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

Maat, J. van de

CHAPTER 1

Introduction

'The fact that large proportions of the citizenry have no opinion may create (...) different interactions between citizenry and government than when awareness and opinion preferences on an issue pervade the population' (Key, 1961, pp. 77-78)

1.1 Abstract

This is a study on item nonresponse and non-substantive answers. Item nonresponse means that a respondent did not answer particular survey questions (or 'items'). Substantive answers are missing for specific items. Studying these missing substantive answers is important, because of two reasons: First, a high item nonresponse rate threatens the validity of the results and the ability to generalize the findings, because a) less data are available, which limits available statistical analyses; and b) item nonresponse may not be randomly distributed, potentially resulting in bias and invalid findings. Secondly, non-substantive answers provide valuable information about which part of the public is unable or unwilling to answer individual survey items.

Most studies focus on the effect of non-substantive response options on item nonresponse (e.g. Bishop, 2005; Bradburn, Sudman & Wansink, 2004; Schuman & Presser, 1996). Alternatively, a body of literature exists about whether the missing data, i.e. item nonresponse or non-substantive answers, are missing at random (Tourangeau et al, 2013; De Leeuw et al., 2003). These scholars focus on the way the 'missingness' of answers is distributed. If they are not missing at random, a bias of the survey outcome may occur. What is missing by and large is a focus on the resulting picture of public opinion. What does public opinion look like when non-substantive answers are registered in a different way?

The research question in this study is: *how does question design regarding non-substantive response options affect survey outcomes?* Specifically, the study focuses on the use of non-substantive response options, i.e. the Don't Know option, the filter question and the follow-up question. Both non-substantive answers and the actual distribution of opinions, i.e. the substantive results, are examined as outcomes. The goal of this study is to see whether various ways to register non-substantive answers affect the results for specific substantive response alternatives.

The introductory chapter develops an argument about why it is important to study public opinion, and particularly why it is important to examine in more detail how question design affects the outcome, before arriving at a more extensive discussion of the puzzle and research question.

1.2 Surveys in the Public and Political Debate

Suppose a national newspaper reports: '66 percent of the Dutch want to introduce the death penalty!'. Such a report would likely receive attention from other media and politicians and become part of the public and political debate – especially in the

aftermath of a grave incident, when the public feels the death penalty is warranted. 66 percent would be a relatively large majority preference. But suppose next that another newspaper also asks citizens what they think about the death penalty, and 40 percent of those citizens does not answer the question because they do not want or cannot answer it. Most of the other respondents (40 percent of the total sample) may support the introduction of a death penalty, but the overall public's preference now seems much less clear.

This example illustrates two things: first of all, the role and potential impact of surveys in representative democracies by representing (some form of) the public's voice; and second, the importance of having information about citizens who do *not* report their opinions in response to a survey question. The role of polls and surveys at election time, but also in the broader public and political debate is evident in both the growing number of organizations doing survey research and their political and policy importance (Kohut, 2009; Lepore, 2015). This trend is discernible in Figure 1.1, where the number of times an opinion poll was mentioned in American newspapers is displayed. Whether it is the design of a new American banknote (Greenhouse, 2015), the replacement of judge Scalia during Obama's final year of his presidential term (Agiesta, 2016; Quealy, 2016), the Brexit in Great Britain (Crosby, 2016; Gripper, 2016; Kirk & Wilkinson, 2016), or Merkel's decision (in April 2016)

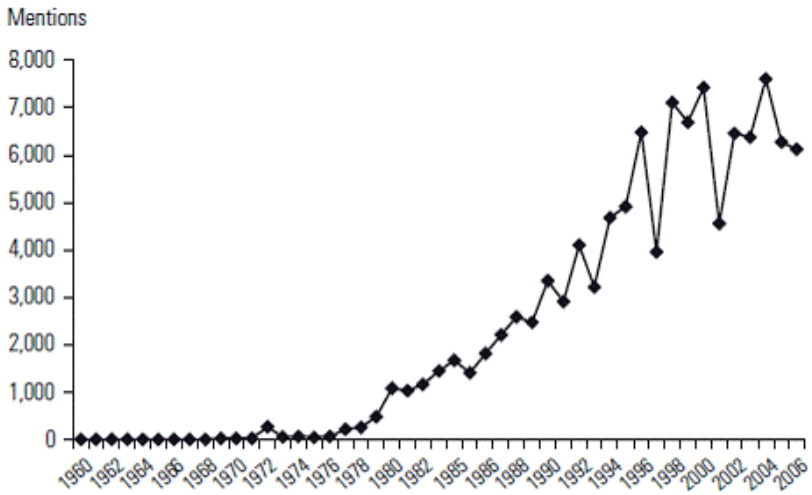

Figure 1.1: Opinion Poll mentions in U.S. news and wires

Source (Kohut, 2009): 'Data drawn from Lexis-Nexis search of newspaper and wire service reports using the term *opinion poll*. [Kohut] searched newspapers published in the United States and wire services where more than 60 percent of the stories originate in the United States for a total of 463 news sources'.

to allow Turkey to prosecute a German comedian who had insulted the Turkish president Erdogan in a poem (Hellemann, Hollstein, Peters, Pfeffer, & Niehus, 2016; Lindhout, 2016)[1]; the general public was asked what they thought of the issue and the outcome became part of the public and political debate.

The prominent place of surveys in the public and political debate is also apparent in the Netherlands. Recent Dutch examples of surveys becoming part of the political debate include the potential introduction of the kilometre tax (which tied road taxes to the actual use of the car)[2], a government commissioned research to gauge opinions about nuclear energy energy (Van Keken, 2010a, 2010b) and a survey by the most famous Dutch pollster, Maurice de Hond, in 2013 about whether Romanians and Bulgarians should be allowed to work without a permit in the Netherlands leading to the Socialist Party demanding action from the minister (Mikkers, 2013). Such examples clearly suggest that representations of public opinion are a force to be reckoned with, at least sometimes for some issues. Second and more importantly, the examples point toward a key role for surveys in monitoring what the public wants. What these examples do not show, however, is how the respondents' opinions are measured and what part of the public is unwilling or unable to answer particular survey questions. These questions are at the core of this dissertation, but before addressing them, the concept 'public opinion' and its role and impact in advanced western democracies will be discussed.

1.3 Defining Public Opinion and Surveys

In order to examine the problems of measuring public opinion with surveys, it should be clear how the concepts *public opinion* and *surveys* are used in this study. It is not easy to come to grips with the concept public opinion. According to Herbst (1998, p. 1): 'Public opinion assessment [is] one of the most frustrating and challenging aspects of democratic practice' and that 'defining public opinion is an exceedingly difficult and complex task' (Herbst 1998, p. 2). That public opinion is considered vague and hard to describe is also evident in the image of public opinion by Lippmann (1927) as a 'phantom' and by V.O. Key (1961) as the 'holy ghost'.

1 Technically it was not solely Merkel's decision, but the German laws which enable criminal prosecution.

2 The Dutch minister of Transport, Eurlings, tied the continuance of the kilometer tax explicitly to public opinion. More specifically, he wanted to know what stakeholders – in this case the car users – thought of the tax, which tied payment to the actual use of the car, and vowed to adhere to their opinions (De Graaf, 2010). In the end, the introduction of the tax was postponed.

Moreover, the meaning of the concept public opinion has changed over time. Where it was once used in a way that it 'transcends individual opinion and reflects an abstract, common good' (Price, 1992, p. 11), which means that the public as a whole has more wisdom than the separate individuals, it has transformed towards majority rule: 'Public opinion as the result of a confidential, scientifically conducted survey of unconnected individuals' (Herbst, 1993b, p. 172). This aggregation of individual opinions has become 'the most common conception of public opinion today' (Price, 1992, p. 22). Also, public opinion in the 18th and 19th century was equated with the opinion of the elites, but over time it came to refer more to mass opinion (van Ginneken, 1999, pp. 25-26).

No consensus exists about what public opinion exactly entails; a general notion prevails that public opinion should be considered as a social construct (Herbst, 1998, pp. 150-151). According to Herbst (1998, p. 14), public opinion originates from four sources: '1) the model of democracy shared by members of a community or nation; 2) the types of technologies or methodologies available for opinion assessment; 3) the rhetoric of our leaders; and 4) the evaluation of public opinion by journalists'. The second element should be considered crucial in examining what surveys measure, since surveys have become so prominent in assessing what the public wants. The rise of the survey as a technique to gauge public opinion is discussed below, but it should already be noted here that the choice for an aggregative definition of public opinion as measured by surveys is consequential: 'Public opinion expression, in this case, is categorical in nature (individuals may choose among two or more options), unattributed, statistically representative of the populace, and directed by the survey researcher and his or her choice of survey form' (Herbst, 1998, p. 16).

While this aggregative definition of public opinion is dominant, other definitions are also used, depending on the type of research that is conducted, the historical conditions and the technology available (Glynn, Herbst, O'Keefe, Shapiro, & Lindeman, 2004, pp. 31-32). Herbst (1993a, pp. 439-440) for example, distinguishes between four categories regarding the meaning of public opinion. The first definition of public opinion is aggregation: the opinions of individuals are aggregated with polls, surveys, elections and referenda. Secondly there is the majority opinion, which is congruent with a democratic principle that the majority rules. Third is the discursive/consensual definition, based on the idea that public opinion is formed through discourse among members of the public. The final category holds that public opinion is a 'reification or fictional entity' and thus does not exist at all (Herbst, 1993a, p. 440). Glynn et al (2004, pp. 19-25) concur with Herbst's categorization, but they add one more category of public opinion: the opinion of the media and elite.

Table 1.1: Techniques for the Expression and Assessment of Public Opinion

Techniques	Time of Appearance
Oratory/Rhetoric	5th century B.C.
Printing	16th century
Crowds	17th century
Petitions	Late 17th century
Salons	Late 17th century
Coffeehouses	18th century
Revolutionary Movements	Late 18th century
Strikes	19th century
General Elections	19th century
Straw Polls	1820s
Modern Newspapers	Mid-19th century
Letters to Public Officials & Editors	Mid-19th century
Mass Media Programming (Political)	1920s –1930s
Sample Survey	1930s

Source: (Herbst, 1993b, p. 48)

The adopted or preferred definition of public opinion is contingent on several factors: time, since the definition of public opinion has changed over the course of history, the object of the study and the way one wants to measure (Herbst, 1998). This also shows from the historical development of the techniques used to assess public opinion (see Table 1.1).

While (old) techniques like petitions and strikes are still used to express and assess public opinion, the sample survey adheres most to the dominant aggregative definition of public opinion. The reasons for the contemporary dominance of the aggregative definition of public opinion are arguably the apparent straightforward way to measure it, the resemblance with the electoral democratic system and its principle of one-man-one-vote and the possibility to analyze causal relationships which may affect the public's opinion (Glynn et al., 2004, p. 20). Price (1992, p. 72) states that the advancements in the collection and analysis of data in large populations contributed to the rise of aggregation as the dominant conception. Furthermore, he notes a shift towards 'the individual side' which 'starts with a representative sample of individual opinions "in all its narrowness and firmness"' (Price, 1992, p. 72), meaning that the focus is nowadays more on individuals and individual opinions rather than a focus on public opinion as an outcome of a societal or political process or public discourse. In other words: operationalizing public opinion as the outcome of surveys is directly related to the dominance of the aggregative

definition, with public opinion as the sum of individual opinions (Herbst, 1993b, p. 43).

The aggregative definition of public opinion, which is also adopted in this study, goes hand-in-hand with the use of polls or surveys. The increasing use of the survey technique has led to the 'one person, one vote" tally of opinions (...) as (...) a baseline definition of public opinion' (P. E. Converse, 1987, pp. S12-13). This is part of the rise of quantification in general and in survey research in particular. Quantification is attractive, because of its 'objective and seemingly decisive nature' (Herbst, 1993b, p. 2). Partly for this reason, survey outcomes have become valuable information in the political debate and decision-making process.

What is a *survey*? 'In a good survey, the sample that has been studied represents the target population, and the information that has been collected represents the concepts of interest. The standardised procedures with which data are collected are mostly, but not always, questionnaires which are either presented to the sample persons by an interviewer or completed by the sample persons themselves' (Stoop & Harrison, 2012, p. 8). Before conducting a survey, according to Stoop and Harrison (2012, pp. 8-16), decisions must be made about the (target) population, the sampling procedure to find members of the target population, the topic of the survey, the survey agency executing the survey, the survey mode and the timing of the survey. In other words: who is studied, how, about what and when? The goal is to 'obtain a composite profile of the population' – not the individuals in the sample (Scheuren, 2004, p. 10).

The notions *polls* and *surveys* are often mixed up and used interchangeably. Both polls and surveys gather individual opinions with a questionnaire. While polls are (at least in the American literature) often described as election forecasts by asking the respondents which party or candidate they are intending to vote for (Brettschneider, 1997; Levy, 1983), some scholars also use this notion for more general measurements of opinions (Blumer, 1948; P. E. Converse, 1987; Erikson & Tedin, 2015). Herbst (1998, p. 48) calls the latter type of polls 'issue polls'; other authors coin these measurements of individuals' opinions (opinion) surveys (e.g. Traugott & Lavrakas, 2007, pp. 1-2). As said, polls are sometimes differentiated from surveys by their use to predict election outcomes, but in other cases a distinction is made between polls as questionnaires which are shorter, with a smaller sample and typically commissioned by commercial organizations and longer, more scientifically conducted surveys (Traugott & Lavrakas, 2007, pp. 2-3).

This study will use both notions of *polls* and *surveys* interchangeably, for two reasons: 1) it is difficult to make a clear distinction between the two notions as suggested by the disagreement in the literature; and 2) the exact delineation of the

notions does not have much added value to this study. This study is about question design effects, or more specifically the effect of non-substantive response options in *any* poll or survey collecting individual opinions via questionnaires.

1.4 Public Opinion and Surveys in a Democracy

With a clearer picture of public opinion and its measurement with polls or surveys in mind, the next and arguably key question is: why bother? The answer is directly related to the potential role of surveys in a democracy. 'Unless mass views have some place in the shaping of policy, all the talk about democracy is nonsense' (Key, 1961, p. 7). Democracy nowadays almost equals elections, but that used to be different. Manin (1997) argues that what we call representative *democracy* would not qualify as a democracy for people like Rousseau, Madison and Siéyès, who would rather speak about a republic, since democracy was equated with what we now would likely call direct democracy. 'The modern meaning and the eighteenth century meaning (...) share the notions of political equality among citizens and the power of the people' (Manin, 1997, p. 4). What has changed, according to Manin (1997, p. 4), is how this notion is transformed and translated into 'principles of representative government', including elections and independent decision-making by elected MPs[3].

In representative democracies, the responsibility for making policies and governing the country is delegated to a very small number of individuals who are elected or, if appointed, at least derive this authority from elections. The formal structure of electing representatives has barely changed in recent history (Przeworski, Stokes, & Manin, 1999a, p. 3). Elections have at least two functions: to provide a mandate 'to select good policies or policy-bearing politicians', and to ensure accountability 'to hold governments responsible for the results of their past actions' (Przeworski, Stokes, & Manin, 1999b, p. 29). The latter function is consistent with Fiorina's retrospective voting theory in which representation is a

3 The argumentation in this section focuses on a 'vertical' relation between citizens and MPs. Another way to think about the role of surveys is by looking at their impact on the formation and discussion of public opinion, i.e. a 'horizontal' relation of public opinion on the mass public. The public can form opinions based on the outcomes of surveys about subjects they do not have (extensive) personal experience with (e.g. Koopmans & Erbe, 2004). Consequently, citizens may also use survey outcomes to find out whether their opinions are prevailing in society; if this is not the case a citizen may be less inclined to express the opinion. A 'spiral of silence' may ensue where citizens do not express the opinions less approved (or less heard) by society at large (see e.g. Noelle-Neumann, 1974; Scheufle & Moy, 2000). The horizontal line of argumentation entails that (the formation of) public opinion is affected by survey outcomes, which reinforces the need to examine what we are measuring in surveys.

mechanism of the people to control their representatives (Fiorina, 1978, 1981; Key, 1966; Miller & Stokes, 1963).

According to Stimson *et al* (1995, p. 557), elections are just one mechanism for public opinion to directly influence public policy. Another more indirect mechanism resulting from elections is rational anticipation by policymakers who adjust their proposals if that leads to positive future results, e.g. reelection. In order to improve their chances of being reelected, policymakers anticipate how policy proposals are judged by their voters and subsequently may adapt their position on crucial issues. In this way both public opinion and rational anticipation are mechanisms for representatives to respond to their people's wishes. Anticipating for elections by responding to changes in public opinion, also called 'responsiveness', fits the dynamic model of representation (see Arnold & Franklin, 2012); 'congruence' or 'concurrence' refers a more static process where changes in policymaking reflect electoral turnover (Miller & Stokes, 1963; Verba & Nie, 1987). The level of responsiveness actually deployed varies (see for example Erikson, Mackuen, & Stimson, 2002; Jacobs & Shapiro, 2002), but 'listening to the public' is at the heart of any representative democracy.

Polls could be seen as 'broadly representative' of what the public wants by measuring public opinion (Gallup & Rae, 1940/1968), but their preferred role is contingent on the model of democratic representation that is adopted. Eulau (Eulau, 1962; Eulau, Wahlke, Buchanan, & Ferguson, 1959) distinguishes three representational role conceptions: the delegate, politico and trustee model. These conceptions differ in the level of discretion the politician or representative has in deciding what policy he will pursue. The trustee pursues what he deems right, 'his convictions and principles, the dictates of his conscience'; delegates agree 'that they should not use their independent judgment or convictions as criteria of decision-making' (Eulau, 1962, pp. 749-750). Politicians might employ different levels of responsiveness based on the issue and/or their own disposition.

There is discussion about the freedom of the MP or representative to act independently or as part of a collective (Thomassen & Andeweg, 2004), i.e. to what extent an MP is able to individually represent other interests than as a party member. Although party unity is very high in the Netherlands (see e.g. van Vonno, 2016) s/he may still be able to represent other interests as well and switch to a different role (Andeweg, 2012; Thomassen & Andeweg, 2004; van Vonno, 2012). Furthermore, Andeweg and Thomassen (2005, p. 508) argue that 'empirically, all representatives can be classified as politicos in Eulau and Wahlke's typology' and they propose a new typology consisting of two dimensions: 'direction' and 'control mechanism'. Finally, there is discussion among scholars as to whether (parliamentary) role theory contributes at all to explaining legislative behavior (Andeweg, 2014). The main point

here is that some form of representation is present in all the MP's roles.

Surveys have become an increasingly used means for assessing public opinion between elections. This is easy to understand, since elections only provide a general sense of direction rather than specific preferences about particular issues. Furthermore, elections only take place once every couple of years. In order to be able to responsive or to anticipate rationally, elected representatives and policymakers have to know what the people's will is and public opinion can increase their responsiveness which is 'central to democratic theory and practice' (Manza, Cook, & Page, 2002, p. 3). Generally speaking, 'government acts upon public opinion and public opinion acts openly and continually upon government' (Laswell, 1941, p. 15). And indeed, most people expect politicians to pursue the wishes of the public[4]: 'In a democracy... policy is supposed to flow from the preferences of the public' (Erikson et al., 2002, p. 33). Polls are the tool used most to assess such preferences. Polls can be valuable, because they form a practical means to gather information, they make comparison possible to what others think and they deal with issues which voters care about (Shiraev & Sobel, 2006, pp. 8-11).

It should be noted here that even though surveys have become an important measurement instrument of public opinion, it is not generally agreed upon that 'public opinion' can be measured by surveys, or at all. Scholars like Dewey (1954) are concerned about the public's ability to reason and participate in the democratic process. Others hold that '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). And this critical view is not restricted to the American context. Bourdieu (1973) argues that *'l'opinion publique n'existe pas'*, because of the underlying assumptions about individual opinions being available and holding equal weight, when measuring public opinion with a survey. Champagne (2004, p. 73) maintains that polls have become 'an instrument for the rational manipulation of election campaigns' which undermine the debate and reflection needed among citizens in a democracy (see also Champagne, 1990).

So there are concerns about the measurement of public opinion using surveys (see i.e. Bishop, 2015; Tiemeijer, 2008). Interesting as this debate may be (Ginsberg, 1986; Tiemeijer, 2006), the assumption in this study is that since public opinion as measured by surveys is in fact part of the public and political debate, it is worth investigating what it is that we measure. In this dissertation the assumption is that public opinion consists of what is measured with polls or surveys, which is consistent

4 Who makes up the public is the topic of another discussion. It could be the voters for a party, the party members, the voters for a specific politician, the majority and so on. For a more elaborate discussion, see Blumer (1946) and Price (1992).

with the dominant aggregative definition and interpretation of public opinion. And since it is assumed that public opinion is the sum of individual opinions collected with surveys and questionnaires, we should learn more about what it is that we are measuring.

1.5 Rise and Dominance of Surveys

According to Gallup and Rae opinion polls could "provide a continuous chart of the opinions of the man in the street" (Gallup & Rae, 1940/1968, p. v). Over half a century later the editor in chief of the Gallup Poll wrote his own plea as to why political leaders 'must listen to the wisdom of the people' (Newport, 2004). Following their line of reasoning the wisdom of the people - i.e. public opinion - manifests itself through mass opinion polls that use surveys and questionnaires to collect data on opinions and attitudes. Subsequently these 'numbered voices' should impact on public debate and democratic politics (cf. Herbst, 1993b). And even though "the United States is more poll crazy than other countries, politicians in other nations have much access to polling results when making decisions (...)" (Geer, 1996, p. 82). Dutch politicians do not form an exception to this general rule (see e.g. Dixhoorn, 2006; Koop & Van Holsteyn, 2008; Tiemeijer, 2006, 2008).

Polls have become increasingly important in media and politics, which was recognized as early as in 1936: 'Not only are the polls assuming increasing importance on the American political and social scene; they are also demanding more and more attention from the social scientist' (Katz & Cantril, 1937, p. 155). Since then, the number of polls executed and reported has only increased. According to Herbst (1993b) the rise of survey research is associated with the quantification of (American) politics. 'Quantitative techniques for expressing and measuring public opinion are attractive because of their "objective" and seemingly decisive nature, as well as their ability to account for a multitude of individual opinions. Political leaders, pollsters, journalists, interest groups, and members of the public have been increasingly drawn to these methods of estimating public opinion because numerical data tend to communicate authority: The data provide, in theory, an undistorted portrait of the common man's convictions' (Herbst, 1993b, p. 2).

While surveys may have become prominent in the contemporary public and political debate, they have been around for a much longer time. Bethlehem (2013, pp. 4-5) refers to censuses as one of the oldest means to gather statistics, which occurred even thousands of years ago. Jean Converse identifies three ancestors of modern day surveys: the English social survey, 'early psychological studies of attitudes, and marketing research' (Converse in Herbst, 1993b, p. 11). Techniques

to aggregate and express opinions appeared from the late 18th century, like general elections which '[required] private communication of opinion' (Herbst, 1993b, p. 57). Where before interaction was deemed necessary and 'public opinion was thought to be a consensus of individuals', aggregation of individual opinions became more and more popular (Herbst, 1993b, p. 59).

Collecting and counting opinions expanded with the early straw polls in the 1820s and 'taking note of attendance at political rallies' grew during the 19th century (Herbst, 1993b, p. 11). Straw polls are polls held by newspapers which tried to gather large numbers of ballots in order to forecast election results, without random sampling (Lusinchi, 2015; Robinson, 1937). The term, according to Bethlehem (2013, p. 6), refers to the straws that were cast into the wind by farmers to assess the direction of the wind; the straw polls were used to assess 'how the political wind blew' (Bethlehem, 2013, p. 6). The first straw poll was held in the US in 1824 during the 'first contested presidential election that would be largely decided by popular vote' (Smith, 1990, p. 23). These polls included counting at meetings or soundings at other elections. Smith (1990, p. 30) argues that the straw polls are an example of a bottom-up approach where the public wanted to know what the popular sentiment was about the presidential candidates. Contrary to Smith, Beniger (1983, p. 482) argues that 'survey research does not arise from a need to speak one's mind (...) but rather from the need to find out what is on people's minds – whether they intend to speak them or not'. According to this latter line of thought, surveys result from a top-down approach where businesses and governments wanted to know what the public wanted. Either way, the straw polls gained popularity since they gave an indication of which candidate or policy was favored by the public at large.

The 1930s and 1940s were crucial in terms of the development and subsequent acceptance of polls by congressman and journalists. After the 1948 polling debacle, in which the Truman victory was not correctly predicted, the pollsters continued to work on improving election predictions. Probability sampling was introduced and election polls were held until the final moments before the elections, to take last minute shifts into account (Bogart, 1972, p. 26). The more systematic approach to public opinion research intensified when Gallup started using random sampling (Bethlehem, 2013, pp. 5-6). This approach spread to other Western developed countries, like Great Britain and France in the late 1930s (Heath, Fisher, & Smith, 2005; Worcester, 1987). Other sources of the increasing importance of polls were the surveys sponsored by governments and academic surveys resulting in for example *The American Voter* (Campbell, Converse, Miller, & Stokes, 1960). More and more countries started to do pre-election surveys as well (Heath et al., 2005, p. 311).

The rise of polling is closely tied to the rise of electoral research; the focus in polling was at first on predicting election outcomes and explaining them afterward.

An example is the founding of the American National Election Studies in 1948 (ANES, 2014). Other major players in American public opinion research were founded around the same time, like the National Opinion Research Center in 1941 and the Roper Center for Public Opinion Research in 1947 (NORC, 2014; Roper Center for Public Opinion Research, 2014). In Europe, the founding of election studies followed with the Germans and Swedes in 1953/54 and the British Election Studies in 1964 (British Election Studies, 2014). This development 'was a deliberate effort by the Michigan group' who wanted to compare between countries, but also came about by the enthusiasm of European scholars to learn from American pollsters (Thomassen, 1994, p. 239) Two European comparative surveys started in the 1970s: the Eurobarometer, set up by the European Commission in 1973 to '[help] the preparation of texts, decision-making and the evaluation of its work', and the European Values Study, which was initiated in the late 1970s by Jan Kerkhofs and Ruud de Moor prior to the first direct elections for the European Parliament in 1979 (Eurobarometer, 2014; EVS, 2014).

Studies in the Netherlands on voting behavior and political participation and attitudes started after the Second World War, but remained limited in scale and ambition until the end of the 1960s (Van der Eijk & Niemöller, 1994, pp. 323-324). After the tumultuous elections of 1966 and 1967, the interest of politicians, journalists and the public at large grew and more and more ambitious voting studies were organized (Van der Eijk & Niemöller, 1994, pp. 325-327). The Dutch Parliamentary Election Studies, the NKO, started in 1971 because of 'the large influence the American election studies administered on the development of political science' (NKO, 2014, own translation JvdM). Election studies developed and became institutionalized. Note that commercial survey organizations were founded already during the 1930s and 1940s (Bethlehem, 2013; IPSOS, 2014; TNS-NIPO, 2014).

While voting research is a major form of public opinion research, it is not the only form of survey research that has grown tremendously. This growth in survey research is evident for example in the number of references to polls in the news, as illustrated in Figure 1.1 (Kohut, 2009), but also in the increase of government-sponsored surveys in the US between 1984 and 2004 (Presser & McCulloch, 2011). The rise of survey research is a result of societal developments like more interest in mass opinion and technological improvements like the sample survey (van Ginneken, 1999, pp. 26-27). Furthermore, marketing research stimulated the rise of surveys to examine the public's wants and needs, both in the US and in other countries, including the Netherlands (J. M. Converse, 1987; van Ginneken, 1993).

Opinion research in the Netherlands amplified after the Second World War (van Ginneken, 1993, pp. 54-56). Marketing and budget research were already executed in the 1930 and Statistics Netherlands (CBS) started gathering statistical data already

in 1899 (Kuijlaars, 1999; van Ginneken, 1993; van Maarseveen & Schreijnders, 1999), but the first major opinion research *Vrije meeningen in een vrij land* was done by the NSS in 1946 (Sweers & Lous, 1946). As said, more (opinion) research agencies were established during the 1940s and 1950s and opinion research became a booming industry (van Ginneken, 1993).

Interest in public opinion was not new, but its prominence in the public sphere was made possible by the introduction and success of surveys as a means to gauge public opinion (van Ginneken, 1993, pp. 186-188). Its rise and dominance only reinforces the need to explore potential problems in survey research and their consequences for the assessment of public opinion.

1.6 Problems in Survey Research

Jill Lepore examined the role of polls during presidential elections in the US and noted in the New York Times in November 2015: 'Lately, the Sea of Polls is deeper than ever before and darker' (Lepore, 2015). The discussion in the New York Times addressed polls and surveys about more general preferences, opinions and beliefs of citizens. The contributors varied in their evaluations which is apparent from the titles: 'Creating an Illusion of Public Opinion' (Bishop, 2015), 'Polls Can Give People a Stronger Voice' (Lupia, 2015) and 'Politicians Use Polls to Adjust Their Message' (Heith, 2015). The discussion about what polls measure and how this information should be used is of course not limited to the US. The same concerns can be heard in countries like the UK (see e.g. "The Guardian View on Opinion Polling: Quality Before Quantity," 2015; Silvera, 2015 for a take on the 2015 UK election polls) and the Netherlands (see e.g. Kanne, 2016; van der Meer, 2016; Vermeulen, 2015).

While journalists and politicians worry about the quality of surveys as a means to express the public's wants and needs, survey methodologists look at a more detailed level at what it is that surveys measure and the potential problems associated with that measurement process and technique. The public debate about surveys often centres around the general usefulness of surveys and whether one should listen to the people; the scientific debate is more concerned with specific elements of the data collection process, like sampling and response rates. These perspectives, however, do interact: a substantial part of the debate on the usefulness and value of opinion polls from a democratic perspective focuses on the quality (validity and reliability) of the information collected via questionnaires.

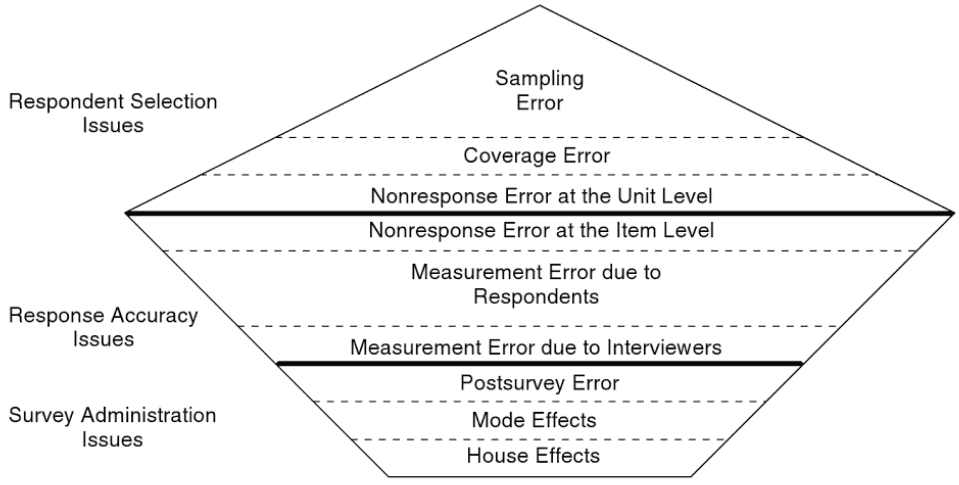

Figure 1.2: Types of Survey Error
Source: (Weisberg, 2005, p. 19)

What are these problems in survey research that affect the information about public opinion? De Leeuw *et al* (2008, pp. 6-13) specify in their 'International Handbook of Survey Methodology' four sources of data collection error: coverage, sampling, nonresponse and measurement errors; see also Weisberg's overview in Figure 1.2. *Coverage* refers to the (mis)match between the sample frame and the target population. *Sampling* refers to the way respondents were selected for the sample, with random selection (or probability sampling) as the most often preferred method to select a sample. *Nonresponse* most often refers to unit nonresponse, i.e. selected respondents do not participate in the survey. When non-respondents differ from respondents and this nonresponse is selective or biased in this way, a nonresponse bias occurs. These three types of error – coverage, sampling and unit nonresponse – are related to the fact that a sample of the population is targeted and all are 'error[s] associated with who answers' (Fowler, 2014, p. 9).

The fourth source of error (besides the 'survey administration issues' in Figure 1.2), *measurement*, refers to problems with the data collection process, or 'error associated with answers' (Fowler, 2014, p. 11) and includes four potential sources: the design of the questionnaire; respondents and their provided information; survey mode or the way the data are collected; and (if applicable, depending on the survey mode) the role of the interviewer. This is categorized in Figure 1.2 as 'response accuracy issues'. These potential data collection errors in surveys overlap

with the methodological problems in web surveys as described by Bethlehem (2013, pp. 9-16). He sees undercoverage of the population, self-selection of respondents rather than random sampling, unit nonresponse and measurement errors as aspects that could lead to unreliable and wrong conclusions. These potential problems are also visible in non-web surveys, but manifest themselves stronger or in a different manner in web surveys. Undercoverage is a bigger threat, for instance, because not everyone has internet access; and self-selection rather than random sampling is quite common in the composition of internet panels (Bethlehem, 2013, pp. 9-16; Bethlehem & Biffignandi, 2011).

All these survey problems or errors are part of the Total Survey Error paradigm (Biemer, 2010a, 2011; Weisberg, 2005). The Total Survey Error paradigm helps identify potential survey error sources to maximize data quality by statistically estimating the impact of the various survey errors on the survey outcome (Biemer, 2010b; Smith, 2011). While all these problems are important to address when conducting surveys, question design arguably has the most impact on data quality because it concerns a fundamental basis of survey research: conceptualization (Fowler & Cosenza, 2008a; Fowler & Mangione, 1990; Sudman & Bradburn, 1974). The focus in this study is on one particular aspect of this problem, i.e. the effect of non-substantive response options as part of the question design: item nonresponse.

1.7 Research Question

In this study, item nonresponse or 'item missing data' (see Tourangeau, Conrad & Couper, 2013, p. 53; Groves et al., 2009, p. 45) means that the respondent did not provide substantive information in response to a particular individual survey question. 'Data on particular items are missing' (De Leeuw et al., 2008, p. 17) or to be more precise: *substantive* answers are missing for specific survey questions. Respondents may have used a non-substantive response option, such as 'don't know', 'unsure' or 'no opinion', or they might have skipped the question. The use of these non-substantive response options is usually called item nonresponse.

Respondents may use a non-substantive response option for various reasons; because they cannot or do not want to answer a survey question or to lower the cognitive burden (e.g. Schuman & Presser, 1996; Krosnick & Presser, 2010). There is disagreement in the literature whether a non-substantive response option should be offered: 'Some argue that [response options like] "don't know,", "no opinion," and "undecided" provide those who cannot put themselves into one of the offered categories a way to register an honest response (Converse & Presser, 1986). Without a non-substantive response option, these respondents would have to select an untrue

answer or skip the question, neither of which is a desirable outcome. Others argue that providing these response options makes it easier for respondents to satisfice; that is, that respondents will select the non-substantive response option rather than doing the mental work necessary to report their true response (Krosnick, 2002)' (Dillman, Smyth & Christian, 2014, pp. 135-136). Furthermore, not offering a non-substantive response option in a web survey may result in more break-offs (Tourangeau, Conrad & Couper, 2013, p. 54).

That the design of separate questions affects the responses and subsequently the outcome of a survey is an established fact (e.g. Bradburn, Sudman, & Wansink, 2004; Schuman & Presser, 1996). More specifically, there is evidence that offering a don't know option explicitly, as a response option or filter question, results in more non-substantive answers, i.e. more item nonresponse (e.g. Bishop, 2005). It remains, however, an open empirical question how these non-substantive answers affect the actual distribution of opinions or survey outcome. Why respondents give a non-substantive answer or which respondents are more prone to use a non-substantive response option is not part of this study. The use of the non-substantive response option is treated as a given. Furthermore, the aim is not to discuss whether a non-substantive response option *should* be offered, but to investigate the impact of various ways to register non-substantive answers on the results for the specific substantive response alternatives.

In this dissertation I look at two specific aspects of the picture of public opinion: 1) non-substantive answers, i.e. item nonresponse and permissive opinions and 2) their impact on the substantive results or actual distribution of opinions. The distribution of opinions reflects the public's stance on a given issue, i.e. the plurality or majority supporting a particular policy position. The goal of this study is to see whether the number of non-substantive answers and the public's stance changes when a different question design is applied. Specifically, the effect of various question design choices on item nonresponse and the absence or presence of substantive opinions in public opinion surveys is examined. The general research question is:

How does question design regarding non-substantive response options affect survey outcomes?

Three question design elements are applied to identify non-substantive answers: the Don't Know option, the filter question and the follow-up question. The Don't Know option is offered as either an explicit response option or as an implicit possibility to skip a question. The filter question is a question posed before the substantive opinion question to give respondents the option to provide a non-substantive answer. In this study, two variants are tested: 'Have you already heard or read enough about [it] to have an opinion' and 'Do you have an opinion on this or not'. The third question

design element is the follow-up question. When a respondent gives a substantive response to an opinion question, the follow-up question asks: how upset would you be if the previously expressed opinion did not prevail when the issue was ultimately decided? If the respondent answers with 'not upset', the answer to the substantive opinion question is categorized as ' permissive', which means that the respondent permits the decision-making to go either way. Strictly speaking this is not item nonresponse, but for the public and political debate information on the strength of opinions, or whether the respondents really care about what happens, it is important background information.

No differentiation is made between types of non-substantive answers. The use of a non-substantive response option, either as a DK answer or by saying 'no' to a filter question, is a given. Refusals, 'no opinion' and DK answers may result from different mechanisms. The aim here, however, is to identify non-substantive answers. The non-substantive response options are offered as generic categories to enable a respondent to give a non-substantive answer. The aim of this study is to look at levels of item nonresponse, regardless of the respondents' reasons for using non-substantive response options.

The empirical part of this study consists of a series of three survey experiments exploring the effects of question design on survey outcomes, in particular in internet or web surveys. The general aim of the project is

1) to investigate the impact of various ways to register non-substantive answers on the general picture that emerges in terms of majorities or pluralities within public opinion, both including and excluding non-substantive answers; and
2) to investigate the effects for substantively different issues, that are assumed to be easier of more difficult for various respondents.

The project has been conducted in the Netherlands, where many pollsters are active, and polls and surveys are part of the public and political debate. In the US an abundancy or research is available about question design effects and the resulting picture of public opinion, but such research is largely lacking for the Netherlands – at least concerning item nonresponse. If the findings in the Netherlands are in line with the literature from the US, the broader applicability of question design effects and the resulting public opinion can be argued.

This study does not contain a normative, philosophical argumentation about what public opinion essentially is or should be[5]. Rather, it is an empirical study of the

5 See for example Tiemeijer (2006) and Yankelovich (1991) for such normative accounts of public opinion and polls in a democracy.

effects of question design. The research questions will be answered by conducting three different survey experiments, which are 'a deliberate manipulation of the form or placement of items in a survey instrument, for purposes of inferring how public opinion works in the real world (...). Comparing the decisions, judgments or behaviors of the respondents in the treatment group to those in the control group reveals the causal effects under investigation' (Gaines, Kuklinski, & Quirk, 2007, pp. 3-4). The experimental conditions in the study comprise alternative ways of offering respondents the possibility to decline giving a substantive response, or give a permissive response. In this way the impact of non-substantive opinions on survey outcomes can be assessed. By employing this design, this study hopes to contribute towards a deeper understanding of the impact of survey question design on item nonresponse and other non-substantive answers.

1.8 Outline of the Book

The aim of this study is to explore and analyze how the way questions are asked and response alternative are offered in a survey affect the outcome in general and level of non-substantive answers in particular. The focus is on elements of question design: the Don't Know option, the use of filter questions, and the use of a follow-up question. These three aspects are central in three separate internet survey experiments which form the empirical part of this dissertation. The underlying question is: what do we measure as public opinion when this question design element is (not) applied?

This introductory chapter is followed by a chapter which gives an overview of some relevant studies in the field of (internet) survey research methodology. In chapter 3 hypotheses are developed on the basis of this literature. Chapter 4 describes the design of the experiments: data and methods. Subsequently, each chapter contains the results of one internet survey experiment. These separate studies in chapter 5 to 7 show to what extent the outcome regarding certain issues is affected by the wording of a question and response alternatives. More specifically, the focus is on how the offered non-substantive response options affect the survey outcome. The focus in chapter 5 is on the Don't Know option; chapter 6 focuses on filter questions; in chapter 7 the 'so what' follow-up question is examined. In the 8th chapter the results from three survey experiments are brought together and compared. Finally, in the concluding chapter 9 the general conclusions and implications are discussed and some suggestions for future research are made.