Public opinion without opinions? Item nonresponse and (the absence of) substantive opinions in public opinion surveys

Maat, J. van de

CHAPTER 3

Hypotheses on Non-Substantive Response Options

3.1 Introduction

In the following sections, hypotheses will be developed which follow from theory and previous findings. These hypotheses will be tested with the original data from survey experiments. The section details hypotheses about question design effects in general, which are expected to affect all three survey experiments, and specific hypotheses which only apply to a single experiment. Overall, the key research question refers to the effect of manipulating non-substantive response options. Additional expectations about the follow-up question and its effect on respondents giving (non)substantive answers are developed in chapter 7.

3.2 Hypotheses for All Survey Experiments

3.2.1 Question Design and Level of Item Nonresponse

The first expectation is about the relation between question design options and the level of item nonresponse. This hypothesis refers to the relation between question design and missing substantive answers. The first and rather obvious expectation is that offering a DK option or using a filter question explicitly will result in more item nonresponse.

Logic dictates that if a non-substantive response option is not offered or only implicitly or at least less explicitly, less item nonresponse occurs. An example of an implicit way of offering a non-substantive response option is the possibility to skip a question, without explicitly mentioning this possibility in the question or offering an explicit (don't know) response category. Compared to variants in which a non-substantive response option is part of the question and/or explicitly mentioned and showed as a response option, the respondent will be less inclined to use such an implicit option (e.g. Bishop, 2005; Schuman & Presser, 1979). The explicitness of the non-substantive response option affects the item nonresponse rate.

> *H1a: The more explicit a non-substantive response option is presented, the more item nonresponse will be measured.*

The second hypothesis about question design and item nonresponse concerns the type of non-substantive response option. Two explicit non-substantive response options are tested: the don't know (DK) option and the filter question. Previous research indicates that in general the use of a filter question results in about 20 to 25 percent item nonresponse (see e.g. Bishop, 2005; Bishop et al., 1983; Schuman & Presser, 1979). Using a DK option in web surveys results in item nonresponse

somewhere between 10 and 20 percent, depending on issue content and the form of the questions (see e.g. Couper et al., 2001, p. 247; Fricker et al., 2005, pp. 387-388; Heerwegh, 2009, pp. 115-116; Heerwegh & Loosveldt, 2008, p. 842). The expectation is that since a filter question is posed before the substantive opinion question and explicitly asks whether a respondent has an opinion, respondents are more likely to use that non-substantive response option than when a DK option is offered as part of the opinion question.

H1b: A filter question results in more item nonresponse than an explicit DK option.

3.2.2 Question Design, Missing Data and Distribution of Opinions

After exploring the relationship between the implicit or explicit way a non-substantive response option is offered and the amount of missing data, the next step is to look at the relation between missing data and the overall distribution of substantive opinions. Missing data may be considered problematic because of a loss of potential information, but it is in particular problematic when it results in bias (De Leeuw et al., 2003). A variant with missing data is compared to a forced choice question variant which does not generate missing data. The answers in the forced-choice variant contain valuable information (Hippler & Schwarz, 1989; Krosnick, 1999) and can be used for comparison since there is no item nonresponse and no nonresponse bias. For the versions of the questionnaire where a non-substantive response option is offered, a higher item nonresponse rate, 'provides greater opportunities for bias' (Lynn, 2014, p. 319), which is why versions with different response rates will be compared as regards their substantive outcomes. What is meant by 'substantive outcomes' is the overall picture of public opinion resulting from a survey. The question is whether one would paint a different picture of public opinion and the public's preferences when a different non-substantive response option is used.

If the data are missing at random for the response categories, no differences will be found between the various variants of the questionnaire in terms of bias. If the data are, however, not missing at random and item nonresponse is not equally distributed over the available substantive response categories, the overall distribution of opinions varies between the variants due to item nonresponse bias and results in a different impression of public opinion. Do majorities or pluralities change or disappear when a question design is used which measures more item nonresponse? Two hypotheses will be tested, based on the data 'missing at random' (MAR and MCAR) and data 'not missing at random' (NMAR) premises respectively (De Leeuw et al., 2003, p. 155; see also Lynn, 2014).

H2a (based on MAR): An increase of the level of missing data does not affect the distribution of opinions.

Alternatively H2b (based on NMAR): An increase of the level of missing data results in a different distribution of opinions.

3.2.3 Question Design and Question Content

After establishing the relationships between question design and item nonresponse and the distribution of opinions, the next step is to move beyond general effects into a more detailed comparison of the content of the question. Here the focus is on the effects of question design for specific items on item nonresponse. The key question is whether there is a relation between the topic and content of the question and the level of item nonresponse resulting from different questionnaire variants.

The first hypothesis, which also formed a starting point for issue selection, is inspired by the general idea that people may have opinions about almost everything, but there are some core issues that they feel much more strongly about (McClosky & Zaller, 1984; Wittkopf, 1990). Respondents may not have a strong opinion about each and every issue and some opinions may seem inconsistent, but there are probably 'underlying principles' (Feldman, 1988, p. 416) structuring opinions. Such 'core beliefs', 'predispositions' or 'general orientations' (see e.g. Everts & Isernia, 2015, p. 35; Feldman, 1988) arguably give a more structured and stable view of public opinion. Various dimensions or cleavages may organize individual attitudes; they arguably are also central to public and political debate. Therefore, one may expect people to be more aware of issues that are directly related to a major dimension, which subsequently results in more substantive opinions and less item nonresponse. For this study, the main dimensions or cleavages in Dutch politics are socio-economic, ethical or moral, and multicultural (Aarts & Thomassen, 2008; Pellikaan, 2010; Pellikaan, de Lange, & Van der Meer, 2007). These dimensions and issue selection are discussed more extensively in the Data and Methods chapter.

H3a: If the topic of a survey question is related to a major political dimension, then the item nonresponse is lower compared to a survey question that is not related to such a dimension.

There seems to be consensus (e.g. Alvarez & Brehm, 2002, p. 214; Everts, 2008, pp. 8-14) about the fact that with respect to foreign policy opinions are less coherent, less stable, and less informed (but see Marquis & Sciarini, 1999). Even though the picture may not be as desolate as was once thought (Aldrich, Gelpi, Feaver, Reifler, & Sharp, 2006; Holsti, 1992), and Everts and Isernia (2015: 35) noted 'a revisionist wave (...) that challenged the pessimistic view of the public', many scholars still

agree that foreign policy opinions are relatively lacking and highly volatile. This ties in with Tiemeijer's (2006, pp. 92-95) suggestion that the general public should only be asked about issues that refer to central values and/or issues that have to do with the individual's personal environment, because such issues do not require 'tacit knowledge' and these opinions are less likely to change since they are rooted in personal experience. Foreign policy issues require specific knowledge and are out of the realm of the average citizen; it is expected that such foreign policy issues result in a higher nonresponse rate.

H3b: The item nonresponse for questions about foreign policy issues is higher than for questions about issues related to the core dimensions.

3.2.4 Question Design and Response Categories

When looking at variations in item nonresponse rate and their impact, one should look further than only at the non-substantive response options and include the substantive response categories. In general, 'a larger number is better than just two response categories (...) Four to seven categories are optimal' (de Leeuw, 2001, p. 153; see also Krosnick & Fabrigar, 1997; Leigh & Martin, 1987). Alwin (1997) compared a 7-point and 11-point scale and showed that a larger number of response categories results in a more precise measurement, but Kroh (2007, p. 210) states that 'too many scale points may also reduce data quality', because it may not be clear to respondents how the points differ from each other and deciding what to answer takes more cognitive effort. In political knowledge tests, the number and nature of plausible distractors is a central concern (Haladyna & Downing, 1993; Owen & Froman, 1987). This is an important element of the survey design, because one answer is correct and it is unclear when respondents get confused or when it is too easy to guess right.

When measuring attitudes, the number of response categories is also important, but in a different way. Respondents need enough variation to cover the intensity or subtleties of their opinion and will get frustrated when such variation is not available. That does not mean the more the better, but sufficient variation should be offered. In this study, the general expectation regarding response categories, regardless of the way a non-substantive response option is offered, is that more substantive choice results in less item nonresponse.

H4a: The more substantive response categories are offered, the lower the item nonresponse rate.

A second relevant element with respect to response categories is whether a neutral or middle response category is offered. A neutral category should arguably only be offered in a bipolar response scale (de Leeuw, 2001), but whether it is offered at all is another point of debate. A neutral answer may validly reflect a respondent's neutral stance on an issue, but it may also hide what is essentially a nonattitude or nonresponse (Inglehart & Klingemann, 1976; Kroh, 2007; Schuman & Presser, 1996). According to Raaijmakers *et al* (2000, p. 208), the meaning of this middle category can be grouped into two subcategories: '(1) True neutral meanings besides "neither/nor", such as "neutral" and "indifferent" (...), and (2) meanings that refer more to a kind of nonresponse, such as "undecided", "don't know", "never thought about it", and "no opinion"'. Research shows that both subcategories do exist, with less item nonresponse when a middle response category is offered and more use of the middle of the scale in the absence of a non-substantive response option (Ayidaya & McClendon, 1990; Lambert, 1983; Presser & Schuman, 1980).

It is expected that for survey questions with a middle response category at least part of the item nonresponse is substituted by a middle answer; in the absence of a midpoint category respondents would use the non-substantive response option and vice versa. Part of the item nonresponse may hide midpoint answers whereas midpoint answers may indicate an absence of opinion, particularly when no non-substantive response option is offered. In the absence of a non-substantive response option, for example in a forced choice variant, it is expected that respondents use the midpoint option as a non-substantive response option.

H4b: A midpoint in the absence of a non-substantive response option results in more use of this midpoint option than when a non-substantive response option is offered.

H4c: A midpoint combined with a non-substantive response option results in less item nonresponse as compared to offering no midpoint category.

3.3 Specific Hypotheses for Individual Experiments

3.3.1 Response Time

In chapter 5 (about the Don't Know option), the response time of respondents is examined as an indicator of how the DK option is used. Response time or completion time (Malhotra, 2008) is a specific form of paradata, which can be easily gathered with a web survey. Paradata or process information or additional data are 'data generated in the process of conducting a survey' which can help analyze survey errors and survey costs (Kreuter, 2013: preface). Since the rise of the web

survey, the amount of paradata has exponentially increased and the interest in analyzing paradata has followed suit. Besides response time, other types of paradata include the tracking of eye movements, mouse movements and the registration of a respondent opening other (internet) windows while conducting the web survey (Revilla & Ochoa, 2015, p. 98).

There are two strands of thought regarding response time: 1) 'shorter response times indicate stronger attitudes and measurement of these attitudes are less affected by question order or response order' or 2) 'response time is an indicator of the amount of cognitive effort invested in solution behavior' (Callegaro, Yang, Bhola, Dillman, & Chin, 2009, p. 6). In both cases, response time is an indicator of cognitive processing in answering a survey question (Bassili & Fletcher, 1991), but does a shorter response time reflect better or worse data? Revilla and Ochoa (2015, p. 109) argue that 'a worse quality of answers is directly related with shorter RT, that is, with more speeding'; Zhang and Conrad (2013) find a relation between speeding and straightlining, i.e. respondents 'giving nondifferentiated (identical) ratings to a series of questions with the same answer choices' (Schonlau & Toepoel, 2015, p. 125), which also points towards the second strand of thought. Moreover, 'particularly quick responses, the so-called speeding, 'might indicate minor data quality' (Greszki et al., 2014, p. 238). The assumption here is that answering a question is a time-consuming process and quick responses indicate that the respondent has not completed all steps of the response process model (Greszki et al., 2014, 2015; Yan & Tourangeau, 2008). Following this line of reasoning, it is the expectation (hypothesis H5) that the more explicit a DK option is offered, the less response time will be registered.

H5: The more explicit the DK option is presented, the less response time will be registered.

3.3.2 Break-offs

In chapter 6 (about filter questions), a hypothesis is added which concerns partial nonresponse caused by respondents breaking-off the survey, also called drop-outs (Peytchev, 2009). The analysis of break-offs is not included in the two chapters, because of a lack of data. The filter question experiment does provide data about break-offs. By analysing when and how many respondents drop out of the survey, the question whether forcing respondents to answer survey questions (rather than providing them a non-substantive response option) results in more break-offs can be addressed. A breakoff occurs 'when a respondent starts the survey but stops prior to completing it' (Peytchev, 2009, p. 74). Missing data in this case are not missing at random, but are 'time dependent' (De Leeuw et al., 2003, p. 157) since it concerns data after a certain point in time, i.e. when the respondent dropped out.

Many factors may influence the decision to drop out, including question characteristics like the length of the questionnaire and the content of the questions, survey mode and (cognitive) burden (Crawford, Couper, & Lamias, 2001; de Leeuw, 2001; Peytchev, 2009). At the respondent level, motivation plays a role in predicting break-offs (Steinbrecher, Roßmann, & Blumenstiel, 2015). Web surveys are particularly prone to break-offs, because respondents decide on participating after scanning some questions. The introduction by an interviewer is the decisive factor in face-to-face and telephone surveys; in web surveys the respondent uses the questionnaire itself to decide whether s/he wants to participate (Ganassali, 2008, p. 22; Vicente & Reis, 2010, pp. 253-254). Even when only one (introductory) screen is viewed by the respondent, a drop-out is registered (Ganassali, 2008, p. 25); the respondent did decide to open the survey and is therefore not considered unit nonresponse.

Factors influencing break-offs in web surveys generally relate to perceived survey burden. *Perceived* survey burden relates to the fact that it is not (only) the actual time and effort necessary to complete the questionnaire, but the perception the respondent has of what is required. Crawford *et al (2001)*, for instance, conclude that more respondents proceed after the first (introductory) screen when less estimated time is reported – even when the actual survey completion time is higher. These factors influencing break-offs include questionnaire length and the use of a progress indicator; if a survey seems to require more time and effort, more respondents drop out (Conrad, Couper, Tourangeau, & Peytchev, 2010; Crawford et al., 2001; Galesic & Bosnjak, 2009). It is also an established fact that question design elements influence the number of break-offs in web surveys, including the question format (open or closed) and visual aspects like radio buttons (Vicente & Reis, 2010, pp. 260-262). The effect of non-substantive response options on partial item nonresponse is however unknown.

The empirical question is when holding all other factors constant – e.g. questionnaire length, question order and content, and number of substantive response categories (see Bosnjak and Tuten, 2001) – whether the number of break-offs is higher in variants where no non-substantive response option is offered explicitly. The mechanism is that if respondents are unable to show they don't have an opinion, this results in frustration and ultimately break-offs.

H6: When respondents are forced to answer survey questions, the number of break-offs is higher than when a non-substantive response option is available.

The hypotheses are summarized in Table 3.1.

Table 3.1: Summary of Hypotheses

			In chapter...
Question design	H1a	The more explicit a non-substantive response option is presented, the more item nonresponse will be measured	5, 6
	H1b	A filter question results in more item nonresponse than an explicit DK option	6
Missing data	H2a	(Based on MAR) An increase of the level of missing data does not affect the distribution of opinions	5, 6
	H2b	(Based on NMAR) An increase of the level of missing data results in a different distribution of opinions	5, 6
Question content	H3a	If the topic of a survey question is related to a major political dimension, then the item nonresponse is lower compared to a survey question that is not related to such a dimension	5, 6
	H3b	The item nonresponse for questions about foreign policy issues is higher than for questions about issues related to the core dimensions	5, 6
Response categories	H4a	The more substantive response categories are offered, the lower the item nonresponse rate	5, 6
	H4b	A midpoint in the absence of a non-substantive response option results in more use of this midpoint option than when a non-substantive response option is offered	5, 6
	H4c	A midpoint combined with a non-substantive response option results in less item nonresponse as compared to offering no midpoint category	5, 6
Response time	H5	The more explicit the DK option is presented, the less response time will be registered	5
Break-offs	H6	When respondents are forced to answer survey questions, the number of break-offs is higher than when a non-substantive response option is available	6