lower the effect of question polarity.

Finally, we turn to the parties. As there is a seemingly one-on-one relation between the response of the respondent and the congruence (as the position of the party does not change here), we expect that any change in the response will also lead to a change in the degree of congruence. As an extension of the first hypothesis, therefore:

H4 The polarity of the item influences the congruence between respondents and parties.

Research Design

To measure the influence of the question wording on the degree of congruence, I will turn to an instrument that has congruence at its heart: Voting Advice Applications (VAAs). These are online questionnaires that compare the answers of their respondents with those of political parties on the same issues. This comparison is then shown as the degree of agreement between the respondent and the party - in other words, their congruence (see also Costello et al., 2020). Using VAAs to calcu-

late issue congruence has several advantages. First, the items are the same for both respondents and parties, thus avoiding the complications other methods have when comparing the positions (Krosnick & Presser, 2010; Costello et al., 2020). Also, we can calculate the match without requiring any further scaling analysis. Finally, as VAAs are popular online instruments, reaching the desired number of responses is easy.

As VAAs, in general, have only a single version of the question wording, to test for the effect of question polarity, we have to run two similar VAAs. Even so, as the wording of the question could potentially influence the respondent and thus the outcome of the VAA, it would be unethical to run two versions with either only positive or only negative items. To get around this, I designed a VAA with two different versions (hereafter Version A and Version B). Both versions had 25 questions in total, with 13 questions being similar for both (these questions all had a positive wording). Of the remaining 12 questions, the polarity of the wording differed for each version. Of the negative questions, 4 of them were explicit negatives and 8 were implicit negatives. The topics of the questions were decided upon in cooperation with other members of the design team based on their expected relevance during the elections. For a full overview of the questionnaire, see Appendix A.

The VAA designed for this study – Stem-Consult – launched several weeks before the elections for the House of Representatives in the Netherlands on March 15, 2017¹ (Gemenis et al., 2017). I reached out to potential respondents through word-of-mouth and targeted Facebook advertisements. Upon entering the website, the VAA assigned respondents at random to either Version A or B of the VAA. Besides the main questionnaire, optional questions asked respondents for their age, political interest, education, and gender. The VAA included 14 of the 28 parties taking part in the elections, which were included either because they were already represented in parliament or showed in the polls a consistent chance of gaining at least a single seat. As some parties were favored only by a small percentage of the respondents, I decided to only include here the 8 largest parties. These are the CDA, a Christian-democratic party, the CU, a social Christian party, D66, a social-liberal party, GL, a green party, the PvdA, a social-democratic party, the PVV, a radical right populist party, the SP, a radical left party, and the VVD, the main center-right liberal party. For an overview of these parties and their abbreviations, see Appendix B.

1 Stem-Consult was designed and launched in cooperation with the PreferenceMatcher consortium.

Voter and Party Positions

In a VAA, respondent positions come from the main questionnaire. Here, the VAA presents the respondent with 25 items - one item at the time - and asks them to respond on a 5-point Likert scale. This scale ranges from *Completely Disagree* to *Completely Agree*, with a *No Opinion* option included. For these questions, there is no time limit and respondents have the opportunity of returning to a previous item if they wish to change their response.

For the party positions, I employed a team of coders that coded the positions of the parties. This coding takes place on the same issues and uses the same response options as the respondents are presented with. For the coding, I used the iterative Delphi-process (Gemenis, 2015). This process presents the coders with the issues in a first round and asks what they think would be the party's position on it. Together with their answer, they are then asked by the moderator to back up their idea of the party's position using any source of information they please. Also, they have to tell how confident they are of their opinion. This information is then collected, anonymized, and fed back to the coders. This allows them to see how the other coders positioned a party on an issue, as well as how confident they are of this position. Then, the moderator asks them if they want to reconsider their original position. The idea then is that those respondents who were not very confident of the position will alter theirs to be more in line with those that are confident. This process then repeats until the coders reach a certain degree of agreement between them. Here, I calculate this agreement following Gemenis (2015) and use Van der Eijk's coefficient α (van der Eijk, 2001). I do so as the ordinal rating scales used to position parties in VAAs do not lend themselves well to other common methods of agreement, such as the standard deviation, as these often reflect the skewness of distribution in addition to the dispersion (van der Eijk, 2001, p.328). Van der Eijk's coefficient α circumvents this by taking a weighted average of the degree of agreement that is there for the individual categories. Though Van der Eijk (2001) does not offer any cut-off point for α , I follow Gemenis and Van der Ham (2014), who carried out a similar analysis during the Dutch elections of 2014 and require the agreement to be higher than $\alpha > 0.7$ for the process to finish. When at this point the positions of the coders still differ, I take the average between them as the position of a party. For this VAA, the coders positioned all parties on the questions as they appear in version A of the VAA. The positions for Version B were then generated by reversing the positions where necessary. While this is in no way ideal, constraints of both time and resources led us to settle on this approach.

Congruence Measures

To calculate the agreement between a respondent and a party, VAAs can draw on two schools of thinking. The first draws on Downs (1957) and holds that agreement is the distance between a party and respondent. The second finds its origin in Rabinowitz and MacDonald (1989) and supposes it is not the distance, but the intensity that counts - relevant is thus if the respondent and the party are on the same side of the argument, regardless of the distance between them. As I deem both to be relevant, I opt to use a hybrid model that splits the difference between both methods (Mendez, 2017). With this algorithm, the degree of congruence can range between -100 to 100. Here, -100 means that when the respondent completely disagreed the party completely agreed, or the respondent completely agreed the party completely disagreed. On the other hand, +100 means that both the respondent and party completely disagreed or completely agreed. Also, 0 means that either the party or the respondent was neutral while the other completely agreed or disagreed (cf. Mendez, 2014). For a complete overview of how this algorithm works, see Appendix C.

Political Sophistication

For political sophistication, I create an additive scale using the education and political interest variables, both of which are measured on a five-point scale. The result of this is an additive scale running from 2 to 10, which lower values indicating a low level of political sophistication and higher values indicating a higher level of political sophistication.

Issue Salience

To measure the degree of issue salience, I will make use of the data supplied by the Dutch Parliamentary Election Study 2017 (van der Meer et al., 2017), which was fielded around the elections during which the VAA was implemented. Here, I only used the responses collected by CAPI, which provided 927 respondents who completed the questionnaire. Then, for issue salience, I used the question “What do you think are the most important problems in our country?”². This question resulted in an open answer where the respondent could name more than one problem, they considered important. These answers were then re-coded into nine categorical variables ranging from the most important problem to the ninth, sorted into twenty-two different issue categories. Of the 927 respondents, 857 mentioned at least a single problem. I used the average frequency of mentions of an issue over the total number

2 The original wording in during the survey was: “*Wat zijn volgens u de belangrijkste problemen in ons land?*”

of issues mentioned as the average importance of that issue (Hosch-Dayican et al., 2013). I then related the twenty-two categories from the DPES to the topics of the items in Stem-Consult. The results of this are in Table 1. Here, the first column refers to the item in the VAA, the second to which of the twenty-two different issue categories it belongs, and the third shows the percentage of respondents that mentioned that issue.

Table 1 Overview of the issues and their percentage of issue salience

#	Item	Type of Issue	Salience (%)
1	Art subsidies	Media	0.2
2	Public broadcasting	Media	0.2
3	Development cooperation	Inequality/Poverty	5.4
4	Pension Age	Social security	3.1
5	Mortgage relief	Housing	1.5
6	Anonymous application	Inequality/Poverty	5.4
7	Insurance	Income/Price levels/Taxes	1.5
8	End of life	Norms and values	5.0
9	Threat of Islam	Norms and values	5.0
10	Soft drugs	Norms and values	5.0
11	Remain in EU	European integration	1.4
12	EU Expansion	European integration	1.4
13	Immigration	Norms and values	5.0
14	Acceptance of refugees	Minorities	18.3
15	Environmental measures	Environment	7.9
16	Energy saving measures	Regulation/Big government	0.9
17	Coal Plants	Environment	7.9
18	Loan system	Education	7.3
19	Binding referendum	Politics	3.8
20	Defense	Defense	0.9
21	Spending on social work	Social security	3.1
22	Own risk in healthcare	Healthcare	19.4
23	Healthcare and market	Healthcare	19.4
24	Mileage charge	Traffic/Mobility	1.3
25	Multicultural society	Minorities	18.3

Data

Given that VAA data are online data, Mendez et al. (2014) and Andreadis (2014) advise cleaning up the data before using it. Following their advice, I removed respondents when: the time taken to complete the total of 25 issue statements was less than 75 seconds; they answered at least one issue in less than 2 seconds or when they answered 12 or more consecutive statements in the same way. Besides, I removed returning respondents – identified by similar entries from the same computer – as well as all respondents taking the VAA after March 15, which was the date of the election, and those of the VAA between 10-12 March, when the VAA was taken offline for a security update. Finally, I removed the first 50 entries as these most likely were filled out during initial testing. After doing so and selecting those respondents who filled out all the items and for which data on political interest and education was available, 2674 respondents remained.

Of these, 1328 were in Version A and 1346 in version B. Both groups were not different with regard to Sex ($X^2(1)=0.07$, $p=0.79$), Age ($t(2670)=-0.61$, $p=0.54$), or Education ($X^2(1)=0.39$, $p=0.53$), and though there was a significant difference in Political Interest ($X^2(1)=4.75$, $p<0.05$), the actual differences of 2.40 for Version A and 2.47 for Version B are too small on a 5-point scale to be expected to make any conceivable difference.

Within the sample, 45.8% was male and 53.9% was female. Compared to the general population with 49.6% male and 50.4% female, this indicates that Stem-Consult attracted more females than males. As for age, the mean age for Stem-Consult was 38.2 for men and 43.0 for women, while in the general population these are 40.7 and 42.5. For education, I find that 51.7% of the respondents had a graduate or post-graduate education, compared to 23.5% in the population. This means that Stem-consult reached younger and higher educated respondents, as is common for most VAAs (van de Pol et al., 2014).

Results

Before we look at the hypotheses, it might be instructive to look at those items not included in the analysis - that is, those items that were positive in both versions. For these items, in contrast to those that had different versions, there should be no difference in the response between the two versions of the VAA.

Table 2 Independent samples t-test for the unaltered items of the VAA. A and B refer to either of the two versions of the VAA.
For significance, * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$)

Item	Mean		t	p	Sig.	CI	
	A	B				Low	High
q3	3.51	3.57	-1.32	0.19	-	-0.14	0.03
q4	2.57	2.53	0.79	0.43	-	-0.06	0.14
q6	3.39	3.52	-3.06	0.00	**	-0.22	-0.05
q7	2.66	2.73	-1.76	0.09	-	-0.15	0.01
q8	2.07	2.08	-0.34	0.73	-	-0.09	0.07
q9	2.81	2.83	-0.30	0.76	-	-0.11	0.08
q12	3.68	3.71	-0.70	0.48	-	-0.10	0.05
q13	1.58	1.59	-0.43	0.66	-	-0.07	0.04
q14	3.65	3.67	-0.48	0.63	-	-0.10	0.06
q19	2.97	2.97	-0.05	0.96	-	-0.10	0.09
q20	2.60	2.56	0.80	0.42	-	-0.05	0.11
q23	3.61	3.73	-2.71	0.01	**	-0.20	-0.03
q24	2.71	2.78	-1.67	0.10	-	-0.16	0.01

To see if degree this is the case, Table 2 shows the independent samples t-test for each of these thirteen items. As expected, in all but two cases, there is no statistically significant difference in the respondents' responses. The two exceptions to this are item 6 ("Anonymous application must become the norm for government jobs") and item 23 ("Through free-market operation, healthcare functions better"). In these two cases, the differences between both versions are 0.13 and 0.12 respectively. While there is no clear evidence on what causes these differences, given their small size it seems quite unproblematic to ignore them.

Table 3 Table of Coefficients for the first linear model, as well as the percentage of salience for each of the items

Item	Intercept		Pos/Neg		Salience (%)
q1	2.91***	(0.03)	-0.03	(0.05)	0.2
q2	3.38***	(0.03)	-0.1**	(0.04)	0.2
q5	3.65***	(0.03)	0.01	(0.04)	1.5
q10	2.5***	(0.03)	0.2***	(0.04)	5.0
q11	3.64***	(0.03)	-0.06	(0.05)	1.4
q15	3.26***	(0.03)	-0.18***	(0.04)	7.9
q16	2.38***	(0.03)	0.17***	(0.04)	0.9
q17	3.34***	(0.03)	0.01	(0.04)	7.9
q18	3.4***	(0.03)	-0.29***	(0.04)	7.3
q21	3.88***	(0.03)	-0.05	(0.03)	3.1
q22	3.52***	(0.03)	-0.02	(0.05)	19.4
q25	3.4***	(0.03)	-0.09**	(0.04)	18.3

Effect of Question Polarity

Turning then to the hypotheses, Table 3 shows the results for the first hypothesis. For this, I used a linear model with the response of the respondent as the dependent variable and the polarity as the independent (binary) variable. Note that the response scale runs from 1 (negative) to 5 (positive) and the polarity is either negative (0) or positive (1). The results show that the polarity only had a significant effect on the response for six of the twelve items. The effect is the largest for item 18 (“The loan system for students should be abolished/maintained”) with -0.29. This means that if one would alter the wording of the item from “abolished” to “maintained” this would lead to respondents giving a more negative response. There are lower (but significant) values for items 15, 16 and 10, with coefficients of -0.18, 0.17 and 0.2. As with item 18, these values are small considering the scale. Besides, there is little consistency in the direction of the effect. In two cases (items 10 and 16) a positive wording instead of a negative one led to a more positive response. In the other four (items 2, 15, 18 and 25) a similar change would lead to a more negative response. Thus, not only do only some of the items show an effect of polarity, of those that do so the effects are small and of inconsistent direction. Taken together, this means that we can only partly confirm Hypothesis 1.

Table 4 Table of Coefficients from the linear models with the interaction term

Item	Intercept		Pos/Neg		Soph.		Pos/Neg x Soph.	
q1	2.88***	(0.16)	-0.39*	(0.23)	0	(0.03)	0.06	(0.04)
q2	2.99***	(0.15)	0.48**	(0.21)	0.06***	(0.02)	-0.09***	(0.03)
q5	3.55***	(0.15)	0.02	(0.21)	0.01	(0.02)	0	(0.03)
q10	2.86***	(0.16)	-0.05	(0.22)	-0.06**	(0.02)	0.04	(0.03)
q11	2.4***	(0.17)	0.21	(0.24)	0.19***	(0.03)	-0.04	(0.04)
q15	3.71***	(0.15)	-0.3	(0.21)	-0.07***	(0.02)	0.02	(0.03)
q16	2.3***	(0.14)	0.25	(0.19)	0.01	(0.02)	-0.01	(0.03)
q17	3.26***	(0.13)	0.03	(0.19)	0.01	(0.02)	0	(0.03)
q18	3.74***	(0.16)	-0.65***	(0.22)	-0.05**	(0.02)	0.06*	(0.03)
q21	4.09***	(0.11)	-0.19	(0.16)	-0.03*	(0.02)	0.02	(0.03)
q22	4.64***	(0.16)	0.07	(0.23)	-0.18***	(0.02)	-0.01	(0.04)
q25	2.3***	(0.15)	0.26	(0.21)	0.17***	(0.02)	-0.05*	(0.03)

Effect of Sophistication

We now turn to the second hypothesis. Here, we considered whether the effect of the polarity is lower when the sophistication of the respondent is higher. To see if this is the case, we first run a second linear model to see if we can expect an interaction between the two to begin with. Thus, in this model, we include both sophistication and its interaction with polarity. Table 4 shows the results for this. Here, we find a significant interaction for only three of the items: 2, 18 and 25. Thus, only in three cases is the effect of sophistication different for either the positive or negative version of the item.

Yet, to see how different, we have to look at the effect at various levels of sophistication. For this, we must look at the average marginal effects at representative cases (MERS). The representative cases here are all those cases that have one of the nine levels of political sophistication. The marginal effects are the contribution of the polarity to the response. Table 5 shows these marginal effects at each of the nine levels of political sophistication. As higher levels of sophistication should lead to a lower effect of the polarity, the marginal effects there should tend towards zero. In other words, at those levels, the polarity contributes little to nothing to the eventual response of the respondent.

Table 5 Table of Margins

Item	2 (Low)	3	4	5	6	7	8	9	10 (High)
q1	-0.28	-0.22	-0.17	-0.11	-0.05	0.00	0.06	0.11	0.17
q2	0.30	0.21	0.12	0.03	-0.06	-0.15	-0.24	-0.33	-0.42
q5	0.02	0.01	0.01	0.01	0.01	0.00	0.00	0.00	-0.00
q10	0.03	0.07	0.11	0.15	0.19	0.23	0.27	0.31	0.35
q11	0.13	0.09	0.05	0.01	-0.03	-0.07	-0.11	-0.15	-0.19
q15	-0.27	-0.25	-0.23	-0.21	-0.19	-0.17	-0.15	-0.13	-0.11
q16	0.22	0.21	0.20	0.19	0.17	0.16	0.15	0.14	0.13
q17	0.02	0.02	0.01	0.01	0.01	0.00	0.00	-0.00	-0.00
q18	-0.54	-0.48	-0.42	-0.37	-0.31	-0.25	-0.20	-0.14	-0.08
q21	-0.15	-0.12	-0.10	-0.08	-0.06	-0.04	-0.01	0.01	0.03
q22	0.05	0.03	0.02	0.01	-0.00	-0.01	-0.03	-0.04	-0.05
q25	0.16	0.10	0.05	-0.00	-0.06	-0.11	-0.17	-0.22	-0.27

From the Table, we see that in the case of items 2 and 25, the marginal effects run from positive to negative. Thus, at low levels of political sophistication, changing the item from negative to positive leads to a more positive response. On the other hand, those with a high level of political sophistication would provide a more negative response. Thus, not only does the effect not disappear at high levels, but it also even flips to behave opposite from expected. Even for item 18, where the negative effect does decrease, it remains negative even at the highest level.

The other items fare not much better. Items 5, 17 and 22 are closest to 0 (no effect) but are stable for all levels of political sophistication. Other items are equally stable and are either positive or negative for each case. Finally, item 10 shows the opposite of what we expect - that is, values that tend towards zero at the lower end of political sophistication. Note though that as before the size of all these effects is small (-0.54 being the largest for item 18). Taken together, this means that we cannot confirm the second hypothesis.

Effect of Issue Salience

As per Hypothesis 3, the effect of the polarity should be smaller for questions with a high salience than for those with a low salience. Looking at Table 3 we find that in the three cases where the effect is significant, the salience is rather low: (0.2% for

questions 1 and 2, and 7.3% for question 18). Also, for the question with the highest salience (question 22, with 19.4%) the effect is both small and insignificant. At the same time, another question with high significance (question 25) does have a high salience, at 18.3%. Moreover, there is no significant correlation between the salience and the effect ($r(10) = 0.02, p = 0.94$). As such, we have to reject the third hypothesis.

Effect on Congruence

The fourth hypothesis asked if question polarity can lead to actual differences in congruence. In other words: does changing the polarity influence the degree of how close or far a respondent is from a party? For this, I calculated how often a party appeared as the “best match” for a respondent. This best match is also that party that appears at the top of the list of matches the respondent receives after filling out the VAA. To then see if polarity made a difference, Table 6 shows the differences between both cases.

Here we see that not only are there differences between the various parties, but there are differences between the items as well. Also, when comparing these results with Table 3, we find that the effect of polarity is not always a good predictor for the change in congruence. For example, while item 18 showed the highest effect of polarity (at -0.29), the changes in the best match are average. Yet, item 5, which did not show any effect of polarity, shows a large “swing” for the GreenLeft. That is, when switching from the negative to the positive version of that item (“Mortgage relief has to be abolished/maintained”), the party lost 18.64% of its best matches, with their loss being equally distributed over the other parties. Something similar occurs for the PVV in the case of item 11, where they lose around a third (34.49%) of their best matches. While in both cases this loss is equally shared between the other parties, in other cases, it is clearer which parties’ profit. For example, for item 25 (The multicultural society is a not good thing/is a good thing), where the PVV loses 26.86%, these losses benefit both CDA and CU in an equal manner.

On why certain parties gain and lose matches, the results are mixed. For item 25, we can argue that the cause of the losses of the PVV is that the party is often associated with the negative version of the item. The same goes for item 11 (“The Netherlands has to remain in/leave the European Union”), as the party is well known for favoring leaving the EU. Yet, in other cases, such as for the Green Left for item 5, or why the CDA and CU profit from the change for item 25, the results are less clear. Overall though, we can conclude that the polarity of the item does influence the congruence, though not always in a consistent manner. Thus, we can confirm the fourth hypothesis.

Table 6 Differences in the percentage of best match for each party, when switching from the negative to the positive version

	CDA	CU	D66	GL	PvdA	PVV	SP	VVD
q1	-2.21	-2.21	2.24	2.24	-2.32	2.24	2.24	-2.21
q2	5.86	5.86	-7.44	-7.44	5.86	-1.13	-7.44	5.86
q5	2.67	2.67	2.67	-18.64	2.67	2.67	2.67	2.67
q10	16.04	3.78	-6.78	-6.78	-15.85	16.04	-6.78	0.32
q11	4.72	4.72	4.72	4.72	4.72	-34.49	6.21	4.72
q15	1.24	1.86	4.22	1.86	4.22	-7.63	1.86	-7.63
q16	-9.05	11.28	-9.05	2.39	11.28	-9.05	11.28	-9.05
q17	6.63	-2.71	-2.71	-2.71	-5.88	6.63	-5.88	6.63
q18	-5.38	-5.38	3.81	3.81	3.81	1.25	-5.38	3.64
q21	1.16	-3.50	1.16	1.16	1.16	-3.50	1.16	1.16
q22	8.81	8.81	7.20	-8.01	-8.01	-8.01	-8.01	7.20
q25	13.48	13.48	-0.01	-0.01	-0.01	-26.86	-0.01	-0.06

Conclusion

One of the cornerstones of the field of issue voting are accurate measures of the congruence between voters and parties on issues. In this paper, I focused on one threat to the accuracy of these measures by looking at how the formulation of an issue can influence it. In other words, could one get alternative degrees of congruence by doing no more than altering the formulation of an issue? For this, I focused on a common alteration of the polarity of the issue, that is, whether an issue has a negative or positive formulation. After selecting the items, I provided two groups of respondents with either formulation using a Voting Advice Application, which launched during the 2017 elections in the Netherlands. This led to mixed results. While for some items there were significant differences between the responses for the two versions of the item, for others there were not. The same was true for the influence of political sophistication and issue salience: sometimes it was influenced by the effect of polarity, other times not. Yet, in all cases, the actual influence on the mean response was low. That is, even when the polarity led to a different mean response, this difference was very small (especially on a 1-5 scale). Yet, these small differences did often have a significant effect on the congruence between the party and the respondent. Thus, for some items, parties received a higher congruence with a respondent with a different formulation of an item.

These findings are interesting for several reasons. To begin with, that a different item wording can lead to different responses is well-known. Yet, what these differences are and whether they matter is much less studied. With regards to congruence, it is not only important to look at the actual change in the response, but also at the change in congruence. Here, we saw that even small differences in the mean response could lead to large differences in the congruence.

As such, these findings can find application in several fields. An obvious one is that of VAAs, which provide respondents with matches between them and various parties, based on the degree of congruence between them. Designers of VAAs should thus consider that not only can the initial selection of the items matter, but their formulation can also as well. Given the increasing evidence that VAAs can influence party choice, this makes it even more relevant for designers to pay close attention to the wording of the issues. Another complication for designers of VAAs is the conditional effect of political sophistication. In some cases, the effects of changing the polarity of the question were higher when the respondent had a lower level of political sophistication. This is a challenge for designers of VAAs as it is for those respondents that VAAs are most beneficial. As these findings show that they are vulnerable to such design choices, there should be an increased focus on these choices. This then leads to the question if they should or should not include negative questions at all. The problem here is that it is difficult *not* to include any negative questions at all. Some of the questions only exist in their negative form in the debate and including them in their positive form could be confusing for the respondents. Also, there is no reason to assume that negative questions are inherently problematic. They are only different from their positive counterparts. The best designers could do is at least to consider which questions to make negative and why. Besides, designers would do well not to use a certain wording if it would favor a certain party.

Apart from VAAs, we can also extend our conclusions to the measurement of congruence in general. In this case, the main conclusion is that congruence not only depends on the content of the item but also on its formulation, with the effect being influenced by the political sophistication of the respondent and the party one calculates the congruence with. Here we saw that while these differences may be small, their influence can be large. As such, ignoring the effect of the formulation of an item can lead one to draw conclusions based on measurement variation instead of substantive variation. Besides, changing the polarity of an item is a simple change. More rigorous changes - such as including or not including examples in the item - are likely to cause equal, or even larger, differences.

From here, there are several avenues for further research. The first one is to extend the current research to other countries. The country here – The Netherlands – is in many ways a unique case. The country has a representative system of government that is one of the most proportional in the world. As a result, there are a large

number of political parties (Lijphart, 1999). In such systems, parties are more likely to adopt issues as their own and stress them during the campaign (Kim, 2020). A second avenue is to carry out similar research, but also pay attention to the positions of the parties. In this study, the coders coded the party on only one version of the question. The position for the opposite wording was nothing less than the reverse of the score. Yet, given that parties might own the wording of certain issues it might be that the positioning of the party is different under different wording. This could explain at least some of the variations found here. A third avenue might be to change other aspects of the wording of the questions or the questionnaire. One example might be the effect of the number or the order of the response options the respondent can use. Another might be to what degree quantifiers or explanations show any influence. Fourth, one could repeat the experiment on dimensions instead of issues, as the wording of the dimensions is most likely affected in the same way.

That question wording is no neutral exercise is clear. Yet, the precise effects of it are often not clear. Here, I showed that even a simple aspect of question-wording could lead to changes in the size of the congruence between voters and parties. Moreover, parties can benefit when the question uses their favored wording. Thus, scholars working with congruence should take not only the effects of question-wording into account, but also realize that no wording can truly be neutral.

References

- Andreadis, I. (2014). Data Quality and Data Cleaning. In D. Garzia and S. Marschall (Eds.), *Matching Voters with Parties and Candidates: Voting Advice Applications in a Comparative Perspective* (pp. 79–92). Colchester: ECPR Press.
- Bassili, J. N. & Krosnick, J. A. (2000). Do Strength-Related Attitude Properties Determine Susceptibility to Response Effects? New Evidence From Response Latency, Attitude Extremity, and Aggregate Indices. *Political Psychology* 21 (1), 107–132.
- Belchior, A. M. & Freire, A. (2013). Is party type relevant to an explanation of policy congruence? catch-all versus ideological parties in the Portuguese case. *International Political Science Review* 34 (3), 273–288.
- Benoit, K. & Laver, M. (2006). *Party Policy in Modern Democracies*. London: Routledge.
- Blasius, J. & Thiessen, V. (2012). *Assessing the Quality of Survey Data*. London: Sage.
- Chessa, A. G. & Holleman, V. (2007). Answering attitudinal questions: Modeling the response process underlying contrastive questions. *Applied Cognitive Psychology* 21 (2), 203–225.
- Clark, H. H. (1976). *Semantics and Comprehension*. Den Haag: Mouton.
- Costello, R., Toshkov, D., Bos, B., & Krouwel, A. (2020). Congruence between voters and parties: The role of party-level issue salience. *European Journal of Political Research*.
- Dalton, R. J. (1985). Political parties and political representation. *Comparative Political Studies* 18 (3), 267–299.
- Dalton, R. J. (2017). Party representation across multiple issue dimensions. *Party Politics* 23 (6), 609–622.

- Diamond, L. & Morlino, L. (2005). *Assessing the Quality of Democracy*. Baltimore, MD: The Johns Hopkins University Press.
- Dinas, E. & Gemenis, V. (2010). Measuring Parties' Ideological Positions With Manifesto Data: A Critical Evaluation of the Competing Methods. *Party Politics* 16 (4), 427–450.
- Dolný, B. & Baboš, P. (2015). Voter–representative congruence in Europe: A loss of institutional influence? *West European Politics* 38 (6), 1274–1304.
- Downs, A. (1957). *An Economic Theory of Democracy*. New York, NY: Harper.
- Gemenis, K. (2015). An iterative expert survey approach for estimating parties' policy positions. *Quality & Quantity* 49 (6), 2291–2306.
- Gemenis, K. B., Bruinsma, C., Djouvas, V., Manavopoulos, & Mendez, F. (2017). *Stem-Consult: Voting Advice Application data for the 2017 parliamentary election in the Netherlands*. DANS. <https://doi.org/10.17026/dans-24r-225b>
- Gemenis, K. & van Ham, C. (2014). Comparing Methods for Estimating Parties' Positions in Voting Advice Applications. In D. Garzia and S. Marschall (Eds.), *Matching Voters with Parties and Candidates: Voting Advice Applications in a Comparative Perspective* (33–47). ECPR Press.
- Giger, N. & Lefkofridi, Z. (2014). Salience-Based Congruence Between Parties & their Voters: The Swiss Case. *Swiss Political Science Review* 20 (2), 287–304.
- Highton, B. (2009). Revisiting the Relationship between Educational Attainment and Political Sophistication. *The Journal of Politics* 71 (4), 1564–1576.
- Hippler, H.-J. & Schwarz, N. (1986). Not Forbidding Isn't Allowing: The Cognitive Basis of the Forbid-Allow Asymmetry. *Public Opinion Quarterly* 50 (1), 87–96.
- Holleman, B., Kamoen, N., Krouwel, A., van de Pol, J., & de Vreese, C. (2016). Positive vs. Negative: The Impact of Question Polarity in Voting Advice Applications. *PLOS ONE* 11 (10), 1–17.
- Holleman, B., van den Bergh, H., Mak, P., Sanders, T. & Kamoen, N. (2017). Why Are Negative Questions Difficult to Answer? On the Processing of Linguistic Contrasts in Surveys. *Public Opinion Quarterly* 81 (3), 613–635.
- Horn, L. R. (1989). *A Natural History of Negation*. Chicago, IL: University of Chicago Press.
- Hosch-Dayican, B., Aarts, K., Amrit, C., & Dassen, A. (2013). Issue salience and issue ownership online and offline: Comparing Twitter and survey data. Paper presented at the American Political Science Association 2013 Annual Meeting; 29 August - 01 September, Chicago, IL.
- Huber, J. D. & Powell, G. B. (1994). Congruence Between Citizens and Policymakers in Two Visions of Liberal Democracy. *World Politics* 46 (3), 291–326.
- Kamoen, N., Holleman, B., Mak, P., Sanders, T., & van den Bergh, H. (2011). Agree or Disagree? Cognitive Processes in Answering Contrastive Survey Questions. *Discourse Processes* 48 (5), 355–385.
- Katz, R. S. (1997). *Democracy and Elections*. Oxford: Oxford University Press.
- Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics* 38 (7), 1033–1050.
- Kim, Y. J. (2020). Issue Ownership and Electoral Rule: Empirical Evidence from Japan. *Government and Opposition* 55 (1), 147–162.

- Krosnick, J. A. & Presser, S. (2010). Question and Questionnaire Design. In P. V. Marsden and J. D. Wright (Eds.), *Handbook of Survey Research (Second ed.)* (pp. 263–313). Bingley: Emerald.
- Lijphart, A. (1999). *Patterns of Democracy*. New Haven, CT: Yale University Press.
- Louwerse, T. & Rosema, M. (2014). The design effects of Voting Advice Applications: Comparing methods of calculating matches. *Acta Politica* 49 (3), 286–312.
- Luskin, R. C. (1990). Explaining Political Sophistication. *Political Behavior* 12 (4), 331–361.
- Mair, P. (2001). Searching for the Positions of Political Actors: A Review of Approaches and a Critical Evaluation of Expert Surveys. In M. Laver (Ed.), *Estimating the Policy Positions of Political Actors* (pp. 10–30). London/New York, NY: Routledge.
- Mendez, F. (2014). What’s behind a matching algorithm? A critical assessment of how VAAs produce voting recommendations. In D. Garzia and S. Marschall (Eds.), *Matching Voters with Parties and Candidates: Voting Advice Applications in a Comparative Perspective* (pp. 49–66). Colchester: ECPR Press.
- Mendez, F. (2017). Modeling proximity and directional decisional logic: What can we learn from applying statistical learning techniques to VAA-generated data? *Journal of Elections, Public Opinion and Parties* 27 (1), 31–55.
- Mendez, F., Gemenis, K., & Djouvas, C. (2014). Methodological challenges in the analysis of voting advice application generated data. In 2014 9th International Workshop on Semantic and Social Media Adaptation and Personalization, pp. 142–148.
- Powell, G. B. (2004). Political Representation in Comparative Politics. *Annual Review of Political Science* 7 (1), 273–296.
- Powell, G. B. (2009). The ideological congruence controversy: The impact of alternative measures, data, and time periods on the effects of election rules. *Comparative Political Studies* 42 (12), 1475–1497.
- Rabinowitz, G. & MacDonald, S. E. (1989). A Directional Theory of Issue Voting. *The American Political Science Review* 83 (1), 93–121.
- Romeijn, J. (2020). Do political parties listen to the(ir) public? Public opinion-party linkage on specific policy issues. *Party Politics* 26 (4), 426–436.
- Rosset, J. & Stecker, C. (2019). How well are citizens represented by their governments? issue congruence and inequality in Europe. *European Political Science Review* 11 (2), 145–160.
- Schuld, J. P., Konrath, S. H., & Schwarz, N. (2011). “Global warming” or “climate change”? Whether the planet is warming depends on question wording. *Public Opinion Quarterly* 75 (1), 115–124.
- Thomassen, J. (2012). The Blind Corner of Political Representation. *Representation* 48 (1), 13–27.
- van de Pol, J., Holleman, B., Kamoen, N., Krouwel, A., & de Vreese, C. (2014). Beyond Young, Highly Educated Males: A Typology of VAA Users. *Journal of Information Technology & Politics* 11 (4), 397–411.
- van der Eijk, C. (2001). Measuring Agreement in Ordered Rating Scales. *Quality & Quantity* 35 (3), 325–341.
- van der Meer, T. W. G., van der Kolk, H., & Rekker, R. (2017). Dutch Parliamentary Election Study 2017 (DPES/NKO 2017). DANS.