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

Maat, J. van de

CHAPTER 2

Doing Surveys and Question Design Effects

2.1 Introduction

In the Introduction, a very brief overview was given of potential problems in survey research to set the stage for the puzzle and research question. In this chapter the process of doing surveys is discussed in more detail. The discussion consists of three parts: 1) setting up and executing a survey; 2) answering survey questions; and 3) results. The discussion of the three parts follows the course of how surveys are done: the survey is designed first and then executed; during the execution the respondents have to think about the survey questions and what to answer; and finally, the results (in terms of nonresponse and substantive opinions) are analysed and published. The part about setting up and executing a survey includes potential methodological problems and a more in-depth examination of one problem that is central in this study: question-design effects. The second part, about answering survey questions, briefly describes how respondents answer survey questions and whether they have opinions. In the third and final part of this chapter the outcome of a survey is discussed, i.e. (item) nonresponse and the distribution of opinions.

Note that the overview below is not exhaustive and merely provides context for the survey experiments examining the relation between a specific problem, i.e. question design, and the outcome of a survey. The chapter consequently pays more attention to aspects that are central to the survey experiments, i.e. question design (as an independent variable that is manipulated in the three survey experiments), item nonresponse or non-substantive answers and distributions of opinions (as dependent variables).

2.2 Methodological Issues in Doing Survey Research

This section contains various elements that have to be considered when doing surveys and addresses a specific problem central to this study: question design (effects).

2.2.1 Potential Survey Methodological Problems

A 'maximizing function of separate individual wills' is 'the most common conception of public opinion' (Price, 1992, p. 13; 22). Surveys have become the dominant means to assess this public opinion. Surveys aggregate individual opinions to explore and express public opinion regarding a specific issue. Initially, at least for some, the hope was that polling would function 'as a technological means for advancing quality in collective decision making' (Price & Neijens, 1997, p. 352). Surveys are assumed to constitute a linkage between the public and their decision-making representatives,

making responsiveness towards the public easier. This survey method of collecting opinions is, however, vulnerable since 'public opinion is created by the procedures that are established to "discover" it. [It] is an artifact of the technical procedures that are designed to capture it' (Osborne & Rose, 1999, p. 382).

The organization and design of surveys is an essential part of measuring or creating public opinion via surveys. Various elements need consideration before designing a survey, while executing it and afterwards. The methodological considerations can be grouped into two categories: the sample and the design of a survey. The sample refers to the particular selection of people answering the questions; the design refers to the way the information is collected from these people (Fink, 2009, p. 5; Fowler, 2014, pp. 9-11). The researcher needs to decide about a target population, the size of a sample and sampling method, survey mode, pre-testing questions, the lay-out of the questionnaire, question order, using open- or closed-ended questions, question formats, scale formats, and the choice of response categories (Bethlehem & Biffignandi, 2011; Bradburn et al., 2004, pp. 283-314; De Leeuw et al., 2008). This list is by no means exhaustive[6] but already shows that there are many things to consider when doing survey research; many things can go wrong.

The Total Survey Error approach has become the dominant paradigm in survey research since the 1990s. 'A full statement of the total survey error approach requires consideration of survey errors, survey constraints, and survey-related effects' (Weisberg, 2005, p. viii) and includes both sampling errors and nonsampling errors (Biemer, 2010a, 2011). What is important in the Total Survey Error approach is that it stresses that there are various elements to be considered when designing, executing and analyzing a survey. When designing a survey, *all* of these elements should be addressed.

The methodological problems in survey research were introduced in the previous chapter. These four sources of data error, as described by De Leeuw *et al* (2008, pp. 6-13), are: coverage of the population, sampling, nonresponse and measurement error. These aspects are relevant for all surveys, but in particular with respect to web surveys – which is the mode of data collection in this study. Particularly the undercoverage of people without an internet connection and the self-selection of respondents are common to internet panels, which are often used for web surveys (Bethlehem, 2013, pp. 9-16; Bethlehem & Biffignandi, 2011). Nevertheless, it is true for all types of surveys that there are many elements to be considered and potential problems may arise. One of those elements is how to design survey questions.

6 The list of elements to consider when doing surveys also only applies to the set-up of the survey. During the execution and afterwards, when the data are analyzed and presented, other problems may arise and a check of data quality is needed. These issues are, however, not part of this study.

2.2.2 Question Design (Effects)

The design of survey questions has a substantial impact on data quality (Fowler & Mangione, 1990; Sudman & Bradburn, 1974): 'From the perspective of total survey design, investing in the design and evaluation of questions is a best buy, one of the endeavors that is most likely to yield results in the form of better, more error-free data' (Fowler & Cosenza, 2008a, p. 375). This is almost a truism: if the survey results are not valid, their results are essentially worthless; in that case the instrument to gather individual opinions does not accurately reflect public opinion. 'A good question is one that produces answers that are reliable and valid measures of something we want to describe' (Fowler & Cosenza, 2008a, p. 376). It is therefore crucial to examine how question design, i.e. the way questions are asked and response alternative are offered in surveys, affects the outcome.

'The way the questions are asked' is a general description for various factors affecting the outcome of surveys. If the way the questions are asked affects the outcome of a survey or poll, these factors influencing the outcome are called design effects. Various factors can be distinguished, including questionnaire length, the choice of response categories, question wording and the order of questions within a questionnaire[7] (Burchell & Marsh, 1992; Couper, Traugott, & Lamias, 2001; Galesic & Bosnjak, 2009; McFarland, 1981; Moore, 2002; Poe, Seeman, McLaughlin, Mehl, & Dietz, 1988; Revilla, Saris, & Krosnick, 2014; Sigelaman, 1981; van Vaerenbergh & Thomas, 2013). In a strict interpretation these question design effects do not include the order of the questions, the order of the responses or the content of the questions; such factors are considered to be part of the questionnaire design but not of the *question* design.

One of the classic collection of studies in the field is by Schuman and Presser, who ask 'how the ways in which attitude questions are asked in surveys affect the results derived from these same surveys' (Schuman & Presser, 1996, p. 2). Five categories of characteristics are examined: open versus closed-ended questions, the use of the Don't Know (DK) option, the use of a neutral or midpoint response category, balanced and unbalanced questions, and attitude strength. These aspects are used in a systematic empirical analysis of how question form, wording and context affect survey results. Based on Schuman and Presser's findings that all these question design choices can and do have impact on survey results, Bishop draws the rather pessimistic conclusion that '[percentages in poll reports] may represent mostly how

7 In addition to these factors, the layout may affect survey results. Particularly in web surveys many options are available and the researcher needs to make decisions about how many items to put on a web page, what colours and buttons to use, the placement of answer scales and many other elements (Ganassali, 2008; Peytchev, Couper, McCabe, & Crawford, 2006).

the reality of public opinion gets constructed through the way in which the questions are framed, worded, and presented to respondents' (Bishop, 2005, p. 67).

That the design of individual questions affects the responses and subsequently the outcome of a survey is an established fact (e.g. Bradburn et al., 2004; Carpini & Keeter, 1993, pp. 1181-1184; Schuman & Presser, 1996)[8]. The aim of this study is to look at the effect of non-substantive response options on item nonresponse and consequently the distribution of opinions. Item nonresponse can be assessed in a number of ways, e.g. offering an (explicit) answer category that captures the absence of an opinion or posing an explicit filter question before the substantive question itself. The DK option and filter question are often treated as variants of a DK or No Opinion filter (e.g. Krosnick & Presser, 2010; Schuman & Presser, 1996). Both are question design elements intended to capture item nonresponse, i.e. respondents not having or giving a substantive answer (an opinion) in response to a particular question.

Using non-substantive response options may affect the survey results in two ways: a higher item nonresponse rate, i.e. more non-substantive answers, and a different overall distribution of opinions. These two effects are discussed elsewhere in this chapter. It should be mentioned at this point that the level of item nonresponse and the resulting distribution of opinions may vary according to the content of the question and the characteristics of the respondent (e.g. Stern, Dillman, & Smyth, 2007; Toepoel & Van Soest, 2009; Tourangeau, Couper, & Conrad, 2007). The goal here, however, is to look at systematic effects of applying a certain question design, i.e. the use of particular non-substantive response options. A more extensive discussion of the individual non-substantive response options can be found in the three empirical chapters.

2.3 Answering Survey Questions

Although this study examines question design effects and not explanations of response patterns of individual respondents in general, the process of answering survey questions needs to be addressed. Some information on 'the psychology of asking questions' (Schwarz, Knäuper, Oyserman, & Stich, 2008) is needed, because

8 There is a body of literature about how to write good survey questions (e.g. Bradburn et al., 2004; Fowler & Cosenza, 2008a; 2008b) and how to evaluate their quality. The evaluation of survey questions should take place prior to actual data collection, by doing pre-tests or cognitive interviewing (Krosnick & Presser, 2010; Presser et al., 2004). Another option is to use the Survey Quality Prediction system, a computer program that systematically assesses survey questions (Saris & Gallhofer, 2007). Both the guidelines and the methods of evaluating survey questions are aimed at improving the measurement of public opinion with surveys.

without knowing anything about how respondents answer survey questions, 'the art of asking questions' (Payne, 1951) is a useless exercise. Furthermore, if question design affects survey results at the *aggregate* level, it happens because *individual* respondents give different answers to questions. Hence the need for a brief exploration of how respondents answer survey questions and why question design impacts on the quality of survey answers (see e.g. Krosnick & Presser, 2010).

2.3.1 How Do Respondents Answer Survey Questions?

Since the outcome of surveys and polls is increasingly considered to be the public's opinion, it should be clear how individual members of this 'public' answer survey questions and how opinions are formed. Surveys aggregate individual opinions by aggregating responses to opinion questions; opinions are considered to be 'observable, verbal responses' to a specific issue or question (Price, 1992, p. 46). Whatever answer the respondent gives to a survey question is by definition regarded as an opinion. There is a vast body of literature about how individual opinions are *formed*. The psychological process of opinion formation is discussed elsewhere (e.g. Schwarz & Sudman, 1996; Tourangeau, Rips, & Rasinski, 2000; Zaller & Feldman, 1992). In this study, the substantive answers to survey questions are treated as opinions.

The generally agreed upon model of how respondents answer closed or pre-coded survey questions (e.g. Krosnick & Presser, 2010; Schwarz, 2007; Tourangeau et al., 2000) consists of a number of steps: 'Understanding the question, recalling information, forming a judgment, formatting the judgment to fit the response alternatives, and editing the final answer' (Schwarz et al., 2008, p. 19). The researcher decides upon a certain question(naire) design in order to simplify and standardize the respondent's interpretation of the individual survey questions. Question design can aide or hinder the process of understanding or interpreting this question (Fowler & Cosenza, 2008a; Groves, 2004, pp. 419-420; Krosnick & Presser, 2010). For example, it is important for data quality that the response alternatives fit the respondent's opinion or judgment[9]. This fit of response alternatives includes the number of response options or 'scale length' (Krosnick & Presser, 2010, p. 268), the use of a midpoint option (Raaijmakers, van Hoof, 't Hart, Verbogt, & Vollebergh, 2000; Tourangeau, Couper, & Conrad, 2004) and the inclusion of non-substantive response options. The point is that question design is an important part of how survey questions are interpreted and answered and subsequently of the individual results and overall picture of public opinion.

9 The assumption is that closed questions are used. Open questions have a number of advantages (Krosnick & Presser, 2010, pp. 266-268; Schuman & Presser, 1996; van Holsteyn 1994), but since closed questions are more common in public opinion surveys, open questions are excluded from this study.

The rather stylized model of answering survey questions is, however, the 'optimal' process of answering survey questions, which requires considerable cognitive effort. Respondents may and in practice often do try to relieve their cognitive burden by putting less effort into each step of the model, or even just select 'a reasonable answer' (Krosnick & Presser, 2010, p. 265). Respondents can employ a response strategy of 'satisficing' rather than 'optimizing' (Krosnick, 1991, 1999; Krosnick & Presser, 2010). Krosnick (Krosnick, 1991; Krosnick & Presser, 2010; Vannette & Krosnick, 2014) argues for a continuum of 'satisficing' which varies in strength: 'weak satisficing' means that respondents are less thorough in going through the steps of the answering process, whereas 'strong satisficing' means that respondents may skip some steps all together and just select an answer. The general point is that 'satisficing' implies that less cognitive effort is put into the process of answering survey questions.

Zaller (1992; Zaller & Feldman, 1992) gives a fundamentally different account of how respondents deal with survey questions. He does not assume that respondents recall information to form a judgment in response to a survey question. His suggestion is that people do not have an opinion beforehand, but that they have several considerations and views available. When asked for an opinion, the view 'on top' is expressed. This view is, however, not completely random or accidental and is influenced by attention for the issue in the media, the phrasing or ordering of the questions and recent personal experiences the respondent might have had related to the issue: '(...) people respond on the basis of whatever considerations are most immediately salient in their minds. The reason that their survey responses are unstable from one interview to the next is that what is at the top of a person's head varies stochastically over time' (Zaller, 1992, p. 365). Zaller's influential perspective on opinion formation and change has developed somewhat in later years, but his model for opinion formation at the individual level – the Receive-Accept-Sample model – has not essentially changed.

Converse (1964) goes one step further than Zaller and argued that a large part of the mass[10] simply does not have an opinion or only holds opinions on specific issues, Moreover, if respondents sometimes do have opinions, these opinions do not fit a coherent pattern or belief system. 'The individual lacks the contextual grasp to understand that the specific case and the general principle belong in the same belief system: in the absence of such understanding, he maintains psychologically independent beliefs about them' (Converse, 1964, p. 230). According to Converse, people give an answer to a survey question even when they do not have an established

10 Converse talks about 'the mass' to distinguish it from 'the public' that would have (consistent) opinions, according to his definition.

substantive opinion: they want to be 'good' and cooperative respondents. This results in nonattitudes, which means that some people 'have no opinion and just pick a response alternative by chance' (Van der Veld & Saris, 2004, pp. 37-38)[11]. Such on the spot 'opinions' formulated as a response to survey questions are called nonattitudes by Converse (1964) or pseudo-opinions by Bishop (Bishop, Oldendick, Tuchfarber, & Bennett, 1980) and these are problematic, because they threaten the validity and consequently the quality of survey data. Rather than valid answers to survey questions which reveal the policy preferences and positions of the public, the occurrence of nonattitudes means that at least part of the public provides answers that are not thought through or uninformed.

Expressing nonattitudes in a survey is related to the individual's information and knowledge about the content of the survey question: 'The likelihood of nonattitudes is inversely related to the level of political information and awareness' (Saris & Sniderman, 2004, pp. 1-2). There are several reasons why information or knowledge is important for the measurement of opinions. 'First, information reduces uncertainty and can persuade (...) Second, information makes predispositions and values relevant for beliefs about policy issues' (Alvarez & Brehm, 2002, p. 50). Hence the common inclusion in the analyses of indicators of political knowledge and information as control variables. Visser *et al* (2008), for example, find that a lack of knowledge results in less stable or even absent opinions. Krosnick (1999, pp. 548-549) shows that satisficing happens more often if the respondent has fewer abilities (which are influenced by the individual's knowledge and information) to answer survey questions. Other factors which increase the likelihood of satisficing are 'task difficulty' and 'motivation to optimize' (Krosnick, 1999, pp. 548-549).

Several attempts have been made to measure more improved, informed, 'better' public opinion, for instance by excluding respondents who do not have enough knowledge and/or do not have an opinion or by informing respondents beforehand; see the deliberative poll by Fishkin (1991) and various simulation models (e.g. Althaus, 1996; Carpini & Keeter, 1997). These models try to assess what public opinion would be if everyone was well and equally informed, see Sturgis (2003) for a comparison of these models. Another device to collect more informed opinions is 'the Choice Questionnaire' by Neijens *et al* (1992). However, it takes more time, effort and money to conduct surveys with these techniques and these designs are only applied occasionally. Furthermore, such surveys do not measure what public opinion is, but what it could be. Another argument for including all citizens regardless of their

11 Other authors argue that Converse is too pessimistic and that his findings about nonattitudes and incoherent belief systems could be attributed to other factors, like measurement errors (see Converse, 2000; Kinder, 1998; Smith, 1984).

level of knowledge is that they are also allowed to vote at elections. 'If the people are too ill informed to take their views into account as measured by polling, then why let them have their input at the time of the vote?' (Newport, 2004, p. 119). In other words: if there is no requirement in terms of knowledge during elections, it could be argued that there also should be no requirements with regards to knowledge or information levels needed to participate in public opinion surveys. This study will therefore focus on 'ordinary' and straightforward surveys, since these are used most often to gauge and represent public opinion.

Why do respondents use a non-substantive response option? The first reason is likely that they actually do not have an opinion. These are the item nonresponses the researcher wants to encourage and collect in order to avoid nonattitudes and less valid and reliable survey data. Other reasons for using a non-substantive response option include vague or unclear question wording, a lack of suitable response options, satisficing or the reluctance to reveal opinions about sensitive issues[12] (Groves, 2004, p. 156; Krosnick, 1999, pp. 556-559; Krosnick & Presser, 2010, pp. 283-284; Shoemaker et al., 2002). Besides respondents' reasons to refuse to answer survey questions, nonresponse is also registered when respondents overlook survey items or are unable to respond or their answers are due to technical or administrative errors not registered (De Leeuw, Hox, & Huisman, 2003, pp. 158-159; Groves, 2004, p. 156). This list is probably not exhaustive, but it illustrates the point that non-substantive response options may be used by respondents for many reasons – not only to diminish nonattitudes. In this study, the reasons for giving a non-substantive answer are not examined; item nonresponse is treated as a given.

2.4 Survey Results 1: Nonresponse in (Web) Surveys & Missing Data

Generally speaking, unit nonresponse is the result of a selected respondent being unable or unwilling to participate in a research (a survey). This particular form of nonresponse refers to 'the failure to obtain measurements from all units in the sample' (Hox & De Leeuw, 1994, p. 329). Since complete response is virtually impossible and may not even be desirable (see Stoop, 2005), all researchers have to deal with unit nonresponse. Unit nonresponse and item nonresponse (Groves & Couper, 1998)

12 This potential 'social desirability bias' means that respondents are for some reason reluctant to reveal their 'real' opinion or to admit to having no opinion. In these circumstances the answers are affected by the sensitivity of questions about desirable behavior and personal preferences (Bradburn, Sudman, Blair, & Stocking, 1978; Shoemaker, Eichholz, & Skewes, 2002). Due to the mode of the survey and the absence of interviewers, however, no major social desirability bias is expected (Heerwegh, 2009, pp. 112-113; Tourangeau & Yan, 2007).

are both discussed below, but most attention is paid to item nonresponse. This discussion of item nonresponse aims to further develop the concept, to review its application in survey (methodological) research and to consider whether item nonresponse should be treated as 'missing data'.

2.4.1 Unit Nonresponse in Surveys

Unit nonresponse consists of all units or potential respondents who were part of the sample but failed to participate (Groves & Couper, 1998). Lynn (2008, p. 37) summarizes the reasons for such unit nonresponse: 'Failure of the data collector to locate/identify the sample unit; failure to make contact with the sample unit; refusal of the sample unit to participate; inability of the sample unit to participate (e.g. ill health, absence, etc); inability of the data collector and sample unit to communicate (e.g. language barriers); accidental loss of the data/ questionnaire'. The nonresponders can be divided into two groups: 'noncontacts' and 'noncooperators' (Stoop, 2005, p. 50). 'Noncontacts' are those people who could not be contacted, whereas 'noncooperators' were contacted but did not cooperate (Stoop, 2005, p. 13). Whatever the reason: no data are obtained from these potential respondents (DeMaio, 1980; Hox & De Leeuw, 1994; Saßenroth, 2013).

The main cause of unit nonresponse is potential respondents failing to cooperate. This could be either because they are unable to, e.g. because of sickness, a language problem or because they refuse (Stoop, 2005, p. 50). Why do respondents refuse to participate? Many reasons are given by respondents, including lack of time and lack of interest (Stoop, 2005, p. 57). Groves, Cialdini and Couper (1992) suggest that a number of factors influence the respondent's decision (not) to cooperate in a survey, including expectations in society, question design and the role of the interviewer (if applicable). Furthermore, they find that respondent characteristics may also explain cooperation.

The unit response rate is an indicator of data quality (Fricker & Tourangeau, 2010; Wagner, 2010). The reasons that unit nonresponse is considered to be problematic are twofold. First, the absolute number of respondents in the survey is reduced which decreases the effective sample size. Secondly and more importantly, if the nonrespondents differ from the respondents, in terms of their opinions, nonresponse bias arises (Groves, 2006; Kohler, 2007). Nonresponse is almost never randomly distributed and is 'typically associated with at least some of the survey variables' (Lynn, 2008, p. 36). As a result, the effective sample may be unrepresentative which makes generalization to the population very problematic.

Over the years, the level of unit nonresponse in survey research has increased (Groves, 2006; Hox & De Leeuw, 1994; Steeh, 1981; Stoop, 2005). Although it depends partially on the data collection method, with face-to-face interviews

usually resulting in the lowest level of nonresponse, the general trend is that fewer respondents participate in surveys and/or fill them out completely (Hox & De Leeuw, 1994; Stoop, 2005). Whether this is a problematic trend depends on the consequences of nonresponse, which primarily depends on the (non)randomness of it. Unit nonresponse need not be problematic per se (see e.g. Groves, 2006), but an unrepresentative sample threatens data quality (Groves & Peytcheva, 2008; Stoop, 2005, pp. 24-25). And the lower the response rate, the more likely that the sample will be unrepresentative.

Since often there is no information available about the nonrespondents, it is difficult to establish whether and how they differ from the respondents and thus whether a good coverage of the intended population is provided by the sample. A small body of literature exists of follow-up research into the nonresponders, however (see Stoop, 2005). For example, according to Voogt and Van Kempen (2002) 'nonrespondents who refuse to cooperate with the survey tend to have had a somewhat lower education, are mostly older and tend to reside in urban areas. The nonrespondents who could not be reached, are relatively higher educated, younger, more often single and are overrepresented in urban areas'. The saliency of a topic, the organization conducting the research, the type of sample and (depending on the data collection method) the number of reminders or visits all affect a respondent's decision to participate, which in turn affects the level of nonresponse (Hox & De Leeuw, 1994; Shoemaker et al., 2002).

There is some general concern about the unit response rate of web surveys which is usually lower than in other survey modes (Bosnjak & Tuten, 2001; Messer, Edwards, & Dillman, 2012), although some studies show comparable results for web and mail surveys (Kaplowitz, Hadlock, & Levine, 2004). In general, meta-analyses show that web surveys result in lower unit response rates (Shih & Fan, 2008). Self-administered surveys have higher unit nonresponse rates than survey modes with an interviewer (Stoop, 2005, pp. 48-50). The number of break-offs is relatively high in web surveys (Lynn, 2008, p. 41), because the decision to cooperate and finish the survey is made when the questionnaire is opened and in survey modes with an interviewer (i.e. telephone and face-to-face) this decision is made during the introduction (Stoop, 2005, pp. 47-48).

Unit nonresponse is a problem that needs attention when doing surveys and 'should be a serious source of anxiety' (Stoop, 2005, p. 5). It is, however, not the focus of this study. This study focuses on the respondents who do participate or cooperate in surveys, but on occasion do not give a substantive answer to a survey question: item nonresponse.

2.4.2 Item Nonresponse

Item nonresponse means that 'data on particular items are missing' (De Leeuw et al., 2008, p. 17); the units or persons participating in the survey do not provide answers to particular items (see also de Leeuw, 2001, p. 147). Three types of item nonresponse can be distinguished: 1) missing by design, when respondents do not answer certain questions because of routing (Huisman & van Der Zouwen, 1998) and vignette experiments; 2) partial nonresponse, resulting from panel mortality or attrition and break-offs; and 3) other item nonresponse. The latter category consists of unusable or lost data or 'info [that] is not provided by a respondent' (De Leeuw et al., 2003, p. 158), the latter source of item nonresponse is central to this study. This type of nonresponse applies to cooperating respondents who cannot or do not want to answer specific individual questions (Mason, Lesser, & Traugott, 2002; Shoemaker et al., 2002). Substantive answers are missing for specific survey questions.

There are various reasons why a respondent gives a non-substantive answer to a survey question. The respondent is, for instance, embarrassed to reveal his or her true opinion and wants to avoid such embarrassment (Kreuter, Presser, & Tourangeau, 2008). Another reason is that the respondent does not have enough knowledge or information to answer the question (Krosnick et al., 2002; Shoemaker

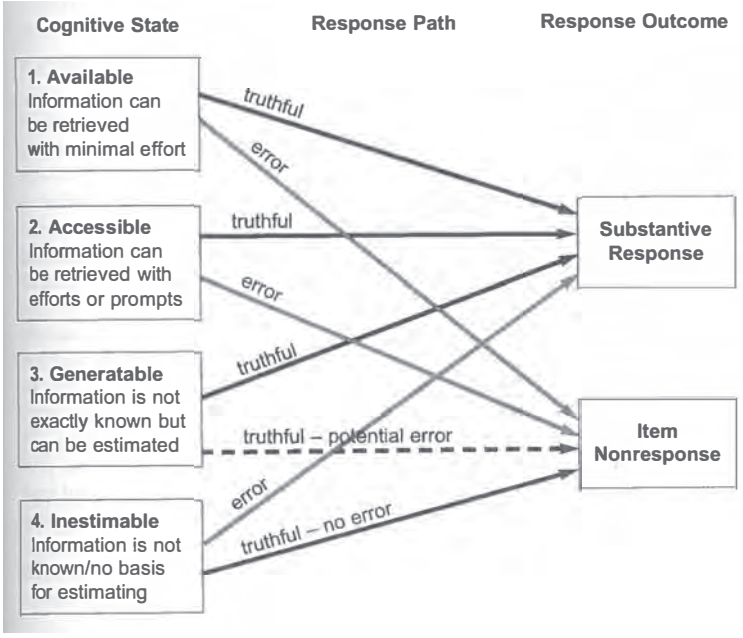

Figure 2.1: Beatty, Herrmann, Puskar & Kerwin's (1998, p. 410) Response Model for Item Nonresponse

et al., 2002; Tourangeau et al., 2000). Alternatively, the respondent could be unwilling to think about the issue asked about, because he or she is uninterested or does not care enough to make the effort. See Figure 2.1 for an overview by Beatty *et al* (1998) of the response process which may result in either a substantive answer or a (non-substantive) item nonresponse.

In this study, item nonresponse is treated as a given; the reasons for using a non-substantive response option are not examined and no distinction is made between various non-substantive response options like don't know or no opinion. Item nonresponse is a repository for all respondents who used a non-substantive response option. The focus is on the consequences of using such a response option on both item nonresponse and the substantive outcome. Whether item nonresponse is problematic, depends on the level of item nonresponse and a potential nonresponse bias. If more respondents answer a certain question it increases the opportunities for using statistical analyses. More importantly, item nonresponse may not be randomly distributed and result in a bias. This item nonresponse bias consists of the fact that if a certain group of respondents does not answer a question, the results may not be representative of the population. 'Nonresponse will not necessarily bias survey estimates, but it will do so if the nonresponders are systematically different from the responders' (De Leeuw et al., 2008, p. 17; Lynn, 2014, p. 319). It is therefore relevant to understand whether the data are missing at random or whether a bias occurs, which is discussed below in the Missing Data paragraph.

So item nonresponse could have two consequences: a smaller amount of data (which limits available statistical analyses) and invalid survey results because of a nonresponse bias. That is why item nonresponse is an important indicator of data quality at the level of individual questions. If the item nonresponse rate is high, the validity of the results and the ability to generalize the findings are threatened.

2.4.3 Item Nonresponse in Web Surveys

Item nonresponse in web surveys is the result of respondents using a non-substantive response option available to them. Depending on the design of the web survey, the non-substantive response option could be a response category (e.g. Don't Know or No Opinion), a filter question or the possibility to skip questions without selecting an answer. Using any of these options is registered as item nonresponse.

Considerably less attention is paid to item nonresponse in web surveys than to unit nonresponse; unit nonresponse is often described as 'nonresponse', without distinguishing it from other types of nonresponse (see e.g. Groves, 2006; Peytchev, 2013; Shih & Fan, 2008). Even though an unrepresentative sample is problematic if the aim is to generalize to the population, more focus should be on item nonresponse and to check data quality at the level of the individual item. Especially when

question design effects are examined, these differences between respondents and non-respondents at the level of individual items should be considered. The same logic applies as for nonresponse at the unit level: it is not necessarily the overall (non)response rate, but in particular the distribution or bias that is potentially problematic (Bethlehem & Biffignandi, 2011; Lynn, 2008).

In order to better understand item nonresponse in web surveys, this self-administration mode should be compared to other survey modes. Comparison across survey modes is difficult however, because typically other factors (like the design or layout of the web survey) also vary (Smyth, Dillman, Christian, & Stern, 2006; Stern et al., 2007). Heerwegh and Loosveldt (2008) collected more DK answers in their internet survey about immigrants than in the face-to-face survey on the same topic. Both Van Ewijk (2004) and Fricker *et al* (2005) compared a web and telephone survey; Van Ewijk had more item nonresponse for the web survey whereas Fricker *et al* found less nonresponse. This may suggest that 'much depends on the nature of the questions and the way the survey is implemented' (De Leeuw & Hox, 2011, p. 61). Generally speaking, however, web surveys result in more item nonresponse than other survey modes, when a non-substantive response option is offered (Messer et al., 2012).

The relatively high item nonresponse rates of web surveys 'suggest that web surveys may be at higher risk of nonresponse error' (Tourangeau, Conrad, & Couper, 2013, p. 6). More item nonresponse increases the potential for bias and is therefore an indicator of worse data quality. Furthermore, if item nonresponse is considered to be useless information or missing data, more item nonresponse is even more problematic: less (valid) data are collected when item nonresponse is excluded as missing data.

2.4.4 Missing Data

Missing data are often treated as an indicator of data quality, since they indicate a loss of information and a potential bias (De Leeuw et al., 2003, p. 153). This could be at either the unit (respondent) or item (survey question) level of data collection; here the focus is on item nonresponse or non-substantive answers as missing data. The loss of information means that potential data entries were lost because the respondent, for whatever reason, did not answer the survey question or the answer was not registered. If the respondent uses a non-substantive response option, this is registered as item nonresponse. This need not be missing data, however: 'An item is missing if the researcher interprets it as such (...) Item nonresponse is defined as the failure to obtain *information* for a question in an interview or questionnaire, so *data* are missing' (De Leeuw et al., 2003, p. 156). In other words: if the researcher thinks nonresponse is valuable information to address 'the big white elephant of

public ignorance' (Moore, 2008, p. 22), item nonresponse should not be considered missing data (see also Groves, 2004, p. 156). This study looks at item nonresponse both as valuable information and as missing data that is excluded to reveal what public opinion looks like.

If item nonresponse is not viewed as valuable data, there are two potential consequences of the occurrence of missing data: a) less data available for analyses because respondents have given fewer substantive answers; and b) less valid results, depending on whether more item nonresponse (or missing data) results in more 'nonresponse error' (Tourangeau et al., 2013, p. 6). To use the terminology of the missing data paradigm: are the data missing at random or not?

It could be the case that regardless of item nonresponse survey results of individual items are representative of the population. When data are 'missing completely at random (MCAR) (...) the missingness of a response to a question is unrelated to its unknown value and also unrelated to the values of responses to other questions' (De Leeuw et al., 2003, p. 155). Another possibility is that the data are missing at random, given covariates (MAR): 'When the missingness is related to the observed data but not to the (unknown) value of the missing response to the question itself, it is said that the data are *missing at random* (MAR)' (De Leeuw et al., 2003, p. 155). Missing data are unfortunate because of the potential for a higher response rate and consequently a larger sample size, but they do not substantially affect the outcome. The missing data may result in inefficiency, but not in bias. It is likely, however, that data are 'not missing at random (NMAR)' and that there is a relation between the fact that they are missing and the questions or some other variable of interest (De Leeuw et al., 2003, p. 155). For example, the respondent does not want to give socially undesirable answers and chooses to use the non-substantive response option, resulting in item nonresponse bias.

In some questionnaires, respondents are encouraged or even forced to give an answer to each separate survey item. In this format, which is called forced-choice, no options to say Don't Know or leave the question blank are offered and no filter question is used (Smyth, Dillman, Christian & Stern, 2006). The result is that no data are missing, but the effect of a forced-choice design on data quality is largely unknown. Other designs are less strict and push respondents to answer, but do register item nonresponse when the respondent is persistent.

Alternatively, there are question(naire) designs which offer the possibility not to give a substantive answer. There are various ways to do this, but the conventional way is to offer an explicit DK category and/or the use of a filter question (Schuman & Presser, 1996). The application of these design choices usually results in a higher level of missing data, but it remains a point of discussion whether valuable information is lost or whether nonattitudes are excluded (De Leeuw, Hox, & Scherpenzeel, 2010;

Gilljam & Granberg, 1993; Krosnick & Presser, 2010). Hippler and Schwarz (1989) for instance argued that filter questions may result in a loss of information, since they seem to suggest that a lot of knowledge or information is required to answer the survey questions. Respondents may use a non-substantive response option as an easy way out (e.g. Krosnick, 1999). Other scholars argue, however, that offering a non-substantive response option improves the validity of survey results, because at least some respondents are in that case willing and able to admit to having no opinion (Schuman & Presser, 1996).

Whether one prefers a certain question design is directly related to how one perceives and interprets missing data. If item nonresponse is seen as missing data because of a loss of potential information and the potential for nonresponse bias (De Leeuw et al., 2008, p. 17; Lynn, 2014, p. 319), more item nonresponse indicates worse data quality. From this point of view item nonresponse threatens the survey's validity and data quality; this is why item nonresponse is included in the Total Survey Error Approach (e.g. Biemer & Lyberg, 2003). If item nonresponse is, however, not seen as missing data but rather as valuable information and as a result of including non-substantive response options to discourage respondents from giving nonattitudes as answers to a survey question (Krosnick, 1991; Krosnick et al., 2002), a question design including explicit non-substantive response options will be preferred. Rather than registering nonattitudes and notice that does not reveal any preferences, a better way of measuring public opinion would be to provide individuals the option not to give a substantive answer if they are unable to do so.

The analysis of item nonresponse needs to include both perspectives on item nonresponse, by focusing on two aspects: 1) the level of item nonresponse, with a focus on differences between the substance of questions and the nature of the non-substantive response option offered; and 2) the distribution of opinions, to see if the missings are randomly distributed or affect the overall outcome of the survey. The first aspect treats item nonresponse as valuable information, while in the second aspect, the analysis of the distribution of opinions, item nonresponse is excluded as missing data from the picture of public opinion. In doing so, both the level (in general and for specific survey questions) and variation (between survey items) can be addressed, as well as the potential nonresponse bias resulting from data (not) missing at random.

2.5 Survey Results 2: Distribution of Opinions

This study explores the effect of question design, i.e. non-substantive response options, on two elements of survey outcomes: item nonresponse or non-substantive answers and the overall distribution of opinions. This distribution of opinions is the general picture of public opinion emerging from the survey. What does public opinion look like? Is a certain policy position supported by a plurality, majority or minority of the respondents? And does the public opinion look different when another question design is used? In other words: how does the use of various non-substantive response options affect the overall picture of public opinion, when item nonresponse is excluded as missing data? The distribution of opinions as a picture of public opinion is viewed in two ways: 1) including item nonresponse, to illustrate the level of 'public ignorance' and other aspects of non-attitudes; and 2) excluding item nonresponse.

The analysis of what public opinion looks like moves beyond merely methodological effects into the realm of responsiveness towards public opinion (by politicians). If the quality of the information as collected by surveys is affected by question design, and specifically non-substantive response options, the usefulness of surveys and whether politicians could and should take their results into account is also affected. This is what McClendon (1986) coined 'unanticipated effects' of offering non-substantive response options. The inclusion of non-substantive response options like a Don't Know option or filter question could not only affect item nonresponse (or missing data), but also 'substantive response distributions' (McClendon, 1986, p. 379). This study explores whether the impression of public opinion from surveys differs when a different question design is applied.

While the literature considers item nonresponse as an indicator of data quality, this subsequent step of looking at the actual survey results is often neglected. Bishop, Oldendick and Tuchfarber concluded over thirty years ago that using a filter question 'can in some instances dramatically affect the conclusions a pollster would draw about the distribution of public opinion on an issue' (Bishop et al., 1983, p. 528). Others found that discouraging nonattitudes by including a non-substantive response option did not improve the reliability, but the substantive responses changed and a different picture of public opinion emerged (McClendon, 1986; McClendon & Alwin, 1993). These findings suggest that a different picture of public opinion results from a different question design. The aim here is to find out whether public opinion, as constructed by surveys, is as robust as is often assumed or that it is at least partially created by the way the questions are asked.