Cover Page

The following handle holds various files of this Leiden University dissertation:
http://hdl.handle.net/1887/76853

Author: Maat, J. van de
Title: Public opinion without opinions? Item nonresponse and (the absence of)
substantive opinions in public opinion surveys
Issue Date: 2019-09-17

PUBLIC OPINION WITHOUT OPINIONS?

*Item Nonresponse and (the Absence of) Substantive Opinions
in Public Opinion Surveys*

Public Opinion Without Opinions?

Item Nonresponse and (the Absence of) Substantive Opinions
in Public Opinion Surveys

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 17 september 2019
klokke 11.15 uur

door

Janna van de Maat
geboren te Rijssen in 1985

Promotoren
Prof. dr. Joop J.M. van Holsteijn
Prof. dr. Rudy B. Andeweg

Promotiecommissie
Prof. dr. Jelke G. Bethlehem
Prof. dr. Galen A. Irwin
Prof. dr. Geert Loosveldt, Katholieke Universiteit Leuven
Dr. Tom Louwerse
Dr. Ineke A.L. Stoop, Sociaal en Cultureel Planbureau

Acknowledgements

If someone would have told me at the start of my PhD project that it would take ten years to finish, I probably would not have started it. But 'life is what happens while you are busy making other plans', as John Lennon sang. The road was long and winding, but the finish line is near.

Over the years, a number of people contributed to my dissertation. I want to thank my supervisors Joop van Holsteijn and Rudy Andeweg for their patience and invaluable feedback. Thank you for all your insights, but more importantly for the personal support in both life and career choices.

My gratitude goes out to the LISS Panel (CenterData), the EenVandaag Opiniepanel (particularly Petra Klapwijk and Nina Hoette) and Team Vier (particularly Remco Frerichs and Jelle Rhebergen) who helped collect the original data for my survey experiments. It is not an exaggeration to say that this dissertation would not exist without their help. I would also like to thank Jelke Bethlehem, for his input on the graphical presentation of my data, and Bas van der Schot, for allowing my to use his cartoon for the cover.

The first years of my PhD project were spent at the Institute of Political Science at Leiden University. I thoroughly enjoyed being part of the Institute as a PhD student and lecturer. What I value the most about this time is not the acquired knowledge and academic skills, albeit valuable, but the network of PhD students and colleagues I was and in many ways still am a part of. A special thanks to Cynthia, Fransje, Lotte, Luzia, Tim, Tom and Wouter for being great colleagues and friends. Tim and Wouter: thank you for being my paranymphs and assisting me in jumping over the final hurdles.

The dissertation was finished while working as a Methodologist at the Netherlands Institute for Social Research (Sociaal en Cultureel Planbureau, SCP). I would like to thank Ineke and Jaco for employing me in my dream job. Ineke, thank you for

giving me time and input to finish this dissertation. Thanks to the colleagues at the SCP, particularly in the Methodology section, for your support.

I would like to thank my family and friends for encouraging me all these years and for not nagging much about the (lack of) progress. The 'p'-word is not forbidden anymore!

Finally, the three loves of my life. Orson and Egon, I hope I made you proud. I am glad I can spend more time with you, now that the PhD is finished. And I promise I will finally finish your baby photobooks.

Joost, even though I doubt you will ever read this book, I could not have finished it without you. You encouraged me, took care of the boys when I needed to write, were always positive and believed in me. As a fundamental member of team J&J, I would like to dedicate this thesis to you.

Contents

LIST OF FIGURES

LIST OF TABLES

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)

I

1.1   Abstract

This is a study on item nonresponse and non-substantive answers. Item non-response 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)
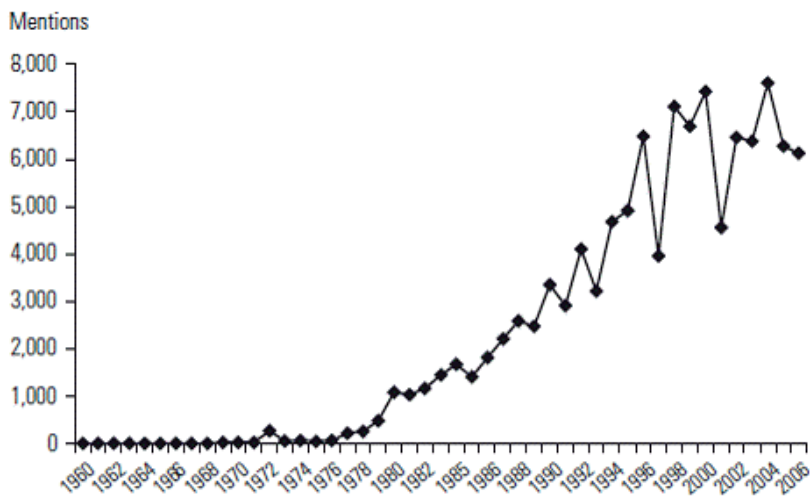
Figure 1.1: Opinion Poll mentions in U.S. news and wires

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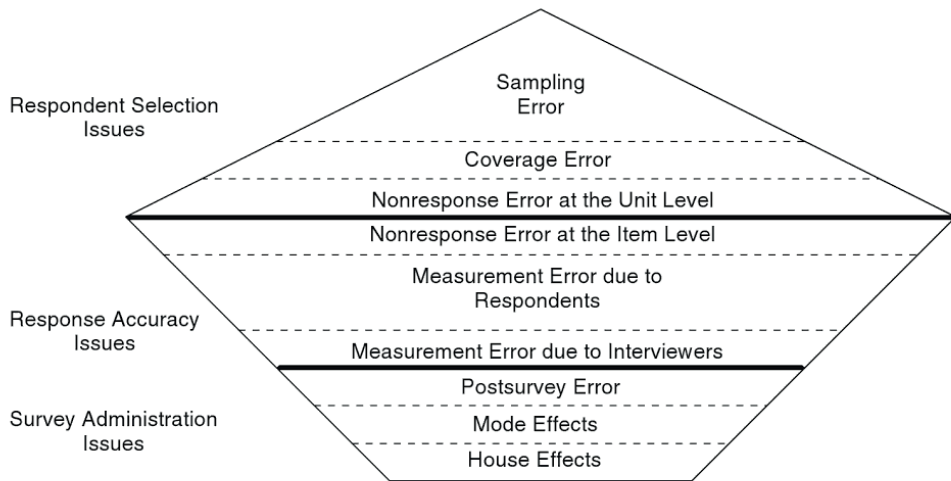
Figure 1.2: Types of Survey Error
Source: (Weisberg, 2005, p. 19)

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.

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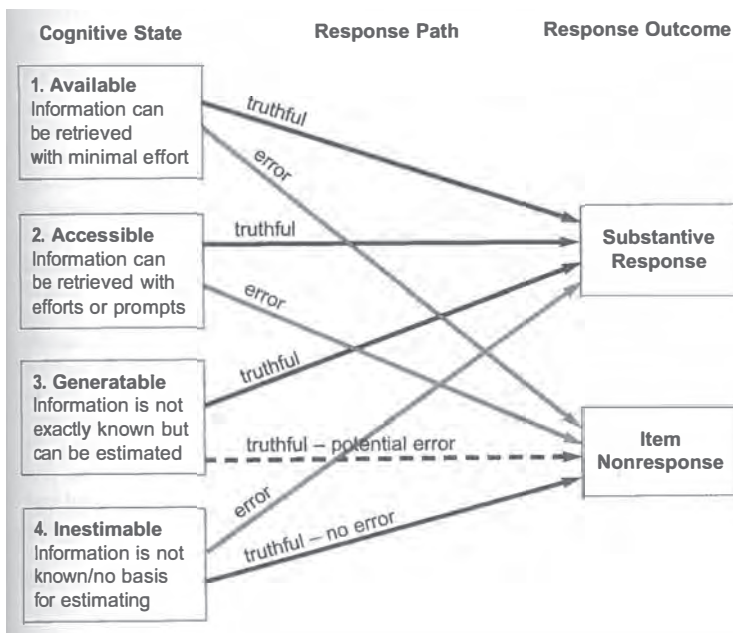
Figure 2.1: Beatty, Herrmann, Puskar & Kerwin's (1998, p. 410) Response Model for Item Nonresponse

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.

CHAPTER 3

Hypotheses on Non-Substantive Response Options

3.1 Introduction

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

3.2 Hypotheses for All Survey Experiments

3.2.1 Question Design and Level of Item Nonresponse

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

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

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

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

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

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

3.2.2   Question Design, Missing Data and Distribution of Opinions

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

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

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

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

3.2.3   Question Design and Question Content

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

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

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

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

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

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

3.2.4 Question Design and Response Categories

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

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

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

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

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

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

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

3.3 Specific Hypotheses for Individual Experiments

3.3.1 Response Time

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

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

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

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

3.3.2 Break-offs

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

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

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

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

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

The hypotheses are summarized in Table 3.1.

Table 3.1: Summary of Hypotheses

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

CHAPTER 4   # Data and Methods

4

4.1   Introduction

All three surveys were conducted via a web panel, but the panels and their composition differ (as will be detailed below). The large N of such web or internet panels made the random composition of subgroups possible to employ a between-subject-design, i.e. a design where 'subjects in an experiment make choices in only one state of the world' (Morton & Williams, 2010, p. 86). The advantage of a between-subject-design is that participants are not aware that they are part of an experiment, as might be the case when they are surveyed twice in a pre-test and a post-test format. The random assignment to subgroups enables the researcher to assume (and check) that the subgroups are similar; differences between the results of subgroups can be ascribed to the manipulation of the independent variable – question design.

The general instruction for all experiments read that the respondents were asked to give their opinion on issues in a survey. They were not aware of the experimental design of the survey; they did know they were participating in a survey to measure public opinion and that the results would be used for scientific research.

In this chapter, several general points are addressed regarding some methodological specifics of this study, including the experimental design and the use of internet panels, before turning to more specific aspects, i.e. the selection of issues for the questionnaire and the question design applied.

4.2   Experimental Research Designs

What are the main advantages of an experimental research design? The textbook answer is that such a design 'engenders considerable confidence in the robustness and trustworthiness of causal findings' (Bryman, 2012, p. 50). By manipulating one or more variables and holding other variables constant, differences in results can be ascribed to the manipulation. 'The unique strength of experimentation is in describing the consequences attributable to deliberately varying a treatment. We call this causal description' (Shadish, Cook, & Campbell, 2006, p. 9). Because of the researcher's ability to manipulate variables and compare between groups, the internal validity of the findings is relatively strong (e.g. Manheim, Rich, Willnat, Brians, & Babb, 2012, pp. 103-104).

A true or full experiment is 'a *randomized trial* in which the researcher randomly assigns units of observation to control and treatment groups' (Druckman, Green, Kuklinski, & Lupia, 2006, p. 628). A crucial characteristic of any true experiment is random assignment. By randomly assigning subjects to either an experimental or a

4

control group, it can be assumed that as regards their composition the groups are similar in all aspects (Manheim et al., 2012, p. 106).

McDermott points to five major advantages of the experimental design: '1) [the] ability to derive causal inferences (...) 2) experimental control (...) 3) precise measurement (...) 4) ability to explore the details of process (...) 5) relative economy' (McDermott, 2002, pp. 38-39). Consequently, and in sync with other developments - technological and methodological innovations, an increasing interest in finding causal mechanisms, the possibility to test and refine theories and the emergence of new research questions (Arceneaux, 2010; Druckman et al., 2006; Morton & Williams, 2010) – experiments have become increasingly popular in political science. Jackson and Cox (2013, pp. 31-32) report an increase from 3 to 8.5 percent of social science articles using experiments between 1990 and 2010.

There are also some disadvantages to doing experiments: '1) artificial environment (...) 2) unrepresentative subject pools (...) 3) external validity (...) 4) experimenter bias' (McDermott, 2002, pp. 39-40). Not all disadvantages apply to each experiment; depending on the characteristics of an experiment some disadvantages may be absent. The main disadvantages are the artificiality of experiments and the limited external validity. The latter disadvantage refers to 'the generalizability of findings from a study, or the extent to which conclusions can be applied across different populations or situations' (McDermott, 2011, p. 34). The external validity is often deemed as lacking, because of the artificiality of the experimental set-up (McDermott, 2011, p. 37) and the composition of the sample which is often non-random and unrepresentative of the targeted population (Jackson & Cox, 2013, p. 35). Three aspects are relevant when assessing the external validity of experimental findings: 'Support from theory or external information (...) 2) use of key characteristics of the studied individuals (...) 3) that the individuals studied are, if not a representative sample in a formal sense, at least broadly representative of the target population' (Jackson & Cox, 2013, p. 35).

Whether the external validity of an experiment is a serious concern, however, depends on the goal of the research (McDermott, 2011). If the research question essentially involves causal inference and internal validity, the experimental research design is the best choice. Internal validity is a prerequisite for drawing more general conclusions with respect to causality and herein lies the strength of an experimental design.

4.3   Survey Experiments

To examine the effects of question design in general and non-substantive response options in particular, doing a survey experiment is appropriate since a single

element can be isolated and studied. A survey experiment is an experiment in the sense that part of the data generating process is manipulated by the experimenter (Morton & Williams, 2010, pp. 30-31). The advantages of doing a survey experiment are summed up nicely by Arceneaux: '[The survey experiment design] possesses strong internal validity, as statistically significant differences in survey responses across question versions constitute strong evidence that differences in question wording are responsible for affecting people's expressed opinions. Yet because survey experiments typically draw on a broader (and sometimes representative) sample of the population of interest, they offer greater external validity than laboratory experiments, which often draw on convenience samples' (Arceneaux, 2010, p. 210). In this way survey experiments combine the best of two methods: causal inference and a realistic setting (Mutz, 2011, pp. 8-13). Survey experiments are artificial because they are set up by the researcher (Jackson & Cox, 2013, p. 43), but the general design corresponds to the way public opinion is usually gauged with polls and surveys. The internal validity of an experimental design can in this way be combined with to some extent externally valid results.

Survey experiments show causal relations. 'A survey experiment (…) is (…) a deliberate manipulation of the form or placement of items in a survey, for purposes of inferring how public opinion works in the real world. The word "experiment" (…) implies random assignment of respondents to control and treatment conditions. Comparing the decisions, judgments or behavior of the respondents in the treatment group to those in the control group reveals the causal effects under investigation' (Gaines et al., 2007, pp. 3-4). Survey experiments are becoming more popular in the social sciences because of the internal ánd external validity of the findings. Barabas and Jerit (2010, p. 226) warn that 'survey experiments generate effects that are observable among particular subgroups, not necessarily the entire population', but even they admit that survey experiments 'can be a valuable tool for studying public opinion' (Barabas & Jerit, 2010, pp. 226-227). If one wants to study the effect of certain question design options, as is the case here, a survey experiment is a most suitable design.

In this study, the internet survey experiment is applied. Examining the effects of question design stems from split-ballot designs, but a more elaborate design is used by, for example, using routing[13] to guide respondents through a questionnaire when they answer filter questions in a certain way. Respondents were randomly assigned to subgroups. Differences between the outcome of the

---

13 Routing (in surveys) means that 'skipping and branching' occurs depending on the individual responses to certain survey questions (Caeyers, Chalmers, & De Weerdt, 2012; De Leeuw, 2008). This process is relatively easy in computer-assisted and internet surveys, because after the questions are programmed correctly the interviewer or respondent is automatically redirected to the fitting next question.

surveys of the separate subgroups should therefore be attributed to the treatment variable: non-substantive response options.

4.4 Internet Surveys and Panels

Web or internet surveys have become popular because they 'allow for simple, fast and easy access to large groups of potential respondents' (Bethlehem & Biffignandi, 2011, p. 2), despite problems like undercoverage of the population, self-selection and nonresponse errors (Bethlehem, 2010; Bethlehem & Biffignandi, 2011; Couper, 2000). While all web surveys use the Internet for data collection, a variety of web surveys can be distinguished. For example, Couper (2000) has made a typology consisting of eight types of web surveys. Two methods are relevant for this study and both are panel-based: the non-probability based 'volunteer panels of Internet users' and the probability-based 'pre-recruited panels of [the] full population' (Couper, 2000, pp. 482-484; 488-490).

Proprietary online panels, also called online or access panels, are panels in which the respondents frequently answer survey questions on the internet (Callegaro et al., 2014, pp. 1-2). Internet panels can differ in the way respondents are recruited. The main distinction is between pre-recruitment and volunteer or convenience panels. Pre-recruitment or probability-based panels aim to include a representative sample of any specified population in the panel by recruiting them via random selection; in volunteer or convenience panels respondents register themselves via self-selection and self-registration (Stoop & Wittenberg, 2008, pp. 8-9). This distinction overlaps with Couper's distinction (2000, p. 477) between non-probability and probability-based sampling which has consequences for the external validity of the findings.

There are several reasons why researchers work with volunteer opt-in panels. Volunteer panels usually have a large sample size; the response rates are often high; surveys are cheap to execute due to the absence of interviewers; and they can be executed very quickly (Couper, 2000; Couper & Miller, 2008; Dillman & Bowker, 2002; Stoop & Wittenberg, 2008). The main problems of volunteer or convenience panels are related to non-coverage and selection bias (Couper, 2000; Hoogendoorn & Daalmans, 2009; Vonk, Ossenbruggen, & Willems, 2008).

The experiments in this study are internet survey experiments executed with panels. This has both negative and positive consequences: on the one hand the respondent may not be the person registered as panel member and the environment cannot be controlled, but on the other hand the response rate is high and interviewer effects are absent. Also, no or at least less social desirability bias occurs (Heerwegh, 2009; Kreuter et al., 2008).

4.4.1 Choice of Panels and External Validity

Three different panels were used for the three survey experiments in this study. For the first experiment the LISS panel was used, which is a pre-recruitment panel composed of a random sample of Dutch households; the panel includes 7,517 potential respondents of over 16 years old. The other two experiments were conducted with convenience or volunteer samples; the second experiment used the EenVandaag Opiniepanel (with about 45,800 potential respondents) and the third experiment the Team Vier internet panel (with about 16,000 potential respondents). The respondents of LISS and Team Vier's panel are paid for their participation, in points or in money, whereas the respondents of EenVandaag do not receive any monetary or other rewards. An overview of the three panels and their characteristics (at the point in time when the experiments were conducted) can be found in Table 4.1.

Even though in only one experiment a random sample is used which allows for generalization to the population – the LISS panel for the experiment with the DK option – any problems with external validity should not be exaggerated. In a report on online panels, the American Association for Public Opinion Research (AAPOR) stated that 'claims of "representativeness" should be avoided'. When generalizing to the population is not the goal of a study, however, '[a nonprobability online panel may be] an acceptable alternative to traditional probability-based methods' (AAPOR, 2010, p. 5). Hence why this study does not draw conclusions about 'the Dutch population' when using nonprobability online panels like the EenVandaag Opiniepanel and the Team Vier internet panel. Furthermore, the limitations of using such panels are discussed both in this chapter and in the concluding remarks in chapter 9. Being transparent about the use and limitations of nonprobability online panels strengthens the findings.

Another reason why external validity does not threaten this study is the focus is on internal validity and causal inference. All panels may suffer from selection bias, due to the recruitment of respondents and panel attrition. Fortunately, the aim of this study is to explore causal effects and not necessarily to generalize to the population, which is the strength of an experimental research design that is employed here (Arceneaux, 2010; Druckman et al., 2006; McDermott, 2002; Morton & Williams, 2010). The experimental design suggests internal validity, which is crucial for the ambition of this research project.

A final point that can be made about the limited generalisability of the findings is that the samples do correspond to the population of the respective internet panel, even if they are not representative of the Dutch population as a whole – depending on the specific panel used. This approach corresponds directly to how mass opinion polls normally would be executed. Internet panels are in practice often used to gauge public opinion and the results are more often than not presented as a representation

Table 4.1: (Internet) Panel Characteristics [a]

	LISS Panel	EenVandaag Opiniepanel	Team Vier Internet Panel
Number of respondents in panel (at the time of survey execution)	7,517	45,780	± 16,000
Sampling	Probability-based random & stratified Over 16 years old	Self-selected, convenience Over 18 years old	Self-selected, convenience Over 15 years old
Non-internet population	Included (via loaned equipment)	Excluded	Excluded
Sample representation	Comparable to high-quality RDD	Overrepresents hyper Internet users	Overrepresents hyper Internet users
Survey frequency	Once a month	Regularly	Regularly
(Financial) incentive	Cash (for each completed questionnaire)	None	Points for cash [b]
Purpose of panel	'Enabling researchers to benefit from existing data, to carry out their own survey or to design a special experiment' [c]	'translate the opinion of thousands of viewers directly to (among others) politics' [d]	'contribute through research to the successful corroboration of policy decisions of customers in the profit and not-for-profit sector' [e]
Average unit response rate	58 to 79 percent [f]	60-70 percent	n.a.
Unit response rate of survey experiment	76.2 percent	64.0 percent	n.a. [g]
Type of issues	Empirical scientific research	Political and social issues; current affairs	A wide range, from retail to magazines, cars and governance [h]

a) The Internet Panel Characteristics table is inspired by the Knowledge Panel (http://www.knowledgenetworks.com/knpanel/index.html; visited on the 3rd of November 2011).

b) The respondent receives 20 points for every 10 minutes of research. After collecting 210 points, the respondent is paid 10 Euros (http://www.teamvier.nl/nl/wie+zijn+we%3F/team+vier+panel, visited on the 27th of May 2016).

c) (http://www.lissdata.nl/lissdata/Home, visited on the 27th of May 2016).

d) Original Dutch text: '*de mening van duizenden kijkers direct te vertalen naar onder meer de politiek*' (http://opiniepanel.eenvandaag.nl/uitleg , visited on the 27th of May 2016).

e) Original Dutch tekst: '...door middel van onderzoek optimaal wil bijdragen aan het succesvol onderbouwen van beleidsbeslissingen van haar klanten in de profit en not-for-profit sector'

f) (https://www.lissdata.nl/lissdata/sites/default/files/bestanden/LISS%20panel%20statistics%202010.pdf, visited on the 27th of May 2016).

g) The unit response rate is unknown, because the survey was closed after the target of 250 respondents completing a variant (for each subgroup) was reached.

h) The respondents can choose whether they want to participate in surveys about all issues or select some of them.

of what the general public wants, regardless of panel characteristics. In other words: this study is not a laboratory experiment where the translation of results to the 'real' world may be difficult (McDermott, 2011, pp. 34-35), but it is rather similar to how public opinion is usually gauged with internet panels.

4.5 Issue Selection

The survey questions for the experiments do not concern prognoses of the outcome of elections, but refer to substantive issues. Questions about facts and/or knowledge are not included. Surveys about substantive issues are done more often than pre-election polls and the results of such surveys are arguably more influential in the political decision-making process since they may provide very specific indications of what the public wants. The surveys deal with 'subjective phenomena' (Turner, 1981) and in line with Schuman and Presser (1996, p. 2), the focus is on 'questions dealing with attitudes, opinions, beliefs, values, preferences, and so on', i.e. 'attitude questions' in which respondents are asked to give their opinion about particular issues and respond to substantive statements. The questions in the surveys covered a range of topics to enable comparison between issues and to test whether the subject of the question matters with regard to the presence or absence of opinions, and the effect of design choices.

All three questionnaires consisted of two parts. The first part contained questions from existing long-term research like the Dutch Parliamentary Election Studies. The same eight questions were included in each questionnaire to enable comparison across experiments. The second part contained questions about current affairs, which were taken from polls and surveys at the time when the experiment was executed. Four general themes were included in the first part of the questionnaire and each theme included at least two questions from existing long-term research. The first three themes (socio-economic, ethical or moral, and multicultural) were included because these are assumed to be indicators of the main dimensions or cleavages in 21st century Dutch politics (Aarts & Thomassen, 2008; Pellikaan, 2010; Pellikaan et al., 2007). The fourth general topic is foreign affairs; this topic was selected because public opinion research suggests that the opinions on such foreign policy issues are often lacking and/or volatile (e.g. Alvarez & Brehm, 2002, p. 214; Everts, 2008, pp. 8-14). The public often does not have (enough) knowledge about foreign policy issues and a feeling of involvement may be missing since these issues usually do not affect the respondent personally; it is thus considered to be a cognitively relatively hard, technical and abstract topic.

The second part of the questionnaire included questions on current affairs

which were selected 'last minute', i.e. shortly before the survey was conducted, to make them as up to date as possible and include questions that were already part of other internet polls, for instance by Peil.nl or EenVandaag. These questions were replications. By replicating 'real'-life survey questions, the effect of manipulating question design could be compared to the original outcome to see whether a different picture of public opinion would be painted when other design choices were made.

The aim of this study was not only to compare between topics, but also between questions on the same topics. For socio-economic affairs, for example, there were differences expected between the general 'income differences' question versus the at the time hotly debated issue 'old-age pension'. The specific questions and design choices are discussed further in the Results chapters where the findings of each experiment are central.

4.6   Question Design

The main characteristic of each experiment is that between subgroups one question design element is manipulated. Everything else, including the content of the questions, the response categories offered, and the question order, is held constant. The experiments all build on the previous one(s) by adding another methodological element, i.e. a non-substantive response option, while maintaining some features of previous experiments. In the second experiment, for example, the DK option that was central in the first experiment is replicated while the filter question is added as a new element. In the third experiment the DK option and the filter question are repeated and the follow-up question is added. This approach of repeating parts of previous experiments has a number of advantages: 1) by reexamining a question design element, the findings can be validated; 2) by combining elements, e.g. the DK option and the filter question, a more detailed and nuanced analysis is possible than when only one element is analyzed in each experiment; and 3) by repeating an element in different panels, the panels can be compared.

Table 4.2 shows the way in which the three question design choices varied. These are a DK option, a filter question and a follow-up question. In the first experiment only the application of a DK option is varied, whereas in the third and final experiment all three elements are included. In each experiment a new element is added, while keeping at least some elements of the previous one(s). The explicit DK option is either offered as a 'Don't Know' or 'No Opinion' response category. These categories are replications of the original questionnaires.

Table 4.2: Question Design Characteristics

	Experiment 1 *Don't Know*	Experiment 2 *Filter Question*	Experiment 3 *Follow-up Question*
DK explicitly mentioned in both question and answer	X		
DK explicitly mentioned as response category	X	X	X
DK implicit as response category	X	X	X
Forced choice	X	X	X
Strongly worded filter question		X	
Weakly worded filter question		X	X
Follow-up Question			X

To illustrate the differences between questions, an example of one question to which all design choices are applied is presented in Table 4.3; see Appendix A for the complete questionnaires in all question design variants.

Table 4.3: Question Design Applied to Example

	Applied in Experiment...	Example Question
DK explicitly mentioned in both question and answer	1	Welfare benefits should be lowered in order to stimulate people to work[14]. Do you agree or disagree with this statement or don't you have an opinion? *Completely agree, agree, disagree, completely disagree, Don't Know.*
DK explicitly mentioned as answer category	1, 2, 3	Welfare benefits should be lowered in order to stimulate people to work. Do you agree or disagree with this statement? *Completely agree, agree, disagree, completely disagree, Don't Know.*
DK implicit as answer category	1, 2, 3	Welfare benefits should be lowered in order to stimulate people to work. Do you agree or disagree with this statement? *Completely agree, agree, disagree, completely disagree, (answer left blank)*
Forced choice	1, 2 3	Welfare benefits should be lowered in order to stimulate people to work. Do you agree or disagree with this statement? *Completely agree, agree, disagree, completely disagree.*
Strongly worded[15] filter question	2	[Introduction] Have you already heard or read enough about welfare benefits to have an opinion? *Yes, No.*
Weakly worded filter question	2, 3	[Introduction] Do you have an opinion on this or not? *Yes, No.*
Follow-up question	3	How upset would you be if the previously expressed opinion did not prevail when the issue was ultimately decided? *Very upset, Upset, A little bit upset, Not upset at all.*

---

14 In Dutch: *'De bijstand moet verlaagd worden zodat mensen gestimuleerd worden om te werken'.*

15 The distinction between the strongly worded filter question and the weakly worded filter question is based on previous research by Bishop (Bishop, 2005, pp. 22-23; Bishop et al., 1983, pp. 530-535), which is discussed more **extensively** in the chapter about filter questions.

The Don't Know Option

5

5.1   Introduction

This chapter describes the first in a series of three survey-experiments as part of a research project which looks at the effects of question design regarding non-substantive response options on survey outcomes, in particular in internet or web surveys. "As far as *web surveys* are concerned, a deeper investigation is needed on the relationships between the questionnaire characteristics and the response patterns" (Ganassali, 2008, p. 22; see also Couper et al., 2001; Dillman, 2007;). In this chapter the focus is on the effects of different ways of offering the DK option. The research question is: *How does the 'Don't Know' option affect the outcome of specific questions and a survey or poll as a whole?* This question will be answered with data from the LISS panel, collected in 2012 by conducting a survey experiment. All other variables are held constant except for the variable of interest, the DK option[16].

5.2   Theoretical Reflection

The DK option is a generic category for respondents who are not able or willing to give an answer, i.e. a substantive answer, to a closed-ended survey question. In this experiment the aim is to explore and analyze what the level of (item) nonresponse is for four different versions of a questionnaire and whether significant differences can be found both with respect to the way this DK option is offered and as regards the topics that are being addressed in the survey questions. All survey questions included in the experiment concern attitudes.

The debate on non-substantive answers and how to handle this problem is old but not yet concluded. There is still disagreement in the literature about how best to handle non-substantive answers, i.e. item nonresponse, for opinion and attitude questions(see e.g. Gilljam & Granberg, 1993; Heerwegh & Loosveldt, 2008; Krosnick et al., 2002; Leigh & Martin, 1987). In practice, respondents are usually encouraged to give their opinion: "The typical practice - what we shall call the *standard question form* - is not to include a DK alternative as part of a question" (Schuman & Presser, 1996, p. 113). Schuman and Presser (1996) state that only spontaneous DK answers are registered in this standard practice; an explicit DK is not part of the response categories. The reason is that researchers want to increase their item response rate and collect as many substantive responses as possible.

Some scholars differentiate between reasons for using a non-substantive response option, e.g. lack of information, or a polite refusal to a difficult or sensitive

---

16   A pilot experiment was conducted among high school students (see Van de Maat, 2009).

question (Bradburn et al., 2004, p. 353), but such a distinction is not made here. The question why respondents use a DK option is relevant but empirically very difficult to analyze. More importantly, the aim in this study is to look at levels of item nonresponse and the resulting picture of public opinion regardless of the reasons for these non-substantive answers. The quantity and not the quality of item nonresponse is examined here. In this experiment the aim is to analyze what the level of item nonresponse is for each version of questionnaire and whether significant differences can be observed. People may give a DK answer for various reasons, but here the aim is first and foremost to establish general *levels* of item nonresponse and not to differentiate. The additional analysis of response time in this chapter may, however, give some indication why respondents use a DK option.

5.3   Hypotheses

In chapter 3 several hypotheses were developed. These hypotheses are summarized in Table 5.1.

Table 5.1: Hypotheses

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

5.4 Data and Methods

For the first experiment, the LISS Panel, Longitudinal Internet Studies for the Social sciences[17], was used. This panel forms the core of the Measurement and Experimentation in the Social Sciences (MESS) project; it was funded between 2007 and 2014 by the Netherlands Foundation for Scientific Research (NWO) to give

---

17   www.lissdata.nl for a description of the panel and examples of previous experiments which have used this panel. The LISS panel is administered by CentERdata (Tilburg University, the Netherlands). The author would like to thank the LISS panel for their support and assistance in setting up and executing the survey experiment.

researchers the opportunity to collect data via surveys. On average, 73 percent of the panel members complete the questionnaires. See section 4.3 for a more extensive description of the LISS Panel characteristics.

The large N of the LISS panel allows to divide respondents into subgroups which were subjected to one particular question design and stimulus (see Table 5.2 below) and to subsequently compare the data. The full panel was targeted, with random selection of the four distinct groups for the various treatments.

The general instruction of the questionnaire read that the respondents were asked to give their opinion on issues in the poll or survey. Scales for self-placement were placed horizontally while the response categories of the other options were ranked vertically. The experiment was carried out between February 1st and February 29th 2012; respondents were given a month to complete the questionnaire.

Table 5.2 contains descriptive statistics of the sample and subgroups to check the comparability of the subgroups and the representativeness of the sample. Overall the subgroups did not differ from each other. The only exception is a small but statistically significant difference between subgroups in terms of gender – see table C.1. in the appendix. On average 47 percent of the respondents was male, but in subgroup 3 they made up 50 percent of the subsample. The average age of the respondents was 49.9 years with no significant difference between subgroups. In terms of the highest level of education, the subgroups also were similar with no significant difference between groups. The average monthly individual income of the household member filling in the survey was 1,495 euro. There was no significant difference in income found between subgroups.

Table 5.2: Descriptive Statistics Subgroups LISS Panel

		1. Double Explicit DK	2. Single Explicit DK	3. Single Implicit DK	4. Forced choice	*Total*
Gender*	Male	45%	46%	50%	46%	47%
	Female	55%	54%	50%	54%	53%
Age (years)	Mean	49.7	50	49.7	49.8	49.9
	SD	17.3	17.3	17.3	17.0	17.3
Education	Basisonderwijs	10%	10%	9%	10%	10%
	VMBO	26%	25%	26%	26%	26%
	Havo/VWO	11%	12%	11%	11%	11%
	MBO	22%	23%	24%	22%	23%
	HBO	22%	23%	23%	23%	23%
	WO	9%	8%	8%	8%	8%
Income	Mean	1426	1494	1609	1456	1495
(Euros/month)	SD	1030	2937	5156	2086	3149
N		1468	1464	1375	1421	

* Any pair-wise comparison with variant 3 is significant at the 0.05 level
Gender: percentage of the sample that is male or female.
Age: Average age in years, with respondents being 16 and older.
Education: Highest level of education completed, recoded into CBS categories.
Income: Average individual income of the household member.

The overall similarity of the subgroups is important to make comparisons between the variants. The comparability of the sample as a whole with the population is less important, since it is an experiment with a between-subjects-design and the key conclusions are drawn after comparing the four subgroups. In order to be able to generalize the findings to the general population, however, the sample should not differ too much from this population. Comparing the sample with the population is problematic, however, due to the lack of population data. Statistics Netherlands (CBS)[18] offers information about the population, but these figures are not directly comparable to the sample[19]. Nevertheless, based on the random sampling method, it is assumed that the sample is comparable to the population and that generalization to this population is possible.

---

18   Statline.cbs.nl with data from 2011.
19   According to CBS, 49.49 percent of the population is male. The average age of the population is 40.3 years, but this includes people who are younger than 16 and these are not included in the LISS Panel. 6% of the population finished Basisonderwijs, 20% VMBO, 42% HAVO/VWO & MBO, 21% HBO and 11% WO. Finally, the average income for the population as a whole is 22,300 euro a year including social security payments, welfare, wages, holiday pay and bonuses.

Table 5.3: Response Rates

Variant	Number of Respondents	Response Rate – Unit	Number of Items in Questionnaire	Average Item Nonresponse
1. Double explicit DK	1468	78%	14	16%
2. Single explicit DK	1464	78%	14	14%
3. Single implicit DK	1375	73%	14	1%
4. No DK, forced choice	1421	76%	14	X
Total	5728	76%	14	

Table 5.3 presents some characteristics of the experiment, including the unit (non)response rate and the average item nonresponse rate of each variant of the questionnaire. Out of the 7,517 individual members of the LISS panel, 5,728 participated in the experiment, i.e. an average unit response rate of 76 percent, with rates ranging from 73 to 78 percent for the individual variants. All respondents answered 14 items relevant for this experiment in their respective variant of the questionnaire.

The questions covered a range of topics. Five general themes were addressed and each theme included three questions of which at least one question came from existing research – see section 4.4 for more information about issue selection. The aim of this part of the experiment is to compare between themes, but also within themes. For the socio-economic theme, for example, there are differences expected between the general 'income differences' question versus the at that time discussed particular issue 'old-age pension'.

The stimulus in the survey experiment is a 'Don't Know' (DK) response category which was offered (1) explicitly in both the question wording and as a response alternative; (2) not mentioned in the question wording but explicitly offered as a response alternative; (3) not mentioned in the question wording nor as an explicit response alternative, but with the possibility to skip the question without giving a substantive answer; and finally (4) not mentioned in the question wording nor as an explicit response alternative, and with no possibility to skip the question without giving any substantive answer[20]. This final forced choice variant is often used in

---

20   The ways the DK option was offered would be labelled as DK encouraging and DK neutral by Luskin and Bullock. In order to examine the effects of several DK designs on the measurement of political knowledge, Luskin and Bullock (2011) applied three ways of offering a DK option: 1) DK discouraging with no explicit DK option and a probe to get more substantive answers; 2) DK encouraging with an instruction to give a DK answer when unsure and an explicit DK option; and 3) DK neutral. In this survey experiment, no variant was used where the respondent was probed for a response, since this is uncommon in standard web surveys.

'interactive self-administered surveys' (Derouvray & Couper, 2002). The number and choice of answer categories was replicated from the original questions and not altered in any way; only the DK option was manipulated. The DK option was the only available non-substantive response option; no comparison was made between the use of either Don't Know or No Opinion as a means to register non-substantive answers.

Table 5.4 shows the four versions of the questions/questionnaire (see Appendix AI for the full questionnaires). And by way of illustration, Figure 5.1 shows the variations of a survey question, depending on the variant offered to the respondent. The parts in italics vary.

Table 5.4: Overview of DK Alternatives [a)]

	DK alternative offered?	
Type of question	in question wording	as response alternative
1. Double explicit DK	+	+
2. Single explicit DK	-	+
3. Single implicit DK	-	±
4. No DK	-	-

a) In the single implicit DK alternative there is no explicit DK response category, but it is possible to skip the question without giving any answer; in the No DK alternative there is no DK response possible and it is not possible to continue with the questionnaire without giving an answer to the question.

'Some people think that euthanasia should be forbidden by law. Others feel that a doctor should always be allowed to end a life, if the patient makes that request. Of course, there are also people whose opinions lie somewhere in between. Suppose that the people (and parties) who think that euthanasia should be forbidden are at the beginning of this line (at number 1), and the people (and parties) who feel that a doctor should always be allowed to end a life upon a patient's request are at the end of the line (at number 7).

Where would you place yourself on the line?' *If you have no idea at all which position a party has, then please feel free to say so.*

1.   Euthanasia should be forbidden.

2 - 6

7.   A doctor should always be allowed to end a life upon a patient's request.

99.   *Don't Know*

Figure 5.1: Survey Question Variations

5.5   Results

In this section, the results are presented of the survey experiment in which a DK option was manipulated. The presentation of results is structured according to the hypotheses, starting with an analysis of item nonresponse. This analysis includes both the number of times the DK option was used and the registered time to answer the question. After that, the relation between missing data and the overall distribution of opinions is examined, followed by an inventory of differences between question content and number of response categories.

5.5.1   Item Nonresponse

Figure 5.2 shows the average item nonresponse for the complete questionnaire for the four question variants.
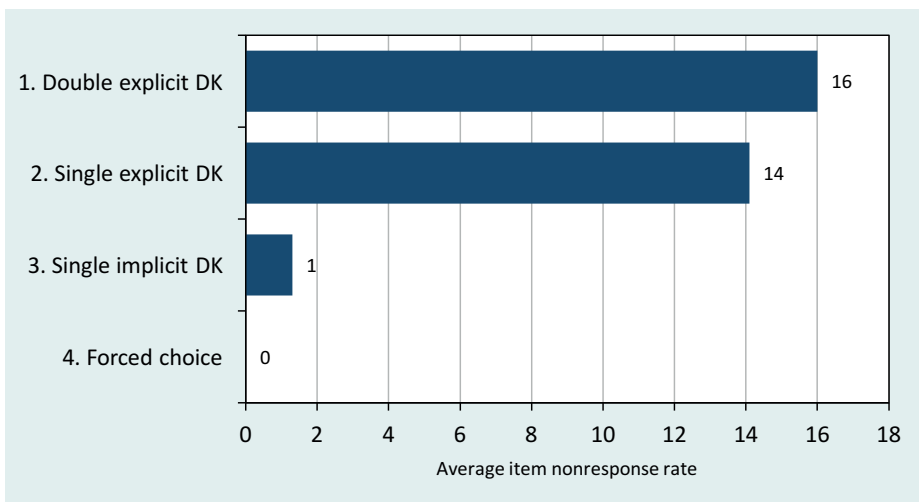
Figure 5.2: Average Item Nonresponse Rate (in Percentages)

Figure 5.2 shows the average item nonresponse rate of the four questionnaire variants; all respondents answered 14 items and the variants only differed with respect to their DK option. There is a substantial difference between variants explicitly offering a DK option, i.e. variant 1 and 2, and variant 3 which made item nonresponse possible by skipping the questions, but did not offer this option explicitly as part of the question. Despite the explicit instruction at the start of variant 3 that skipping questions was possible, this option was not used often. The instruction was given because the respondents of the LISS panel are usually not

given this option and would therefore probably not consider it an option. This could be the reason the respondents did not skip questions often, as they might not have been aware the option existed. The instruction was given in the introduction of the questionnaire.

In variant 1 respondents were twice pointed to the possibility of using a DK option, by explicitly mentioning the DK option in the question and as a response category, as opposed to only a DK response category in variant 2. The difference in average item nonresponse rate is 2 percent between variant 1 (16 percent) and 2 (14 percent). To test whether the difference in item nonresponse rate was statistically different, a t-test was executed; the difference is small but statistically significant at the .001 level[21]. The main difference exists, however, between both these explicit DK variants and the implicit or forced choice variants 3 and 4. When respondents could skip a question but were not reminded of that option (variant 3), the item nonresponse rate dropped to 1 percent. These overall results show that the way a DK option is offered has a significant and substantial influence on item nonresponse, as stated in H1a. Offering a DK option more explicitly results in more item nonresponse; note that offering such an option explicitly once already makes a major difference.

In Table 5.5 the item nonresponse for all individual items in the explicit and implicit DK variants are shown; the forced choice variant is not included, since here respondents by design could not use a non-substantive response option.

All individual items show a large difference between the variants explicitly offering a DK option and the single implicit DK version. The item nonresponse of all items in variant 3 is significantly (at the .001 level) and substantially lower than for the same items offered with an explicit DK option. The highest level of item nonresponse in version 3, which is for the UN item, amounts to 4 percent; this would rank amongst the lowest level of item nonresponse in variant 1 and 2.

The difference in item nonresponse between the implicit and explicit DK variants is highly significant (as indicated by .001 significance levels of all comparisons with implicit DK variant 3) with consistently lower item nonresponse measured for all individual items in the implicit DK variant. This is in line with hypothesis H1a: offering a (non-substantive) DK option more explicitly results in more item nonresponse. At least this is the case when comparing variants which offer DK as an explicit response category with a variant which offer an implicit DK option by giving the respondent the opportunity to skip questions. The

---

21 In addition to the t-test, a negative binomial regression was performed with the number of non-substantive answers as a dependent variable, to account for the fact that the data are not normally distributed. The results were almost identical to the t-test.

Table 5.5: Item Nonresponse (%) of Individual Items

Question	1. Double Explicit DK	2. Single Explicit DK	3. Single Implicit DK	Significance Paired Comparisons		
				1 / 2	1 / 3	2 / 3
Self-placement Income Differences	7	5	0	.094	.000	.000
The old-age pension age should be preserved at 65	7	6	0	.247	.000	.000
The welfare benefits should be lowered in order to stimulate people to work	13	11	1	.044	.000	.000
Self-placement Euthanasia	3	4	1	.755	.000	.000
Adoption by same-sex couples should be possible	12	11	1	.374	.000	.000
It is right that women can freeze their ova to be able to have children at a later age	18	17	1	.321	.000	.000
Self-placement Foreigners	2	2	0	.627	.000	.000
There are too many people of a non-Dutch nationality living in the Netherlands	16	12	1	.012	.000	.000
All people living in the Netherlands illegally for a long time should be allowed to stay here	14	13	2	.466	.000	.000
Self-placement EU	12	10	1	.202	.000	.000
The Netherlands should spend more money on developmental aid	13	12	1	.379	.000	.000
The United Nations has too little power	39	35	4	.082	.000	.000
The Queen can only communicate governmental policies towards journalists	35	29	2	.001	.000	.000
Are you pro or con surrogate motherhood?	33	29	2	.029	.000	.000
N	1468	1464	1375			

Item Nonresponse is measured as a percentage of the total number of respondents not responding to a certain survey item by using the DK option (in variant 1 and 2) or skipping the question (in variant 3).
Significance Paired Comparisons shows the significance level of the differences in item nonresponse between two questionnaire variants.

inclusion of a DK option as an explicit response category results in significantly more item nonresponse than when an implicit DK option is provided.

The effect of an explicit DK option is, however, less clear-cut when the two explicit DK variants (1 and 2) are compared. Out of a total of fourteen items in Table 5.5, ten do not show a statistically significant difference (at a .05 level) in item nonresponse between the double explicit DK variant 1 and the single explicit DK variant 2. In other words: the number of respondents using a DK option is for most items not affected by mentioning the DK option in the question itself (next to explicitly offering DK as a response category).

One can conclude that only a weak relation exists between mentioning the DK option in the question on top of offering it as a response category and the use of the DK option. The main difference is between offering a DK option explicitly or implicitly. The comparison of explicit DK variants with the single implicit DK variant does show large and significant results, consistent with hypothesis H1a: more item nonresponse is generated when a non-substantive response option is offered more explicitly.

5.5.2   Response Time

The analyses above do not disclose any reasons for using the DK option; an analysis of the response time may give an indication how the DK option is used. If variants with an explicit DK option are answered significantly quicker compared to variants without such an option (forced choice) or with only an implicit option (skip the question), this may suggest that the DK option is also used as an easy way out to complete the questionnaire as quickly as possible. Here it is expected (in hypothesis H5) that shorter response times indicate that respondents in web surveys are using the DK option as a short-cut, by taking less time to formulate a response to the survey question and using the DK option as an easy way out (see e.g. Greszki, Meyer, & Schoen, 2014, 2015; Malhotra, 2008; Yan & Tourangeau, 2008; Zhang & Conrad, 2013).

For each respondent in the internet survey experiment the response time in seconds was registered for each question. This enables a number of analyses. First, the differences in total response time per type of questionnaire can be examined to explore whether in explicit DK variants less response time is registered than in other variants. Secondly, the average response time of individual questions can be compared to find out whether certain types of questions result in more use of the DK option as an easy way out than other questions; the focus is then on differences between individual questions and question content. This second analysis is included in Appendix B.

For all analyses, the maximum response time for an individual question has been

Double explicit DK	196
Single explicit DK	186
Single implicit DK	183
Forced choice	181

Average total response time (in seconds)
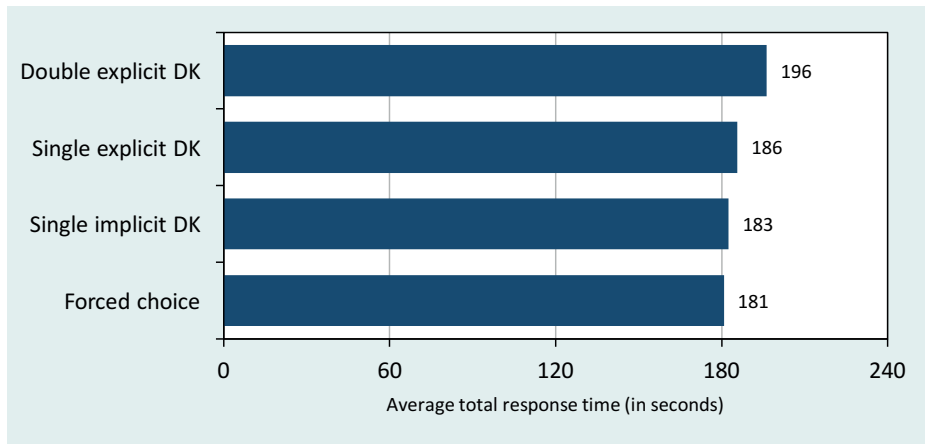
Figure 5.3: Average Response Time per Variant for All Survey Questions

set at 180 seconds and negative values were excluded[22]. Most respondents, however, spent much less time on answering individual survey questions than the maximum of 180 seconds. Figure 5.3 shows the response time of the average respondent for the *total* number of questions, compared between the four variants of the questionnaire. The introduction is not included in the total response time analysis, since it was held constant for all questionnaires. It took the average respondent between 181 and 196 seconds to complete the questionnaire.

If DK is an easy way out, the average response time should be lower in the explicit DK variants than in the single implicit DK and forced choice variants. Figure 5.3, however, shows the opposite relation between explicitness of the DK option and

---

22 There were some anomalies in the data, including negative response times and extremely long response times for individual questions up to 32,341 seconds, which would mean that someone took 539 minutes to answer a single survey question. Negative response times result from opening the questionnaire and/or completing some questions, and returning the next day to finish the survey. The individual response times were calculated by comparing the time the previous question was completed to the completion time of the next question, without including the date: therefore negative values arose. These cases are excluded from the analyses (39 cases). Out of these 39 cases, 26 cases occurred for the first survey question; the respondent read the introduction and opened the questionnaire, but started answering at a later point in time.

The extremely long response times are also very likely caused by respondents pausing at some point and finishing the questionnaire later. Here it is more difficult to establish when a response time is too long; the maximum has slightly arbitrarily been set at 180 seconds per question. Out of 5,728 participants in the study, the response time of 282 respondents (4.92 percent) was excluded from the analysis for one or more individual questions.

response time: the less explicit a DK option was offered, the lower the response time. The average total response time was 196 seconds for the double explicit DK variant, 186 seconds for the single explicit DK, 183 seconds for implicit DK and 181 seconds for the forced choice variant. The response time in the double explicit DK variant is significantly higher than the response time of the other variants[23]; the other questionnaire variants do not differ significantly from each other. So offering a DK option does not seem to result in quick DK answers as an easy way out. To the contrary: more time is spent and arguably more thought is given to the questionnaire in variants where not giving an answer is an (explicit) option. What the response time analysis strongly suggests is that, contrary to hypothesis H5, the DK option is not used as an easy way out, since that would result in shorter response times rather than the longer response times that were observed.

5.5.3 The Overall Distribution of Opinions – Towards Public Opinion

There are two elements important in the examination of the link between question design and survey outcomes: 1) the level of item (non)response which reflects whether respondents actually give a substantive answer; and 2) the actual overall outcome in terms of minorities and majorities as part of public opinion. This second aspect is the focus of this section: how are opinions distributed and is this distribution affected by question design regarding non-substantive response options?

Figure 5.4, Figure 5.5, Figure 5.6 and Figure 5.7 present the distributions of opinions of four survey items. In this section, item nonresponse is treated as missing data and excluded from the findings. The distributions of opinions of all survey items and the significance tests (of the differences between questionnaire variants) can be found in Appendix C.

Does the way a DK option is offered change the resulting overall picture of public opinion? The short answer is no. When item nonresponse is excluded (i.e. missing data), the resulting distributions of opinions and overall public opinion show hardly any substantive differences. Despite some statistically significant between-variant-differences, the response option preferred by most respondents barely varies – see Table C.2 and Table C.3 in Appendix C.

The effect of the DK option on the distribution of opinions is weak or even non-existent, for both self-placement items and other survey items. For example,

---

23 A pairwise comparison with a t-test shows significant differences in average response time between the double explicit DK and single implicit DK and the double explicit DK and forced choice variants (at the .001 level). The double explicit and single explicit DK variants differ significantly at the .005 level. Other comparisons (between single explicit and single implicit DK, single explicit DK and forced choice, single implicit DK and forced choice) do not show significant differences in average response time.
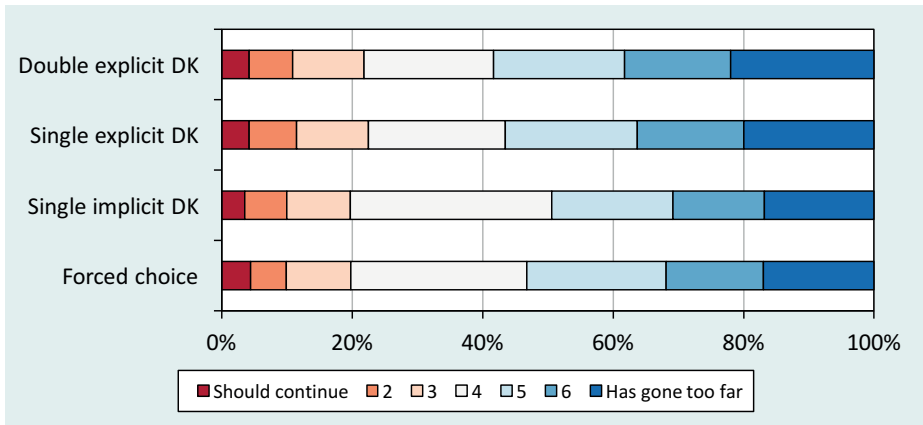
Figure 5.4: Distribution of Opinions Self-Placement European Unification
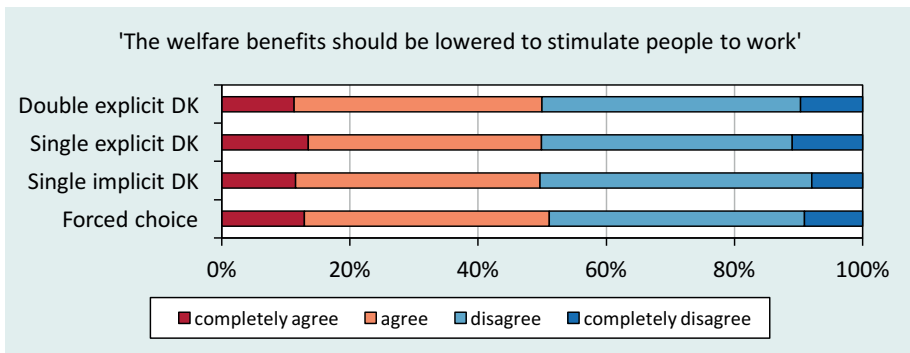
Figure 5.5: Distribution of Opinions Lowering Welfare Benefits
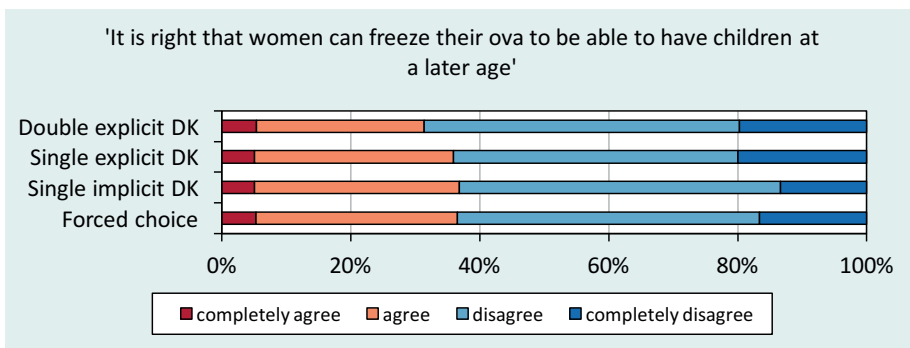
Figure 5.6: Distribution of Opinions Ova Freezing

'Are you pro or con surrogate motherhood?'
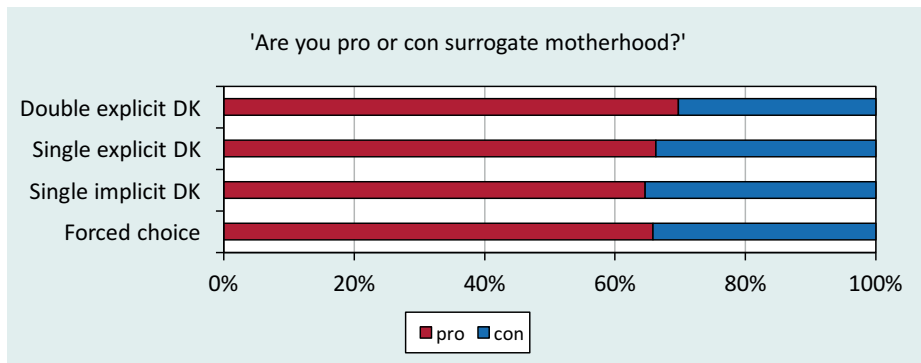
Figure 5.7: Distribution of Opinions Surrogate Motherhood

a majority of respondents disagrees completely or disagrees with the statement 'it is right that women can freeze their ova to be able to have children at a later age', regardless of which questionnaire variant they were assigned to; the outcome was very similar for the four subgroups (see Figure 5.6). This rather surprising finding supports hypothesis H2a: missing data seem to be missing at random. If the data were not missing at random, significant substantive differences would show for the various questionnaire variants. The overall substantive distributions are, however, robust and the question design effect on overall results is (with item nonresponse excluded as missing data) at best a weak effect.

So would it matter which variant was used if a politician wanted to know what the public wants? For most items the answer to this question is no. The numbers may differ a few percentage points, but this does not change the fact that most respondents think for example it is not right for women to be able to freeze their ova to delay having children (see Figure 5.6) or that most respondents (in all variants) are pro surrogate motherhood (see Figure 5.7). The only times that offering a DK option affects the results and majorities change, are when proponents and opponents are tied. The question about lowering welfare benefits (Figure 5.5), for example, results in about 50 percent of the respondents agreeing and 50 percent disagreeing with the statement. In the forced choice variant, the respondents agreeing with the statement add up to 51 percent, i.e. a (very small) majority that does not show in the other variants and may simply not be a majority given the confidence interval. So it could be argued that such a 'majority' is too small for any politician to use as an indication of what the general public wants, except to argue that the public is divided[24].

24 Other survey questions with a tied outcome are non-Dutch nationality and the power of the UN in and .

The preliminary conclusion about the impact of a DK option on the outcome of survey questions is that if item nonresponse is excluded as missing data, the effect is small or even nonexistent. If, however, item nonresponse is included in the distribution of opinions, the picture of public opinion changes substantially, at least for some items. Figure 5.8 - Figure 5.11 display two pictures of public opinion, one excluding DK answers (in variant 1 and 2) and skipped questions (in variant 3) as missing data and the other including item nonresponse as part of the overall outcome. The substantive answers were dichotomized for these analyses. See Appendix C for an overview of all distributions of opinions; the self-placement items are excluded.
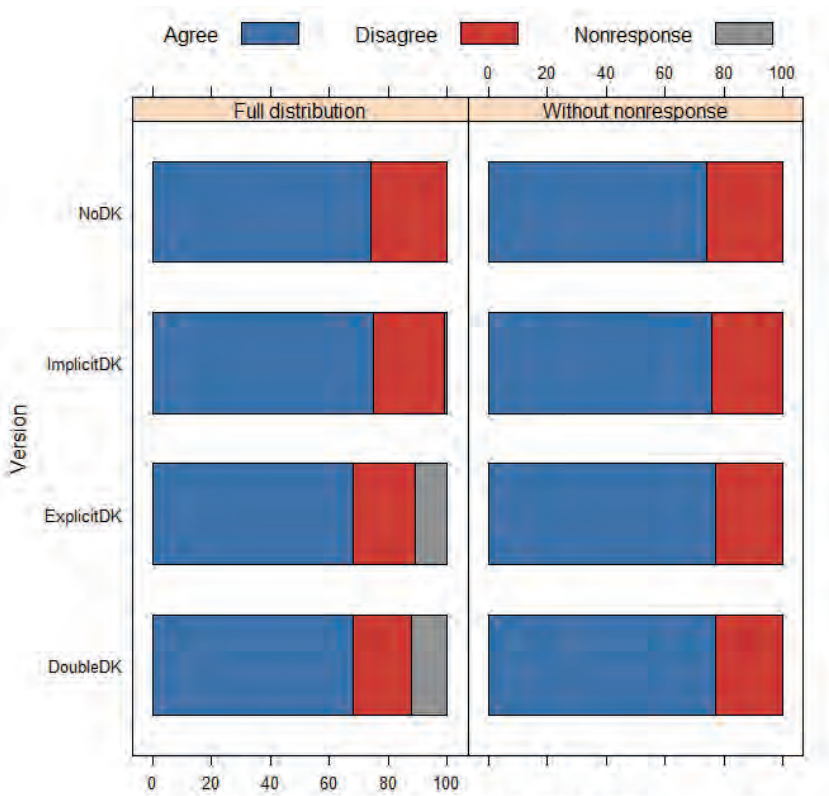
*Figure 5.8: Distribution (%) of Opinions **Same-Sex Adoption** With and Without Item Nonresponse*
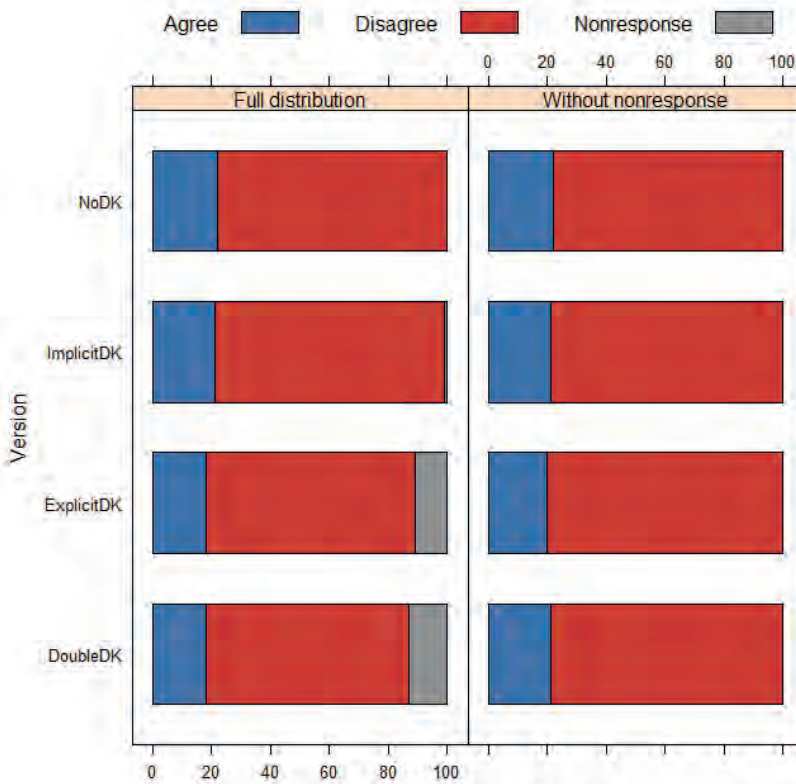
Figure 5.9: Distribution (%) of Opinions **Developmental Aid** *With and Without Item Nonresponse*

When item nonresponse – i.e. DK answers (in variant 1 and 2) and skipped questions (in variant 3) – is treated as a valid and substantively interesting response category, a number of differences may show between the various variants: 1) the preference of the largest group of respondents, either a plurality or majority, is not the same in all four variants; 2) the preference of the largest group of respondents is the same in all four variants, but it is not supported by a majority in all variants; and 3) in some subgroups the response category used by a plurality of the respondents is the DK option. The first effect is arguably the most interesting but also worrisome, because it means that a different choice of question design regarding non-substantive response options would create a different picture of public opinion. Not only is the extent of the support for a particular policy stance affected, but the actual policy stance itself would appear to be different in at least one of the variants. Such an effect of offering a non-substantive response option is visible in three items. Lowering welfare benefits, too many non-Dutch and too little UN power result in
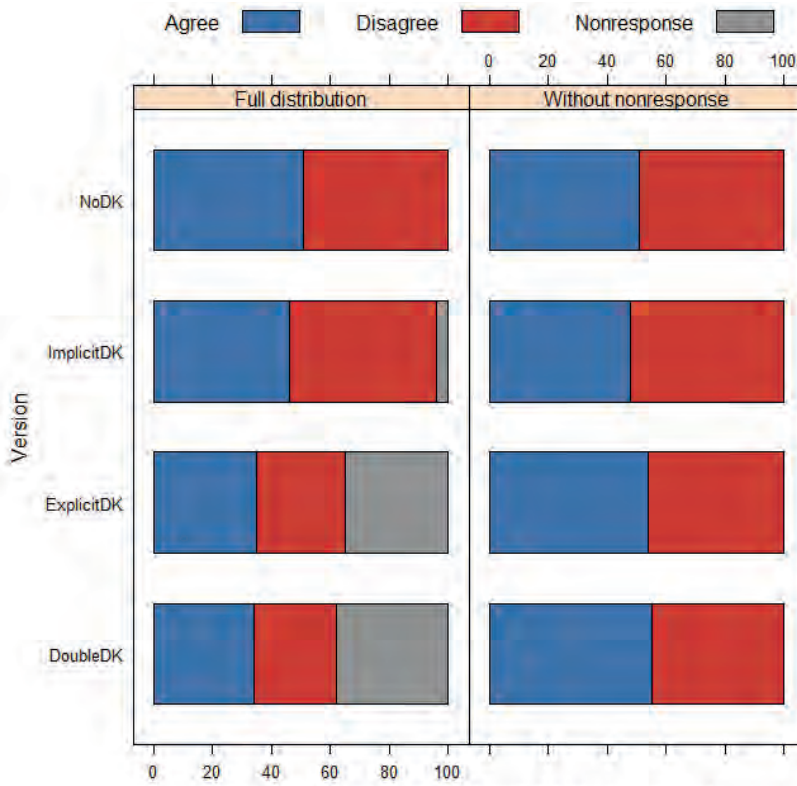
*Figure 5.10: Distribution (%) of Opinions **UN Power** With and Without Item Nonresponse*

a different preferred policy position, but the preferences are not supported by a majority of the respondents in all variants. For example, most respondents of the single explicit DK and forced choice variants agreed with the statement that the UN has too little power (see Figure 5.10), whereas most respondents of the other two (double explicit DK and single implicit DK) variants disagreed. Only in one (forced choice) variant, however, this most preferred position was supported by a majority of the respondents – with 51 percent. The other pluralities ranged between 34 and 50 percent.

The second effect, i.e. same preferred stance in all variants but with varying degrees of support, is visible in two items: the role of the Queen and surrogate motherhood. Depending on the variant, 47 to 66 percent of the respondents were pro surrogate motherhood (see Figure 5.11). The degree of support corresponded to the registered level of item nonresponse; in (explicit DK) variants with more item nonresponse the level of support for surrogate motherhood is lower.
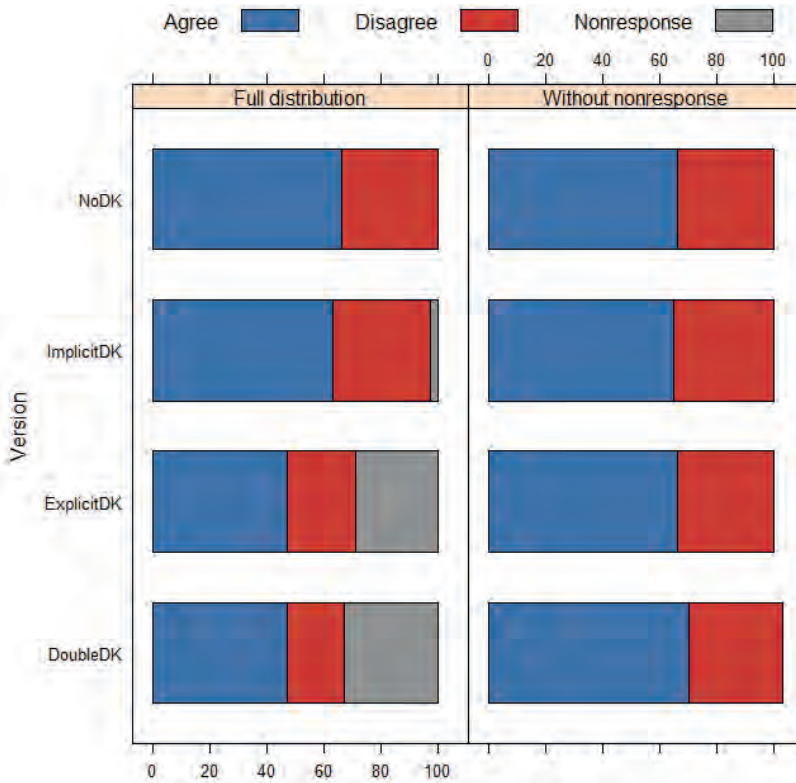
*Figure 5.11: Distribution (%) of Opinions **Surrogate Motherhood** With and Without Item Nonresponse*

The third effect only appears for the UN power item, where a plurality of respondents of the double explicit DK (38 percent) and single explicit DK (35 percent) variants used the DK option (see Figure 5.10). Together with the fact that the most preferred position is not the same in all variants, the conclusion is that these data make it hard to say what the public thinks about the powers of the UN.

All three categories of differences affect the impression of public opinion and eventually affect the decisions a politician makes based on these pictures of public opinion. It should, however, be noted that for five items no substantial differences are visible between the four questionnaire variants when item nonresponse is included. These items show the same outcome in terms of policy preference, which is supported by a majority of respondents in all four variants. So regardless of whether and how a non-substantive response option was offered, 68 to 75 percent of the various subgroups agreed with the statement that adoption by same-sex couples should be possible (see Figure 5.8). Likewise, 58 to 64 percent of the respondents

of the four variants expressed that the Netherlands should not spend more money on developmental aid (see Figure 5.9). The five items without substantial between-subgroup-differences, which amounts to half of the non-self-placement items included in the questionnaire, indicated the same majority preference for all variants of the questionnaire; the only difference was that the majority was slightly smaller in the explicit DK variants, due to the higher item nonresponse rate.

These subgroup comparisons focused at the differences between the variants when item nonresponse is included as a valid and substantively interesting response category. Another way to look at the effect of non-substantive response options on the distribution of opinions is to compare the distributions with and without item nonresponse. The aim is not to look at the preferred position, but whether a majority remains a majority when item nonresponse is included. If we look at the outcome of the same variant including and excluding item nonresponse, does the picture of public opinion change? Six of the ten non-self-placement items show only a plurality supporting the most preferred position in the explicit DK variants, compared to a majority in the other variants. Depending on questionnaire design, a (clear) majority is visible or just a plurality.

All in all, the conclusions about the effect of offering a DK option (explicitly) on the survey results are twofold: firstly, the substantive effect on the distribution of opinions is rather small or even non-existent when item nonresponse is excluded as missing data; secondly, the statistical and substantive impact on the overall distribution of opinions including a DK response category is more substantial. The main difference is whether DK responses are viewed as substantively valuable information; otherwise, i.e. if item nonresponse is excluded as missing data, the picture of public opinion hardly changes at all. For the resulting distribution of opinions, whether and how a DK option is offered in general does not affect survey results if item nonresponse is excluded. The devil may be in the details, however, which is why the analysis continues with a more detailed comparison of question content.

5.5.4 Question Content

Are some themes more susceptible to question design effects than others? Does question content matter? The hypotheses regarding question content were inspired by two ideas: whether or not an issue ties in with a major political dimension, and whether an issue is related to foreign policy. The analysis of question design effects only looks at item nonresponse; it has been established already that the distribution of opinions (when item nonresponse is excluded as missing data) is not substantially affected by the way a DK option was offered, but there were a few significant differences.

Table 5.6: Item Nonresponse (%) for Themes [a]

	Number of Nonresponses	1. Double Explicit DK	2. Single Explicit DK	3. Single Implicit DK
Socio-economic	1	17	14	2
	2	4	3	0
	3	1	1	0
	Total	22	18	2
Ethical	1	20	18	2
	2	5	5	0
	3	1	1	0
	Total	26	24	2
Multicultural	1	19	17	2
	2	5	4	0
	3	1	1	0
	Total	25	22	2
Foreign Affairs	1	30	29	4
	2	12	9	1
	3	3	3	0
	Total	45	42	5
Foreign Affairs without UN outlier	1	17	14	2
	2	4	4	0
	Total	21	18	2
Current Affairs	1	39	37	3
	2	15	11	1
	Total	54	48	4

a) Number of Nonresponses is the number of times an individual respondent used the DK option within a certain theme. The 'Total' rows show the percentage of respondents using the DK option at least once.
The significance of the differences between the means of the questionnaire variants can be found in Appendix C; the comparisons of variant 1 with 3 and variant 2 with 3 are significant at the .001 level for all items in a t-test and a negative binomial regression.

Table 5.6 shows how often respondents used the DK option within a theme; it shows that most respondents only occasionally give a non-substantive answer. Also, when respondents do not answer all items they usually only use the DK option once. No more than 3 percent in the double explicit DK variant 1 and 3 percent of the respondents of the single explicit DK variant 2 respond to all questions within one theme with Don't Know – excluding the Current Affairs theme where only two questions were asked. This suggests that the DK option is used selectively. The option to skip questions (in variant 3) is hardly used at all, which is why it is not

discussed further below. Variant 4 is excluded from Table 5.6, since respondents could not give a non-substantive answer.

What is noticeable is that the question design effect is relatively robust. All comparisons of a single implicit DK variant 3 with an explicit DK variant show significant differences at the .001 level (see Appendix C); a comparison of the two explicit DK variants shows statistically significant differences for only some themes. This is consistent with the main effect.

The percentage of respondents giving a DK answer once within one of the core dimensions ranges between 17 and 20 percent for double explicit DK variant 1 and between 14 and 18 percent for single explicit DK variant 2. This range is narrow, especially considering the differences in question content. Furthermore, item nonresponse for the three themes related to the core dimensions is lower than for the other themes, supporting the assumption that the core dimensions may help organize and express attitudes. Item nonresponse is relatively low for questions related to a major political dimension, which supports hypothesis H3a.

There are only small differences in cumulated item nonresponse for the themes related to the core dimensions of Dutch politics; the last two themes (foreign and current affairs) show substantially higher percentages. This suggests a distinction between the questions about the main dimensions in Dutch politics and the other themes, but further analysis does not fully substantiate this interpretation, since there is an outlier within the Foreign Affairs theme about the UN. The item nonresponse for this outlier is about 20-25 percentage points higher in variant 1 and 2 compared to the other two items. By excluding the UN item, the results are more in line with the first three themes both in terms of total (cumulative) item nonresponse per theme and in respondents only using the DK option once (see Table 5.6). Still, even when the outlier is excluded the number of respondents using a DK option once is 17 percent for the double explicit DK variant and 14 percent for the single explicit DK variant. This is similar to the number of respondents using a DK option once within the socio-economic, ethical and multicultural theme, but with only two survey questions included in the foreign affairs theme (compared to three in the other themes). In other words: the results still support the idea (hypothesis H3b) that item nonresponse for foreign policy issues is relatively high.

The final cluster included in Table 5.6 contains current affairs questions, i.e. questions that were decided upon at the last minute and that are replicated from Dutch pollsters. The two questions within the current affairs theme – on the role of the Queen and surrogate motherhood – are very different in content; the latter could even be considered part of the ethical theme. The item nonresponse as regards surrogate motherhood was, however, high with 33 percent in the double explicit DK variant and 29 percent in the single explicit DK variant. These percentages are much

higher than the item nonresponse of the individual items in the ethical theme, but similar to item nonresponse for the other current affairs item: 35 percent in the double explicit DK variant and 29 percent in the single explicit DK variant for the role of the Queen. Taken together, 54 percent of the double explicit DK respondents and 48 percent of the single explicit DK respondents used the DK option at least once for the two current affairs questions, which is high compared to all other themes.

What does it mean that the DK option is used most frequently for current affairs items? It is difficult to draw definitive conclusions, for instance since the order of the questionnaire and the wording of the questions were not manipulated. It might be the case that if the current affairs questions were posed as the first block of questions, item nonresponse would be lower. Furthermore, the questions within the current affairs theme were relatively short, which requires less cognitive effort but also presents fewer cues for respondents to base their opinions on. This is mere speculation, but some empirical evidence will be given in subsequent chapters. For now, the main point is that item nonresponse does vary according to question content and that question design effects are stronger for certain issues; the main question design effect is, however, whether an explicit DK option is offered, regardless of question content.

5.5.5   Number of Response Categories & Neutral Response Category

The fourth and final set of hypotheses concerns response categories, i.e. the number of response alternatives and the presence of a neutral category. The first hypothesis is that the more substantive response categories are offered, the lower the item nonresponse rate. To examine the relation between the number of categories and item nonresponse, Table 5.7 contains item nonresponse for the individual items in the three variants that included a non-substantive response option, ranked by the number of substantive response categories.

Table 5.7: Item Nonresponse (%) of Individual Items – Ranked to Number of Response Categories

Number of Response categories (Excluding DK)	Question	1. Double Explicit DK	2. Single Explicit DK	3. Single Implicit DK	Significance Paired Comparisons		
					1 / 2	1 / 3	2 /3
7	Self-placement Income Differences	7	5	0	.094	.000	.000
	Self-placement Euthanasia	3	4	1	.755	.000	.000
	Self-placement Foreigners	2	2	0	.627	.000	.000
	Self-placement EU	12	10	1	.202	.000	.000
4	The old-age pension age should be preserved at 65	7	6	0	.247	.000	.000
	The welfare benefits should be lowered in order to stimulate people to work	13	11	1	.044	.000	.000
	Adoption by same-sex couples should be possible	12	11	1	.374	.000	.000
	It is right that women can freeze their ova to be able to have children at a later age	18	17	1	.321	.000	.000
	There are too many people of a non-Dutch nationality living in the Netherlands	16	12	1	.012	.000	.000
	All people living in the Netherlands illegally for a long time should be allowed to stay here	14	13	2	.466	.000	.000
	The Netherlands should spend more money on developmental aid	13	12	1	.379	.000	.000
	The United Nations has too little power	39	35	4	.082	.000	.000
2	The Queen can only communicate governmental policies towards journalists	35	29	2	.001	.000	.000
	Are you pro or con surrogate motherhood?	33	29	2	.029	.000	.000

Item Nonresponse is measured as a percentage of the total number of respondents not responding to a certain survey item by using the DK option (in variant 1 and 2) or skipping the question (in variant 3).

The four self-placement items offered 7 response categories, eight items offered 4 categories and in two cases, (both Current Affairs themes,) only two options were available. Moreover, items with a seven-point scale include a 'Neutral' response category which might 'attract' some DK answers.

Table 5.7 clearly shows a lower level of item nonresponse for most self-placement items compared to the other items within the same theme. For the multicultural theme, for example, the self-placement Foreigners item yields a level of item nonresponse of 2 / 2 / 0 percent [for variant 1 / 2 / 3] which is substantially lower than the 16 and 14 percent of the other two multicultural items in variant 1. This pattern of lower item nonresponse for self-placement items is consistent for all items in the three variants applying a DK option, except for pensions and euthanasia and same-sex adoption in variant 3. These findings suggest that a limited number of response categories may not cover all of the respondents' opinions, and results in a more frequent use of the DK option, although it may be that some items render more DK answers than others because of their substance and regardless of the number of response categories. Nevertheless: in general more respondents use the DK option when the number of response categories is limited and a tentative conclusion is that more response categories result in or at least correlate with less item nonresponse. For more firm conclusions, however, a more extensive analysis is needed.

The final analysis here concerns the availability of a neutral or middle response category. It was expected that at least part of the item nonresponse is substituted by a middle answer. More specifically, a distinction was made between questionnaire variants with and without an explicit non-substantive response option: in the absence of a non-substantive response option (i.e. DK) more use of the midpoint option was expected than in variants offering an explicit DK option. The only items that offered a neutral or midpoint category were the self-placement items. Table 5.7 contains the percentage of respondents using the neutral response category in each of the four variants.

5

Table 5.8: Use of the Neutral Response Category (%)

Question	1. Double Explicit DK	2. Single Explicit DK	3. Single Implicit DK	4. Forced choice
Self-placement Income Differences	20	21	28	30
Self-placement Euthanasia	6	5	8	8
Self-placement Foreigners	18	16	20	19
Self-placement European Unification	18	19	31	27

How often the neutral response category is used varies with question design: for all self-placement items a significant difference is found between some of the subgroups in their use of the neutral option (see Appendix C for the significance tests). The main difference occurs when dichotomizing the subgroups into an Explicit DK category and an Other category with the Single Implicit DK variant and the Forced Choice variant. The midpoint category is used more often in the latter variants. The number of respondents of the Explicit DK variants of the EU item, for instance, using the neutral response category is 20 and 21 percent compared to 27 percent of the Forced Choice and 31 percent of the Single Implicit DK variants. This is consistent with the idea that when a non-substantive response option is not available, the midpoint option is used as a quasi-non-substantive response option. What is also clear is that for some items the midpoint option is used far more often than for others, with the euthanasia self-placement item rendering only between 6 and 8 percent neutral answers as opposed to 21 to 30 percent neutral answers for income differences. This indicates differences in item content, but does not change the fact that all items see a distinct and rather clear midpoint pattern. Whether a DK option is offered explicitly or not is what matters here, which is in line with hypothesis H4b.

5.6   Conclusion

In this chapter the results of a survey experiment on the effects of the way a DK option was offered or not in a questionnaire were presented. The effects were assessed in terms of non-substantive answers, i.e. item nonresponse, and shifts in

the overall distribution of opinions and public opinion. Different types of questions were examined in terms of number of response categories and the subject of the question. The experiment was carried out by the LISS panel.

The DK option effect on item nonresponse is significant and substantive: offering a DK option explicitly as a response category raises the average item nonresponse rate with 13 percent (for the single explicit DK variant as compared to the single implicit DK variant where skipping questions was possible). Secondly, mentioning the DK option in the question itself, in addition to an explicit DK response category, has a small, often insignificant additional effect on item nonresponse. The main difference is between offering a DK option explicitly or implicitly. These findings support hypothesis H1a: offering a non-substantive response option more explicitly results in more item nonresponse.

An analysis of response time reveals more about why respondents use the DK option. Contrary to hypothesis H5, which stated that the DK option would be used as an easy way out and offering the DK option explicitly would result in shorter response times, the response time of explicit DK variants was longer than when a non-substantive response option was offered implicitly or not at all. These results contradict previous findings (see e.g. Greszki et al., 2014, 2015; Malhotra, 2008; Yan & Tourangeau, 2008; Zhang & Conrad, 2013) and suggest that respondents may use the non-substantive response option after careful consideration of the survey question. This is in line with a second strand of thought regarding response time, which argues that 'shorter response times indicate stronger attitudes and measurement of these attitudes are less affected by question order or response order' (Callegaro, Yang, Bhola, Dillman, & Chin, 2009, p. 6). The DK option is used after careful consideration of the survey question.

The second main point is the effect of a DK option on the distribution of opinions, which ultimately reveals the overall picture of public opinion. The question was whether majorities or pluralities would change or disappear when another question design is applied with less or more item nonresponse. Despite some statistically significant effects, the substantive effect of the DK option on the distribution of opinions (with item nonresponse excluded) was small. Furthermore, the majorities only changed when the public was evenly divided on an issue. The story is more nuanced when non-substantive answers are included in the picture of public opinion, with disappearing majorities or even a plurality of respondents giving a non-substantive answer as a result. Overall, the findings support hypothesis H2a: an increase of the level of missing data (as a result of offering a DK option explicitly) does not affect the distribution of opinions.

The finding that offering a DK option explicitly affects item nonresponse is not new (see for example Bishop, 2005; Schuman & Presser, 1979), although it

5

has been substantiated for the Dutch case for the first time. The finding that the picture of public opinion barely changes when a different question design is applied is, however, surprising. More (item) nonresponse arguably leaves more potential for a nonresponse bias, but the data suggest that the missing data are randomly distributed.

Next to the overall difference between the four variants, major differences were found for individual survey items. It was expected that the foreign affairs items would result in relatively high levels of item nonresponse, due to the assumed difficult and abstract nature of these issues. This expectation is partially confirmed, with the UN item resulting in an item nonresponse of 39 and 35 percent for the explicit DK variants. Furthermore, the EU self-placement item did result in the highest level of item nonresponse for self-placement items. These findings support hypothesis H3b: issues related to foreign policy result in more item nonresponse. Hypothesis H3a was also supported: survey questions related to a major political dimension result in less item nonresponse.

In addition to the DK response category, the 'regular' substantive response categories were also examined. Both hypotheses (H4a and H4b) were supported: more substantive response categories rendered less item nonresponse and respondents of variants without a DK option used the neutral or midpoint category more often.

In sum, the main effect of how a DK option is offered is on the level of item nonresponse and not on the overall distribution of opinions. Furthermore, it matters most whether a DK response category is offered explicitly or not, whereas a reminder in the question itself at best only slightly increases the nonresponse.

The Filter Question

6

6.1   Introduction

The focus is in this chapter on the use of filter questions and the effects on the collection and aggregation of substantive and non-substantive answers. According to Krosnick and Presser (2010, p. 264) 'filter questions should be included, to avoid asking respondents questions that do not apply to them'. Whether filter questions improve the quality of survey data is, however, a topic of debate. On the one hand filter questions may discourage respondents from reporting nonattitudes (Krosnick et al., 2002; Zaller, 1992), but on the other hand a filter question could be used as an easy way out to cut the survey short and limit cognitive efforts (e.g. Eckman et al., 2014; Kreuter, McCulloch, Presser, & Tourangeau, 2011). The goal here is to examine the effects of a filter question.

Many authors refer to a 'Don't Know filter', but that does not necessarily entail the use of a separate filter question; an explicit Don't Know option is considered a 'filter' by some as well (e.g. Heerwegh & Loosveldt, 2008; Leigh & Martin, 1987; Loosveldt, Pickery, & Billiet, 2002). This Don't Know option was central in the previous chapter. Here the focus is on an explicit filter question which is posed before the opinion question. On the basis of the literature (e.g. Converse, 1964; Gallup, 1947; Moore, 2008; Schuman & Presser, 1996; Sudman & Bradburn, 1989; Zaller, 1992) we know that using a filter question can influence the level of item nonresponse, but the relationship with the content of the questions is less clear. Furthermore, how a filter question affects the substantive distribution of opinions is not yet fully understood.

The aim of the experiment reported here is to explore and analyze the effect of filter questions on a) the results of the variants of the questionnaire (with and without filter question); b) the outcome of questions in general. The research question is: *How do filter questions influence the outcome of a survey or poll?* An additional question is: how is the level of item nonresponse after the use of a filter question related to the substance of the question?

6.2   Theoretical Reflection

The filter question is an element of question form which together with respondent and interviewer traits forms the three major sources of response effects (Sudman & Bradburn, 1974). The main effect of filter questions is on item nonresponse: the filter results in extra missing data for separate questions. The consensus is that the use of a filter question results in about 20 percent item nonresponse, regardless of question content (Bishop, 2005; Bishop et al., 1983; Eckman et al., 2014; Schuman

& Presser, 1979). Eckman *et al* (2014) find item nonresponse rates (after a filter question) starting at 21 percent and up to 60 (!) percent, depending on the exact filter format and question content.

In web panels respondents sometimes learn how filter questions work when they are used often, and subsequently they use it more often to lower cognitive efforts. This 'panel conditioning' consequently lowers the number of substantive answers, including those of respondents who would otherwise have provided a substantive answer (Eckman et al., 2014). The presence of an interviewer (in telephone and face-to-face surveys) also affects the 'triggering of filter questions' (Kosyakova et al., 2015, p. 418) and as a result the sample size and amount of valid data gathered with the survey (Josten & Trappmann, 2016; Kosyakova et al., 2015). 'Triggering' means that a substantive opinion question is 'triggered' or accessed when the respondent answers 'yes' to a filter question. The 'trigger rate' is 'the proportion of respondents giving answers that trigger follow-up questions' (Eckman et al., 2014, p. 722), which corresponds to the item response rate for individual opinion questions (see also Kosyakova, Skopek, & Eckman, 2015). So the substantive opinion question is only triggered when the preceding filter question is answered with 'yes'.

Using filter questions may affect both the number of non-substantive answers and the distribution of opinions. Previous studies suggest that 'the filtered distribution of opinions sometimes differs from the unfiltered (standard) distribution and sometimes does not' (McClendon & Alwin, 1993, p. 439). Schuman and Presser (Schuman & Presser, 1996, p. 127) concluded that 'filtering can on occasion significantly alter the division of substantive opinion, but that it typically does not'. And Knaüper (1998) finds, for example, that the number of reported crimes is affected by the wording of the filter question. If skipping a question is a random process among respondents, the overall distribution of opinions is not affected by differences in the offered non-substantive response options. If the missing data are, however, related to a refusal to reveal certain information or opinions, a systematic nonresponse bias could arise (De Leeuw et al., 2003, p. 159). This 'bias only occurs if people who do not answer are different from those who do', either because they are different in terms of individual characteristics or because they hold different opinions (see also Stöss, 2009; Weisberg, 2008, p. 225). This difference in the distribution of opinions or 'substantive proportions' (Schuman & Presser, 1996, p. 115) will be analyzed below.

The aim of the experiment reported here is to contribute to the literature by examining how the use of filter questions affects the outcome of a survey. By looking at the item nonresponse rate and the substantive distribution of opinions, both the loss of (valuable) information and the (non)random distribution are explored. Furthermore, not all filter questions render the same results. Filter questions can

be treated as variables: the strength of the filter question may vary between 'Do you have an opinion on this or not?' as arguably the weakest version with the least item nonresponse, and 'Have you already heard or read enough about it to have an opinion?' as the strongest filter (Bishop, 2005, pp. 22-23; Bishop et al., 1983, pp. 530-535).

6.3 Hypotheses

The hypotheses are summarized in Table 6.1. See chapter 3 for a more extensive discussion.

Table 6.1: Hypotheses

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

6.4   Data and Methods

In order to study the effects of filter questions, a survey experiment was executed. The experiment was an internet survey experiment filled in by the respondents of the *EenVandaag Opiniepanel*. *EenVandaag* is a Dutch daily news program on a public broadcasting channel having its own online panel. The respondents are at least 18 years old and registered themselves to participate in surveys about current affairs[25]. The full panel was targeted, with random selection of seven distinct groups for the various treatments. This between-subjects-design made comparison between subgroups of respondents possible; the random assignment resulted in subgroups which were similar on key demographic characteristics (see Table 6.1 below).

The general instruction in the questionnaire read that the respondents were asked to give their opinion on issues in the poll and that the results would be used for scientific research; usually the results are published in the news program[26]. The question wording was identical to the original question which was replicated, including the introduction and choice of response alternatives. Five general themes were addressed and each theme included three questions of which at least one question came from existing research – see section 4.4 for more information about issue selection.

The experiment was carried out in October and November 2011. Five variants of the questionnaire (1A, 1B, 2A, 2B and 3A) were distributed on Friday 28 October 2011, the other two on Monday 31 October 2011 (due to a software problem). After a reminder on Friday the 5th of November, the survey closed on Wednesday November 9th. In total, 29.333 respondents – 64 percent of the EenVandaag Opiniepanel – participated in the experiment.

---

25   See http://opiniepanel.eenvandaag.nl/uitleg for more information about the EenVandaag Opiniepanel.

26   The EenVandaag Opiniepanel has been used before as a source of data for (political) scientists. The results were, for instance, used by Kranenburg and Weimar (2008, p. 500) in their mini review of 'surveys that have been performed to study public opinion on the idea of introducing incentives for living kidney donation' and by Bovens and Wille (2008) in their essay about political trust, who used data that were not specifically collected with the purpose of scientific analysis. More rigorous was the analysis of electoral volatility with data from 2006 till 2010 from EenVandaag (Van der Meer, Lubbe, Van Elsas, Elff, & Van der Brug, 2012; van der Meer, van Elsas, Lubbe, & van der Brug, 2012). These analyses, however, used the original data, whereas the data in this survey experiment were specifically gathered with the purpose of using it for (scientific) research. An example in which the EenVandaag Opiniepanel was used to gather data for research is Van Holsteyn's (van Holsteyn & Cupido, 2013a, 2013b) analysis of political cartoons, to see whether respondents understand the cartoonist's message.

Table 6.2: Descriptive Statistics Subgroups EenVandaag Opiniepanel

		1A	1B	2A	2B	3A	3B	4	Total
Gender*	Male	71%	73%	69%	70%	69%	67%	65%	69%
	Female	29%	27%	31%	30%	31%	33%	35%	31%
Age (years)*	Mean	58.6	57.8	59.2	58.7	56.1	54.1	53.6	56.9
	SD	31.8	32.5	53.1	32.9	14.1	35.8	14.6	33.2
Education*	Basisonderwijs	1%	1%	1%	1%	1%	1%	1%	1%
	VMBO	16%	14%	16%	16%	15%	15%	16%	15%
	Havo/VWO	9%	9%	10%	10%	10%	10%	9%	10%
	MBO	22%	21%	20%	20%	21%	21%	23%	21%
	HBO	36%	37%	37%	36%	37%	36%	36%	36%
	WO	16%	18%	17%	17%	17%	17%	14%	17%
Income*	Below modal	20%	21%	22%	22%	23%	25%	26%	23%
	Modal	20%	20%	20%	21%	20%	21%	22%	21%
	More than modal	60%	59%	58%	57%	56%	54%	51%	56%
N		4412	4329	4327	4283	3931	3591	4460	29333

*Significant difference between the seven subgroups at the .001 level
Gender: percentage of the (sub)group that is male or female.
Age: average age in years, with respondents being 18 years or older.
Education: highest level of education completed, recoded into CBS categories.
Income: recoded into three categories.

6

Some descriptive characteristics of the seven subgroups and the sample as a whole are displayed in Table 6.2. Comparing subgroups, all variables show statistically significant differences between subgroups at the .001 level. Looking at the substantive differences, however, these are very small; the statistical difference can be explained at least partially by the large N. The similarity of the subgroups is important in a between-subjects-design to determine the effect of the manipulation. The sample as a whole is quite unrepresentative and generalization to the population is not possible. See for example the gender and education of respondents: 69 percent of the total sample is male, compared to 49 percent of the population, and 49 percent of the respondents finished HBO or WO, compared to 32 percent in the population. The use of a nonprobability online panel limits the external validity of the findings, but the between-subjects-design does provide internally valid findings with the possibility to explore causal mechanisms.

Table 6.3: Response Rates

Variant	Number of Respondents	Response Rate – Unit	Number of Items in Questionnaire		Average Item Nonresponse
			Filter	Opinion	
1A – Strong filter, explicit DK	4412	67%	15	17	11% (6% DK)
1B – Strong filter, implicit DK	4329	66%	15	17	7% (1% DK)
2A – Weak filter, explicit DK	4327	66%	15	17	9% (5% DK)
2B – Weak filter, implicit DK	4283	65%	15	17	6% (1% DK)
3A – No filter, explicit DK	3931	60%		17	8% DK
3B – No filter, implicit DK	3591	55%		17	1% DK
4 – Forced choice	4460	68%		17	0% DK
Total	29333	64%			

Average Item Nonresponse combines the two categories for item nonresponse, i.e. 'no' to a filter question and the DK option. The figures between parentheses indicate the number of respondents using the second category (the DK option).

Table 6.3 shows the response rates of the respondents of the *EenVandaag Opiniepanel* participating in the experiment. For each variant, 6,570 people were approached. The response only includes respondents who completed the survey; break-offs are not registered in the same dataset. Usually about 60 to 70 percent of the respondents of the EenVandaag panel participate in surveys (Opiniepanel, 2015); the response rate for this survey experiment is about average with response rates for the subsamples ranging from 55 to 68 percent. The software problem that delayed the start of the survey of variant 3B and 4 did not seem to affect the response rate; these variants resulted in the lowest (55) and highest (68) response rates.

The manipulated variable in this experiment is the use of filter questions (and other non-substantive response options). Filter questions were asked prior to the substantive question and were intended to distinguish respondents without an opinion or knowledge – depending on the wording and purpose of the filter question – from those who did have an opinion or knowledge about the subject. Individuals indicating they did not have the relevant opinion or knowledge were filtered out and routed to a next question. Overall, the design options resulted in

an experiment with seven subgroups: two strongly worded filter question variants (with and without DK option), two weaker worded filter question variants (with and without DK option), two variants without filter question (with and without DK option) and a forced choice variant – see Table 6.4 (see Appendix A for the complete questionnaires). The distinction between 'weaker' and 'strongly' worded filter questions is based on Bishop *et al* (1983) and tested in the analysis below.

Table 6.4: Experimental Conditions and Safety Nets for Item Nonresponse

	Explicit DK	Implicit DK	Forced choice
Strong filter	1A: 'no' to filter question or say DK	1B: 'no' to filter question or skip question	
Weak filter	2A: 'no' to filter question or say DK	2B: 'no' to filter question or skip question	
No filter	3A: say DK	3B: skip question	4: no safety net, answer obligatory

6.5   Results

In this section, the results are presented of the survey experiment in which both the use of the filter question and the way a DK option was offered were manipulated. It should be noted beforehand that there is little attention for the statistical significance of differences between subgroups: the N is very large, which leads to statistically significant differences easily. Therefore, the focus is on substantive differences between subgroups.

6.5.1   Item Nonresponse
The analysis of item nonresponse focuses on three key aspects: 1) the *use of filter questions* with a strongly worded, weaker worded and absent filter question; 2) the way in which a *DK option* is offered and 3) the *differences between issues* in level of nonresponse and susceptibility to design choices. This section discusses the first two aspects by looking at the effect of offering a filter question and comparing the resulting item nonresponse to the item nonresponse rendered by the (explicit) use of a DK option.

How does the use of a filter question affect item nonresponse? Two hypotheses were developed: H1a expects more item nonresponse for more explicit non-substantive

response options and H1b compares the item nonresponse of a filter question with the item nonresponse of an explicit DK option. Based on H1b, one would expect the variants with a filter question to render most item nonresponse. Questionnaire variants with either weak or strong filter questions are therefore expected to have a higher average item nonresponse rate than variants without filter questions. According to H1a it is expected that variants with a 'strong' filter question render more item nonresponse than variants with a 'weak' filter question. Finally - and comparing pairs of the same filter variant, e.g. the strong filter variants - according to H1a, variants with an explicit DK category result in more item nonresponse than variants with an implicit DK option and the possibility to skip questions.
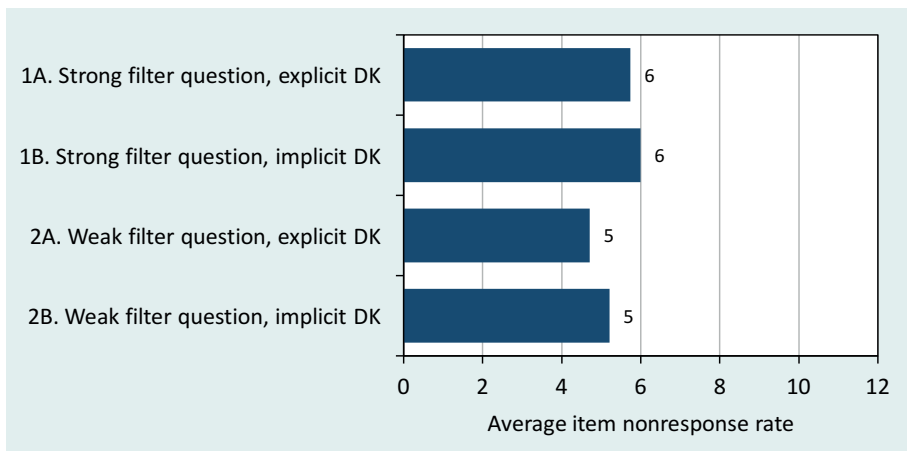
Figure 6.1: Average Item Nonresponse (%) Resulting from Filter Questions
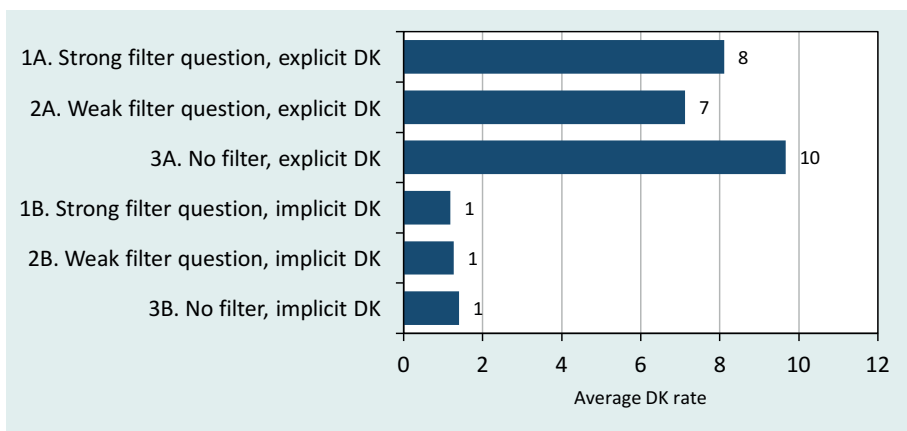
Figure 6.2: Average Item Nonresponse (%) Resulting from DK Option

Figure 6.1 and Figure 6.2 show the item nonresponse resulting from the use of a filter question (Figure 6.1) or a DK option (Figure 6.2). The item nonresponse only shows the isolated effect of a single non-substantive response option. Offering a filter question results on average in about 5 to 6 percent item nonresponse, depending on the exact wording of the filter question[27]. This is a small effect; the numbers are substantially lower than the 20 percent item nonresponse reported in previous studies.

Even more surprising is the comparison of the filter question with other types of non-substantive response options, i.e. the explicit DK option and the option to skip questions. Contrary to hypothesis H1a, the more explicit non-substantive response option, i.e. the filter question, did not have a higher item nonresponse rate than less explicit (DK) non-substantive response options; the item nonresponse rate of the explicit DK option was 7 to 9 percent (see Figure 6.2) as compared to 5 to 6 percent for the more explicit filter question (see Figure 6.1). Even though the filter question constituted the most explicit reminder of the possibility to express a nonresponse, it was used less often than the less explicit non-substantive DK response option. Only the implicit option to skip a question rendered less average item nonresponse: 1 percent (see Figure 6.2). So the least explicit non-substantive response option, i.e. the implicit DK option, generates the least item nonresponse; the most explicit non-substantive response option, i.e. the filter question, does not show the highest item nonresponse rate. *Hypotheses H1a and H1b are not supported.*

The only qualification of the conclusion that more explicit non-substantive response options do not result in more item nonresponse, results from the comparison within type of non-substantive response options. Rather than comparing filter questions with DK options, the two types ('strong' and 'weak') of filter questions can be compared. The expectation that a 'strong' filter– 'Have you already heard or read enough about it to have an opinion' – results in more item nonresponse rate than a 'weak' filter– 'Do you have an opinion on this or not' – is partially confirmed: the item nonresponse of strong filter question variants (1A and 1B) is higher than in the weak variants (2A and 2B). The effect of the wording of a filter question is, however, very small with only 1 percentage point between the 'strong' and 'weak' filter question variants (see Figure 6.1). The same within-type-comparison of the two types of DK options, i.e. explicit or implicit, supports the

---

27 The item nonresponse rates of the filter questions were calculated by adding the item nonresponse rate of individual filter questions and dividing this sum by 15. The item nonresponse rates of the DK option were calculated by adding the item nonresponse (resulting from explicit DK answers and skipping questions) of individual opinion questions and dividing this sum by 17. The difference in the number of items, 15 versus 17, stems from the fact that two filter questions each preceded two opinion questions.

findings in the previous chapter. The explicit DK option results in an average item nonresponse rate of 7 to 9 percent; the implicit DK option results in about 1 percent item nonresponse (see Figure 6.2).

The discussion so far addressed the isolated effect of one non-substantive response option. In reality, the four filter question variants offered two non-substantive response options to the subgroup respondents: a filter question plus a DK option. It is possible that respondents learned throughout the survey that another non-substantive response option was available to them even after they answered the filter question with 'yes'. Following this line of thought, the item nonresponse rate of the filter variants would have been considerably higher if no other nonresponse (DK) option was offered to them other than the filter question.

This point can be addressed in two ways. The first is that the implicit DK variants resulted in an extremely low item nonresponse rate of 1 percent, which is an indication that many respondents probably did not notice this non-substantive response option[28]. Still, the item nonresponse rate of variants offering a filter question followed by an implicit DK option was less than 1 percentage point higher than in filter variants with an explicit DK option (see Figure 6.1). Secondly, while it is true that filter variants with an explicit DK option result in the highest *total* item nonresponse of 9 and 11 percent (see Figure 6.3), the no filter variant with an explicit DK option ranks third with 8 percent. Even though the latter variant only includes one non-substantive response option, it outranks two filter variants in terms of item nonresponse. Using a filter question does not result in more item nonresponse than including a DK option as an explicit response category.

---

28   The combination of a filter question followed by a forced choice opinion question was not included in this experiment.
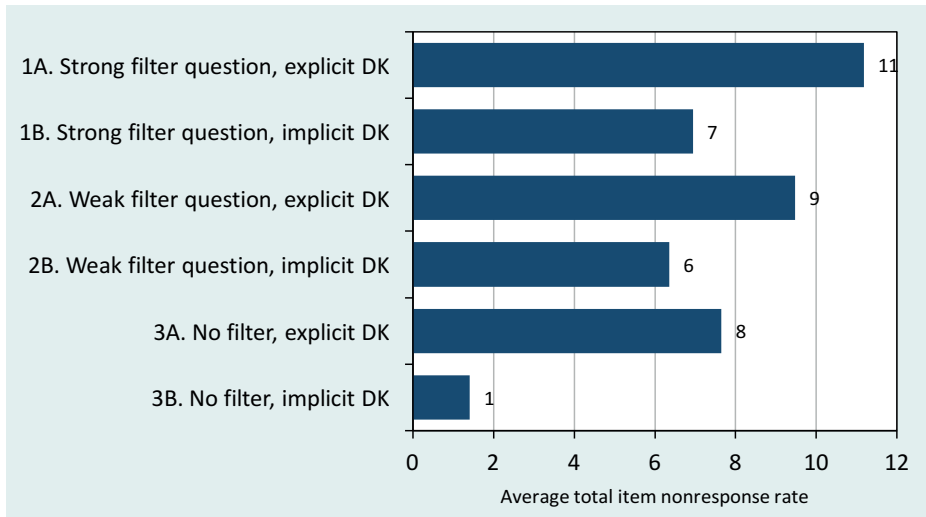
Figure 6.3: Average Total Item Nonresponse (%)[29]

To examine in more detail how question design affects item nonresponse and see both the effect of a filter question on item nonresponse and whether the use of a filter question results in more item nonresponse than offering an explicit DK option, the item nonresponse for all separate items is shown in Table 6.5 (item nonresponse rendered by filter questions) and Table 6.6 (item nonresponse rendered by all non-substantive response options).

---

29   The average total item nonresponse is computed by adding up the level of item nonresponse of each issue in one variant of the questionnaire, both as a DK answer and as a 'no' to the preceding filter question, and dividing this sum by 17. Since no DK option was offered in the forced choice variant, it is excluded from the analysis.

It should be noted that the average total item nonresponse (in Figure 6.3) does not equal the added up nonresponse rates in Figure 6.1 (filter questions) and Figure 6.2 (the DK option). There are two reasons for this: 1) the average DK rate (in Figure 6.2) is based on the respondents answering the substantive opinion question, which excludes the people saying 'no' to the previous filter question; and 2) the average item nonresponse rate resulting from filter questions and a DK option is calculated differently, because of the difference in number of items. .The average total item nonresponse presented here shows how many respondents did not answer the substantive opinion question, either because they used the filter or the DK option.

Table 6.5: Filtered Out Item Nonresponse (%) of Individual Items[30]

Question	1A Strong Filter, Explicit DK	1B Strong Filter, Implicit DK	2A Weak Filter, Explicit DK	2B Weak Filter, Implicit DK
Self-placement Income Differences	19	17	8	8
Welfare benefits should be lowered in order to stimulate people to work	17	15	8	7
Self-placement Euthanasia	2	3	2	2
Adoption by same-sex couples should be possible	8	9	7	7
Self-placement Foreigners	2	2	1	2
There are too many people of a non-Dutch nationality living in the Netherlands	2	2	2	2
Self-placement European Unification	4	5	3	4
The Netherlands should spend more money on development aid	4	4	2	3
What do you think should happen to the mortgage interest deduction?	6	6	6	7
What do you think is the best solution for the impending deficits of pension funds?	6	8	6	9
The Netherlands should in the next year quit the euro and go back to the gulden	4	4	3	4
Do you think that the King or Queen should have political influence, or should s/he restrict herself to ceremonial roles?	2	3	4	4
Do you think that Maxima's father can or cannot be present at the coronation?	6	7	10	10
[Filter question Libya]	3	4	5	7
[Filter question current cabinet]	2	2	2	3
N	4412	4329	4327	4283

30 The last two filter questions (about Libya and the current cabinet) were both followed by two substantive opinion questions. The number of items in this table is therefore lower than in Table 6.6. Another difference between Table 6.5 and Table 6.6 is that the item nonresponse rate in the former only results from the filter question (while ignoring the potential use of a DK option in the substantive opinion question), whereas the total item nonresponse in Table 6.6 combines both non-substantive response options by showing how many respondents did not answer the opinion question – either because of saying 'no' to the filter question or because of the use of a DK option.

Table 6.5 shows how many respondents said 'no' to the filter question, constituting item nonresponse. The filter question could either be 'do you have an opinion on this or not' or 'have you already heard or read enough about it to have an opinion'. It was expected that the latter stronger worded question would result in more item nonresponse, since it assumes more than just having an opinion.

There is an empirical indication that a stronger filter question results in more item nonresponse. On average, 6 percent of the respondents of the strong filter (1A and 1B) said 'no' to the filter question, which is slightly more than the average of 5 percent in weaker variants 2A and 2B. This minor difference of 1 percentage point, however, points to a very small effect of the strength of a filter at best. Moreover, some items show the opposite result, e.g. the questions about the best solution for the deficits of pension funds and Maxima's father[31]. In the latter case the weak filter resulted in 10 percent of the respondents saying 'no' compared to 6 and 7 for the stronger 'Have you heard or read enough?'.

Another noticeable aspect is how often the filter is used over the course of the survey. More respondents in implicit DK variants 1B and 2B say 'no' to the filter question towards the end of the survey than respondents of the explicit DK variants. One explanation is that respondents learn during the survey that the filter question is their only non-substantive response option - or at least that is what they think, because no explicit DK option is included. Rather than answering the opinion question, they may decide not to look at this question and say 'no' to the filter question more often compared to respondents who do have an explicit DK option as a second way out. Also noticeable is that respondents say 'no' to filter questions more often at the start of the survey, especially in the strong filter variants 1A and 1B. The percentages diminish after two opinion questions, possibly because respondents find out how the question format works.

The presence of both non-substantive response options, i.e. the filter question and the (explicit) DK option (see Table 6.6), shows that the expectation that a filter question raises the item nonresponse to about 20 to 25 percent must be qualified. The results here are inconsistent with the at least part of the literature about the effect of filter questions (Bishop et al., 1983; Schuman & Presser, 1979). A filter question in this study results in about 10 percent item nonresponse in both the stronger and weaker variant.

---

31   The survey question fits within the public and political debate about whether the father of (then future Queen) Máxima should be allowed to attend the accession to the throne of Máxima's husband Willem-Alexander. Jorge Zorreguieta was not allowed to be present at their wedding in 2001 because of his involvement in the military junta in Argentina; in 2011 the discussion centered around his future attendance at the throne accession (AD, 2011; de Volkskrant, 2011). According to Maurice de Hond, whose survey question is replicated in this experiment, 56 percent of the Dutch thought that Jorge Zorreguieta should be allowed to attend – 38 percent disagreed (het Parool, 2011).

Table 6.6: Total Item Nonresponse (%) of Individual Items

Question	1A Strong Filter, Explicit DK	1B Strong Filter, Implicit DK	2A Weak Filter, Explicit DK	2B Weak Filter, Implicit DK	3A No Filter, Explicit DK	3B No Filter, Implicit DK
Self-placement Income Differences	20	18	8	10	1	1
Welfare benefits should be lowered in order to stimulate people to work	19	16	10	8	4	1
Self-placement Euthanasia	3	4	2	3	1	0
Adoption by same-sex couples should be possible	10	9	8	8	5	1
Self-placement Foreigners	2	3	1	2	0	0
There are too many people of a non-Dutch nationality living in the Netherlands	5	3	4	3	4	1
Self-placement European Unification	4	5	4	5	2	0
The Netherlands should spend more money on development aid	7	5	5	4	5	1
What do you think should happen to the mortgage interest deduction?	6	6	6	7	3	1
What do you think is the best solution for the impending deficits of pension funds?	32	12	27	13	30	6
The Netherlands should in the next year quit the euro and go back to the gulden	10	6	8	5	8	2
Do you think that the King or Queen should have political influence, or should s/he restrict herself to ceremonial roles?	4	4	5	5	3	1
Do you think that Maxima's father can or cannot be present at the coronation?	10	8	12	11	8	2
I think that Libya will, in time, become a normal democratic country	23	6	23	8	23	2
The Netherlands should be actively involved to help Libya establish a democratic regime	10	6	12	9	9	2

Table 6.6: Continued

Question	1A Strong Filter, Explicit DK	1B Strong Filter, Implicit DK	2A Weak Filter, Explicit DK	2B Weak Filter, Implicit DK	3A No Filter, Explicit DK	3B No Filter, Implicit DK
How long do you think this cabinet will remain in office?	7	3	7	3	6	1
Suppose that next year another 5 billion euro in budget cuts have to be made. Do you think the PVV will stop supporting the government?	18	4	19	4	18	2
N	4412	4329	4327	4283	3931	3591

Total Item Nonresponse is measured as a percentage of the total number of respondents not responding to a certain opinion question by saying 'no' to the preceding filter question (in variant 1A, 1B, 2A and 2B), by using the explicit DK option (in variant 1A, 2A and 3A) or by skipping the question (in implicit DK variants 1B, 2B and 3B).

How does the filter question compare to the explicit DK option? Do filter questions affect item nonresponse more than an explicit DK option? A high level of item nonresponse is observed for the questions about 'pension funds', income and 'welfare benefits' (when using filter questions), one question about Libya and one question about the current government. A low level of item nonresponse, regardless of which variant of the questionnaire was used, can be seen for questions about 'immigrants', 'nationality' and the Queen. (Potential differences in question content are discussed below.)

The most remarkable finding, however, is that it does not seem to matter whether a filter question was used in combination with a DK option or only an explicit DK option, since these options result in comparable levels of total item nonresponse[32]. This is somewhat counterintuitive, since double explicit DK filter variants 1A and 2A apply a double stimulus to give no opinion, whereas in variant 3A only one stimulus was used: the explicit DK option. Still, even though the filter questions have less impact on item nonresponse than expected and suggested by

---

32   An analysis of how often the DK option was used, regardless of whether a filter question was posed before the opinion question, can be found in . The general trend is that the explicit DK option is used more often when no other nonresponse option is available. Furthermore, the implicit DK option was seldom used.

the standard literature, the variants with a filter question do generally result in more missing data than variants without a filter question. This supports H1b; the previous analysis of the use of filter questions, however, indicates that the higher item nonresponse rate of filter question variants is the result of the use of *two* non-substantive response options, i.e. both a filter question and a DK option. A comparison of variants with a filter question and without an explicit DK option with a variant with only an explicit DK option suggests that only offering a filter question does not result in more item nonresponse than only offering an explicit DK option; in the implicit DK filter variants, the filter question effect is not stronger than the effect of offering a DK option explicitly. Hypothesis H1b is rejected.

6.5.2   Distribution of Opinions – Towards Public Opinion

How about the outcomes in terms of majorities and pluralities and the overall distribution of opinions? In this section, item nonresponse is treated as missing data and excluded from the findings that are presented. Two pictures of public opinion will be presented: with and without item nonresponse. The main question is: how is the overall distribution of opinions affected by using filter questions? The hypotheses tested are H2a and H2b, which refer to the data (not) missing at random to see whether the resulting picture of public opinion looks different when no filter question (or explicit DK option) is used.

Figure 6.4 to Figure 6.7 present the distributions of opinions of four survey items. Item nonresponse resulting from the use of a non-substantive response option, i.e. a filter question or DK option, is excluded as missing data. The distributions of opinions of all items can be found in Appendix C.
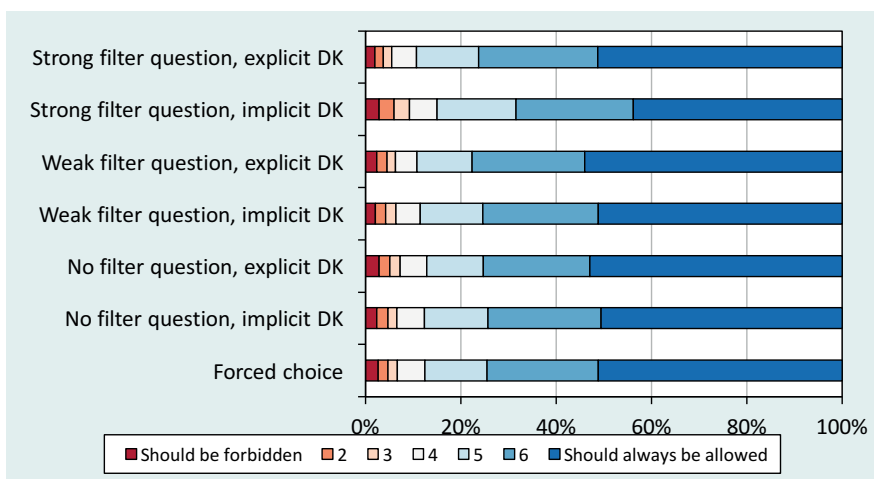
Figure 6.4: Distribution of Opinions Self-Placement Euthanasia
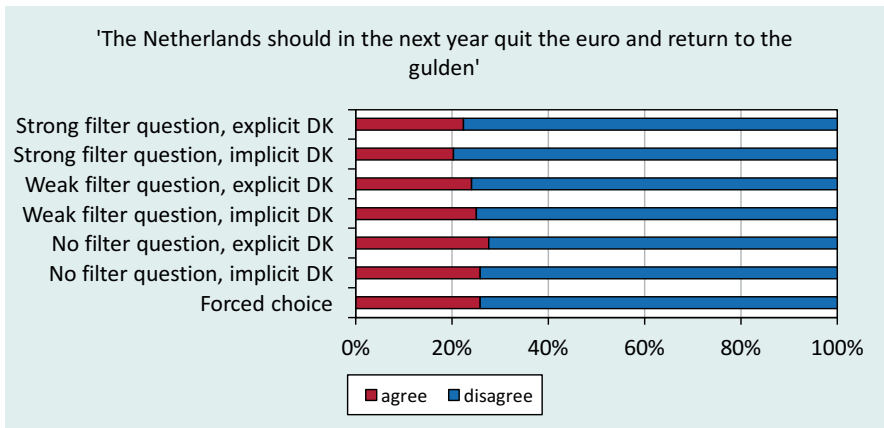
Figure 6.5: Distribution of Opinions Gulden
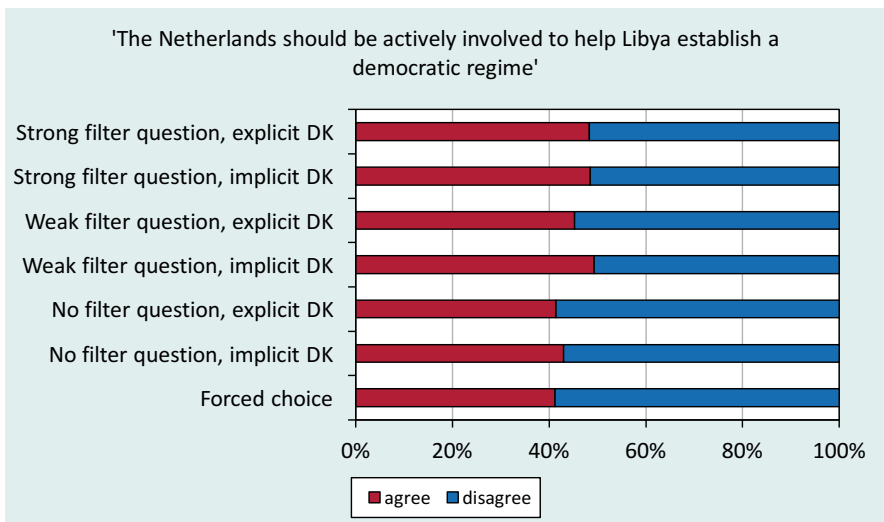
Figure 6.6: Distribution of Opinions Libya 2
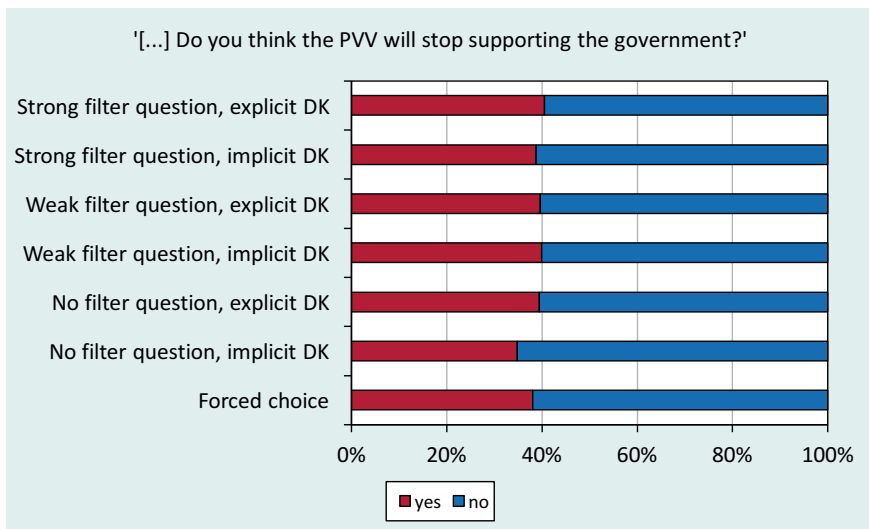
Figure 6.7: Distribution of Opinions PVV

How does the use of a filter question and a DK option affect survey results? Do the distributions of opinions differ for different options? One would expect to see differences when item nonresponse (excluded as missing data) is not randomly distributed. Furthermore, a higher item nonresponse rate resulting from the filter question and an explicit DK option increases the potential for nonresponse bias; items with more item nonresponse should result in a different distribution of opinions.

The results, however, seem fairly robust to question design. Despite the manipulation of non-substantive response options and the consequent variation of item nonresponse, the overall distributions of opinions show hardly any substantive differences. In items with two response categories the same majority preference is held in all seven variants. For example, a majority of 72 to 80 percent of the respondents in all variants thinks the Gulden should not be reintroduced (see Figure 6.5) and 60 to 65 percent disagrees with the statement that the Dutch extreme right-wing political party PVV will stop supporting the government if additional budget cuts were needed (in 2011, see Figure 6.7)[33], regardless of question design. Even in the self-placement items offering seven substantive response categories, the differences between overall distributions are small – see for example euthanasia

---

33   It could be argued that this statement does not measure public opinion, but respondents' expectations. These expectations do, however, give an indication of the respondents' views of the PVV's support of the government.

in Figure 6.4. This is confirmed by the comparison of means of the self-placement items (see Table C.8 in Appendix C).

More item nonresponse results from using a filter question, but the distribution of opinions is robust and pluralities or majorities do not change. These findings support hypothesis H2a: the data seem to be missing at random and no nonresponse bias occurs. It does not matter which design is used to see what the preferred policy option is. The only difference is the size of the plurality or majority. To give an example: a majority in all seven subgroups disagrees that the Netherlands should be actively involved in helping Libya to establish a democratic regime, but this majority ranges between 51 and 59 percent (see Figure 6.6). These differences may however be consequential. If the public is divided (51 versus 49 percent), a politician may not see this result as a guideline to decide on the policy s/he wants to pursue; a majority of almost 60 percent, however, may be a clearer and stronger signal of what the public wants. So even though the majorities do not change, the size of the majority may be valuable information.

The preliminary conclusion is that the filter effect on the overall outcome is extremely limited or even nonexistent. However, if item nonresponse is *not* excluded as missing data and regarded as valuable information about how many citizens do not have an opinion, the picture of public opinion looks different, at least for some items. Figure 6.8 to Figure 6.10 display for three survey items the two alternative pictures of public opinion: one excluding and one including item nonresponse resulting from non-substantive response options as part of the outcome. All other distributions of opinions can be found in Appendix C.

6
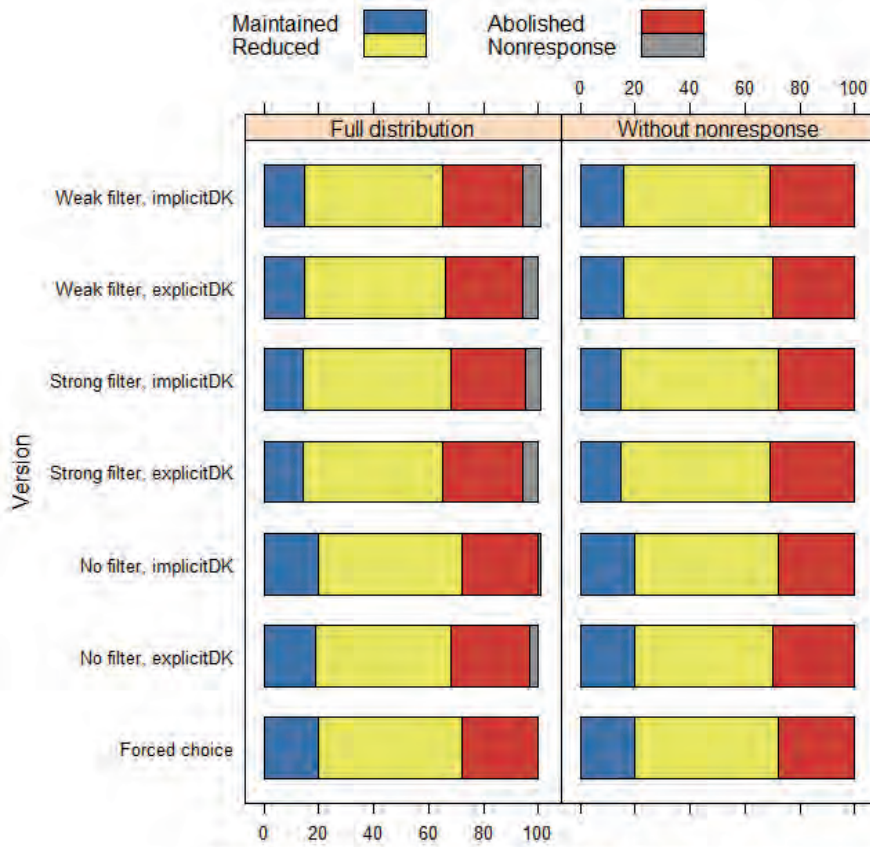
Figure 6.8: Distribution (%) of Opinions **Mortgage Interest Deduction** With and Without Item Non-response
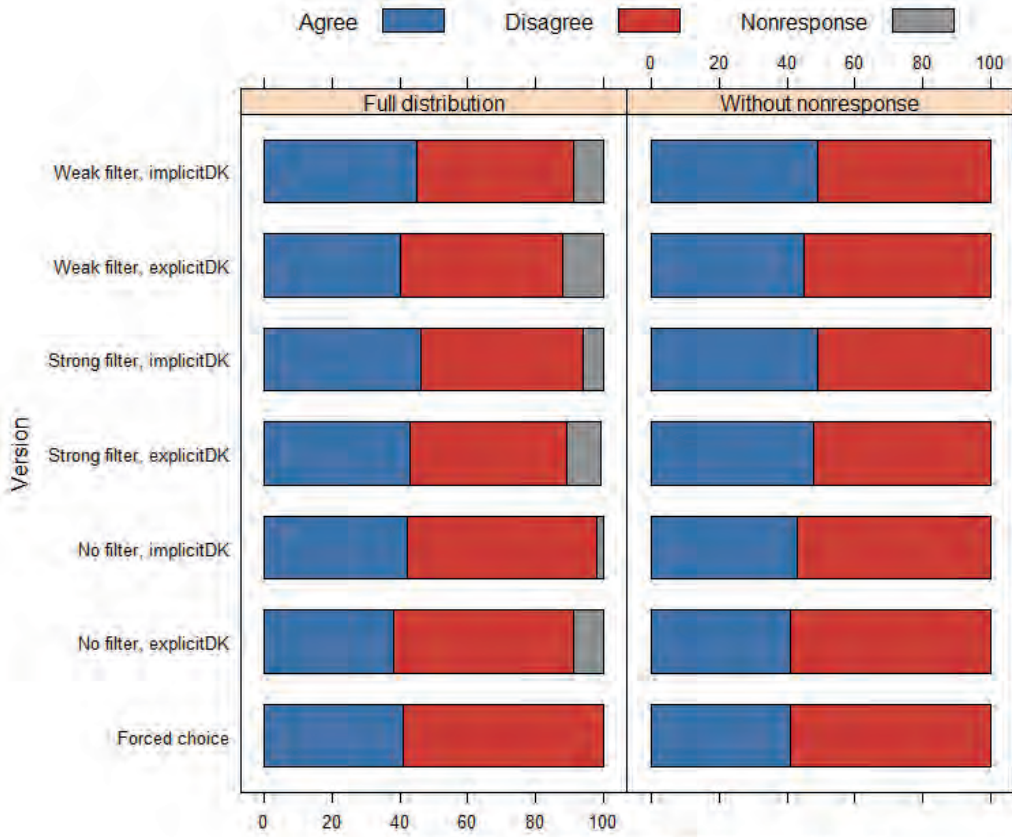
*Figure 6.9: Distribution (%) of Opinions **Involvement Netherlands in Libya** With and Without Item Nonresponse*
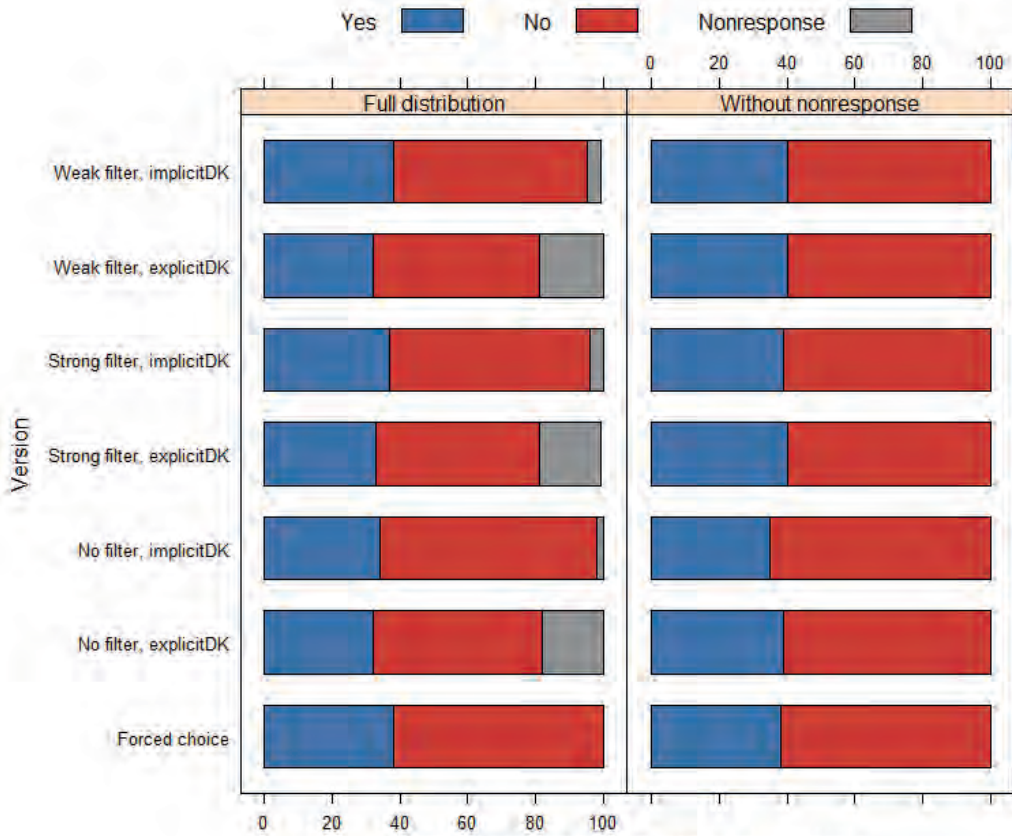
Figure 6.10: Distribution (%) of Opinions **PVV Support Government** *With and Without Item Nonresponse*

If item nonresponse is not taken into account, the distribution of opinions and overall picture of public opinion is not significantly affected by differences in the way non-substantive response options were offered. What happens if item nonresponse is included as relevant part of the outcome? Three effects could occur: 1) the preference of the largest group of respondents, either a plurality or majority, is not the same in all seven variants; 2) the preference of the largest group of respondents is the same in all seven variants, but it is not supported any longer by a majority in all variants; and 3) in some subgroups the response category used by a plurality of the respondents is the non-substantive response option. To illustrate these potential effects, they are applied to a hypothetical example in which respondents are asked whether they agree or disagree with the statement that more tax money should be spend on the military. If the first effect occurs, in some variants of the questionnaire the largest group would disagree with this statement whereas in other variants a plurality or majority agrees. Depending on the variant (and the way a non-substantive response option is offered), a different picture of public opinion results and a different conclusion as to whether the public wants to see more tax money spent on the military or not. If the second effect occurs, all seven variants suggest the same direction – e.g. the largest group answers that more tax money should go towards the military – but this preference is not supported by a majority in all variants. And if the third effect occurs, a plurality or majority of the respondents does not substantively answer the survey question.

The first effect was already disproven above, with no changing majorities or pluralities for any item. The second effect, i.e. the same preference but varying degrees of support when item nonresponse is included, is present in several cases: mortgage interest deduction, Maxima's father's presence at the crowning, the two questions about Libya and the PVV's support of the government. In addition to these five items, two items (non-Dutch nationality and pension funds) had a variant where the preferred policy was supported by only 50 percent of the respondents – not a majority. For these survey questions, a variant with a non-substantive response option (either a filter question and/or an explicit DK option) results in a plurality rather than a majority supporting a particular position. For example: most respondents in all variants answered that mortgage interest deduction should be reduced over time, but this position was not supported by a majority in all variants (see Figure 6.8). 49 percent of the respondents of the explicit DK variant 3A wanted it to be reduced over time, compared to a majority of 54 percent in the strong filter, implicit DK variant 1B. Likewise, a plurality of 46 to 48 percent of the filter variant respondents (1A, 1B, 2A and 2B) disagreed with the statement that the Netherlands should be actively involved in helping Libya set up a democratic regime, compared to a majority of 53 to 59 percent in the other non-filter variants

(see Figure 6.6). A plurality of 48 to 50 percent of explicit DK variants 1A, 2A and 3A said the PVV will not stop supporting the government if additional budget cuts are needed, opposed to a majority of 57 to 64 percent in the implicit DK and forced choice variants (see Figure 6.7). The picture of public opinion does look different when item nonresponse is taken into account; non-substantive response options can change a majority into a plurality.

The third effect of non-substantive response options on the distribution of opinions can be that a plurality uses a non-substantive response option by saying 'no' to the filter question, using the explicit DK option or skipping an opinion question in the implicit DK variants. This effect does not occur. Item nonresponse is not the plurality in any of the items for any variant, although it is the second largest response category for the question about pension funds (see Appendix C).

Hypothesis H2b is partially supported: an increase of missing data results in different overall distributions of opinions. Non-substantive response options affect the picture of public opinion; majorities disappear or become less pronounced. The overall conclusions about the effect of a filter question are twofold. Firstly, the substantive effect on the distribution of opinions is small or non-existent when item nonresponse is excluded as missing data. Secondly, the effect on the distribution of opinions including item nonresponse is more substantial with majorities becoming pluralities and a more divided overall picture of public opinion. The main difference relates to the treatment of item nonresponse as either missing data, or as valuable information about what part of the public does not have an opinion.

6.5.3 Question Design in Practice

To determine the extent to which question design affects the actual outcome of a survey question, this section is devoted to an in-depth investigation of one issue, i.e. the PVV's support of the government (in 2011). The first step will be to compare the 'real' poll result to the result of the survey experiment. The point of reference is a question without a filter question and with an explicit DK option. The comparison is two-fold: both for all respondents and for the PVV voters[34], since this distinction between groups of voters was also made in the actual reporting of the survey results. Second, attention is devoted to the distributions of the seven variants, to see whether they differ from each other and find out whether question design affects the resulting picture of public opinion.

---

34   The PVV voters could be selected on the basis of their reported voting behavior in the national elections of 2010. This is also how Peil.nl selects the voters of a particular party.

Table 6.7: Real Outcome versus Outcome Variant 3A – '... Do you think the PVV will stop supporting the government?'

	Real Result (peil.nl)			Survey Experiment Result		
	Yes	No	DK	Yes	No	DK
All respondents	34%	55%	11%	32%	50%	18%
PVV voters	52%	37%	11%	44%	43%	13%

The survey question read: 'Suppose that next year another 5 billion euro in budget cuts have to be made. Do you think the PVV will stop supporting the government?'. The question was part of a survey about the future of Rutte Cabinet I and about what would happen to this minority VVD-CDA cabinet, supported in parliament by the PVV, although that party did not have any ministers in the government[35]. The results were originally published on the Peil.nl website with an introduction that started as follows: 'Cabinet Rutte exists 1 year. What do the voters expect for the future of this cabinet?[36]'

Table 6.7 contains both the original Peil.nl results and the results of survey experiment variant 3A[37], i.e. the explicit DK variant which is identical to the original question format and offered non-substantive response options. A majority of respondents of Peil.nl (55 percent) expected the PVV not to stop supporting the government if additional budget cuts were necessary. This is rather similar to the 50 percent saying no to the question in the survey experiment, even though item nonresponse is slightly higher in the latter case (18 compared to 11 percent). Looking specifically at PVV voters, however, the conclusions differ: 52 percent of PVV voters in Peil.nl answered 'yes', compared to 44 percent in the survey experiment: the majority has become a plurality. Furthermore, the gap between 'yes' and 'no' has dwindled (from 15 to 1 percent). This means that the poll and experimental results differ substantially. A potential explanation is Peil.nl's weighting procedure. Unfortunately this website does not publish the adjustments that were made to the raw data, so this procedure could not be applied to the experimental dataset.

---

35   The minority VVD-CDA cabinet, also called Rutte-I, governed from 2010 until 2012. The cabinet fell in 2012 after extensive discussion about budget cuts.

36   Original Dutch introduction: 'Kabinet Rutte bestaat 1 jaar. Wat verwachten de kiezers van de toekomst van dit cabinet? Published on the website www.peil.nl on October 16th, 2011.

37   The Peil.nl survey was carried out in the week before the 16th of October, 2011; the survey experiment was carried out from 28 October to November 9th 2011.
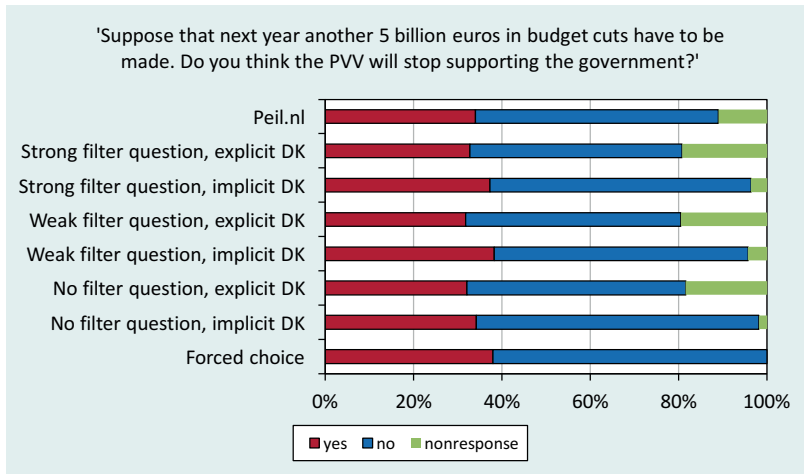
Figure 6.11: Results Seven Variants for PVV Support Government – All Respondents
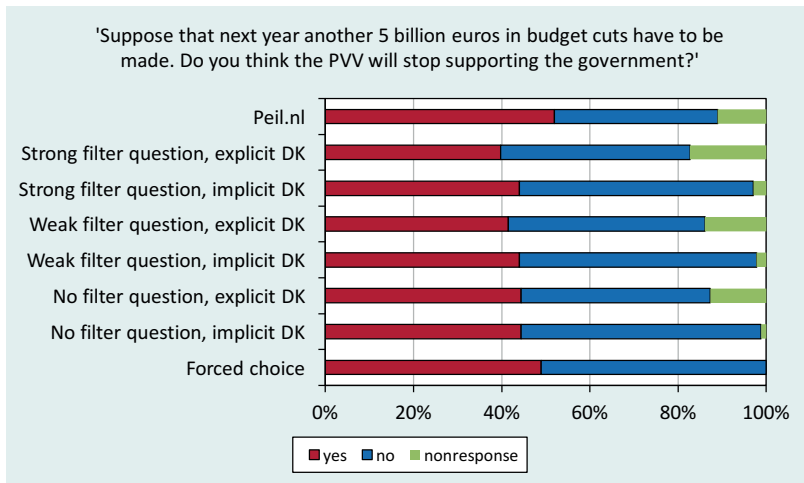
Figure 6.12: Results Seven Variants for PVV Support Government –PVV Voters

To see the effect of offering non-substantive response options on the distribution of opinions, Figure 6.11 (all respondents) and Figure 6.12 (PVV voters) present the results of the seven questionnaire variants.

Figure 6.11 and Figure 6.12 give an indication of what voters in October 2011 expected for the Rutte Cabinet I. The results include the people who used the filter question or said they did not know whether the PVV would stop supporting the government. For all subgroups in Figure 6.11, the largest group expected that the PVV would not stop supporting the government. This is consistent with Peil.nl's outcome of 55 percent, but the largest group is not always a majority: in variant 1A and 2A the percentage of respondents saying 'no' is 48 and 49 and in variant 3A it is 49 percent. So in three of the seven variants half of the respondents seem to hold the position that the PVV would not stop supporting the government.

For the PVV voters (Figure 6.12), the effect is actually the other way around: all subgroups show higher percentages saying 'no' than the 37 percent of the PVV voters reported by Peil.nl. In the implicit DK variants 1B, 2B and 3B it even is a majority saying the PVV would still support the government. In this case, there was therefore also an effect of (not) offering certain response options on the distribution of opinions.

6.5.4 Question Content and Item Nonresponse

Some themes may be more susceptible to question design effect than others. The hypotheses followed from two questions: 1) is the issue related to one of the main dimensions in Dutch politics (McClosky & Zaller, 1984; Wittkopf, 1990) and 2) is the issue related to foreign policy? To examine the relation between question content and item nonresponse, the questions have been ranked per the level of item nonresponse in variant 1A. The analysis of question design effects looks exclusively at item nonresponse, since it is established already that the data seem to be missing at random and the main effect of filter questions is item nonresponse rates.

Table 6.8 shows the ranking based on item nonresponse resulting from the filter question; Table 6.9 shows the total item nonresponse resulting from all non-substantive response options.

Table 6.8: Filtered Out Item Nonresponse (%) – Ranked with Variant 1A

Question	'No' to Filter Question in Variant 1A
Self-placement Income Differences	19
Welfare benefits should be lowered in order to stimulate people to work	17
Adoption by same-sex couples should be possible	8
What do you think should happen to the mortgage interest deduction?	
What do you think is the best solution for the impending deficits of pension funds?	6
Do you think that Maxima's father can or cannot be present at the coronation?	
Self-placement European Unification	
The Netherlands should spend more money on developmental aid	4
The Netherlands should in the next year quit the euro and go back to the gulden	
I think that Libya will, in time, become a normal democratic country	
The Netherlands should be actively involved to help Libya establish a democratic regime	3
Self-placement Euthanasia	
Self-placement Foreigners	
There are too many people of a non-Dutch nationality living in the Netherlands	
Do you think that the King or Queen should have political influence, or should s/he restrict herself to ceremonial roles?	2
How long do you think this cabinet will remain in office?	
Suppose that next year another 5 billion euro in budget cuts have to be made. Do you think the PVV will stop supporting the government?	

At first sight, the results (Table 6.8) seem to falsify hypothesis H3a stating that item nonresponse resulting from a filter question is lower when a question is related to a major political dimension. The three items with the highest percentage of respondents saying 'no' to the filter question are self-placement income differences, lowering welfare benefits and adoption by same-sex couples; these items are supposed to be related to a socio-economic or ethical dimension. The other ethical item (self-placement euthanasia) and the two multicultural items, however, render the least item nonresponse (2 percent). So overall results are mixed: some items related to a major political dimension result in less item nonresponse, i.e. 'no' to a filter question, but other items result in more item nonresponse.

The results are also mixed when the second non-substantive response option, i.e. the explicit DK option, is included (see Table 6.9: Total Item Nonresponse (%) – Ranked with Variant 1A). Three of the four items with the highest total item nonresponse rate (the deficits of pension funds, self-placement income differences,

Table 6.9: Total Item Nonresponse (%) – Ranked with Variant 1A

Question	Item Nonresponse
What do you think is the best solution for the impending deficits of pension funds?	32
I think that Libya will, in time, become a normal democratic country	23
Self-placement Income Differences	20
Welfare benefits should be lowered in order to stimulate people to work	19
Suppose that next year another 5 billion euro in budget cuts have to be made. Do you think the PVV will stop supporting the government?	18
The Netherlands should in the next year quit the euro and go back to the gulden	
Adoption by same-sex couples should be possible	
The Netherlands should be actively involved to help Libya establish a democratic regime	10
Do you think that Maxima's father can or cannot be present at the coronation?	
How long do you think this cabinet will remain in office?	8
The Netherlands should spend more money on developmental aid	7
What do you think should happen to the mortgage interest deduction?	6
There are too many people of a non-Dutch nationality living in the Netherlands	5
Do you think that the King or Queen should have political influence, or should s/he restrict herself to ceremonial roles?	4
Self-placement European Unification	
Self-placement Euthanasia	3
Self-placement Foreigners	2

and lowering welfare benefits) can be related to the socio-economic theme. Three items related to the ethical and multicultural themes are, however, among the items with the least item nonresponse, i.e. the least number of no's to the preceding filter question and DK answers. Hypothesis H3a is rejected.

What is also unexpected is that the questions within the foreign affairs theme are not ranked higher, i.e. questions on EU self-placement, money spent on development aid and questions about Libya. The expectation (hypothesis H3b) was that the item nonresponse for survey questions about foreign policy issues would be relatively higher. This expectation is not supported by the results in Table 6.8 and Table 6.9: Total Item Nonresponse (%) – Ranked with Variant 1A: questions about foreign policy issues are not ranked higher in item nonresponse. The only exception is a question on whether Libya will become a normal democratic country, resulting in the second highest total item nonresponse (23 percent). The percentage of respondents saying no to the preceding filter question, however, is relatively low with 4 percent (see Table 6.8). Hence the conclusion

about foreign policy issues that the overall item nonresponse is not higher.

Finally, the self-placement questions, with the exception of the one about income differences, have the least item nonresponse: 2 to 4 percent. A possible explanation relates to the number of response categories, or the availability of a neutral midpoint category. The number of response categories for the self-placement items is indicated in the introduction of the filter question and respondents were therefore able to decide whether they would be able to place themselves on a seven-point scale. This will be explored in the section about response categories.

6.5.5 Number of Response Categories & Neutral Response Category

It was hypothesized that survey questions with more response categories result in relatively less item nonresponse. Table 6.10 presents the number of DK answers, implicitly or explicitly, for individual survey questions, ranked according to the number of substantive response categories. Strictly speaking, a better test of the hypothesis would have been to manipulate the number of answer categories while holding all other elements (wording, non-substantive response options, etcetera) constant. The answer categories, however, are not a central part of the research question and were not manipulated. The analysis presented here only explores the relation between the number of answer categories and item nonresponse.

Table 6.10: DK Item Nonresponse (%) of Individual Items – Ranked to Number of Response Categories

Number of Substantive Response Categories	Question	Explicit DK Variants			Implicit DK Variants		
		1A. Strong Filter	2A. Weak Filter	3A. No Filter	1B. Strong Filter	2B. Weak Filter	3B. No Filter
7	Self-placement Income Differences	1	1	1	2	2	1
	Self-placement Euthanasia	0	0	1	1	1	0
	Self-placement Foreigners	0	0	0	1	1	0
	Self-placement EU	0	0	2	1	1	0
4	The welfare benefits should be lowered in order to stimulate people to work	2	3	4	1	1	1
	Adoption by same-sex couples should be possible	2	1	5	1	1	1
	There are too many people of a non-Dutch nationality living in the Netherlands	4	2	4	1	1	1
	The Netherlands should spend more money on developmental aid	3	3	5	1	1	1
	How long do you think this cabinet will remain in office?	6	5	6	1	1	1
3	What do you think should happen to the mortgage interest deduction?	0	0	3	0	1	1
2	What do you think is the best solution for the impending deficits of the pension funds?	26	21	30	4	4	6
	The Netherlands should next year quit the euro and go back to the gulden	7	5	8	1	1	2
	Do you think that the King/Queen should have political influence or should s/he restrict herself to ceremonial roles?	2	1	3	1	1	1
	Do you think that Maxima's father can or cannot be present at the coronation?	4	2	8	1	1	2
	I think that Libya will, in time, become a normal democratic country.	21	19	23	2	1	2
	The Netherlands should be actively involved to help Libya establish a democratic regime	8	7	9	2	2	2
	Suppose that next year... Do you think the PVV will stop supporting the government?	18	17	18	2	2	2

DK Item Nonresponse is measured as a percentage of the respondents using a DK option in response to an opinion question. Respondents saying 'no' to the preceding filter question are excluded from the analysis.

In Table 6.10 the questions are ranked according to the number of response categories they offered. Four self-placement questions included a seven-point scale, five questions included four substantive response categories, one question included three and seven questions offered two substantive response options. Only DK answers are included as item nonresponse, because the respondent has to see the substantive response categories in order to determine whether the response categories fit the preferred answer. If no suitable response category is available, the respondent may decide to use the DK option.

The items with the least substantive response categories generate the highest item nonresponse in Table 6.10. Questions about the deficits of pension funds (30 percent DK answers), Libya (23 percent) and the PVV's support for the government (18 percent explicit DK) all offer only two substantive answers and result in high item nonresponse rates. It is interesting to note that two of the three items with two response categories, i.e. Libya and PVV's support for the government, are about expectations and not opinions. If these two are excluded, the effect partially disappears.

The self-placement items, offering more options (and a midpoint category, which is discussed below), are ranked highest. For all self-placement items, the DK item nonresponse is clearly lower than for other survey items; the numbers are also lower compared to the other questions referring to the same theme. The self-placement item addressing the multicultural theme (Foreigners), for example, results in 0 to 1 percent DK answers which is considerably lower than the 4 percent of the related item ('There are too many people of a non-Dutch nationality living in the Netherlands'). These findings suggest that a limited number of response categories results in a more frequent use of the DK option. Hypothesis H4a is supported.

One note is in order, however. The conclusion that more substantive response categories result in less item nonresponse can only be drawn for the explicit DK variants (1A, 2A and 3A). The implicit DK option was used so infrequently, that the differences between items with more and fewer substantive response categories are negligible: for example, up to 2 percent skipped the self-placement income item and up to 4 (implicit filter variants 1B and 2B) and 6 (no filter variant 3B) percent used the implicit DK option for the question about pension fund deficits with two response categories. So hypothesis H4a is supported for the explicit DK option, but no effect of the number of response categories is found for the implicit DK variants.

The second issue in this section is the presence and use of the midpoint category, which was included in the self-placement items and for the question about mortgage interest deduction. Based on the meanings of the midpoint or neutral category as either a truly neutral position or as a disguised nonresponse, it

is expected that for questions where a middle response category is offered at least part of the item nonresponse is substituted by such a neutral answer. In the absence of a non-substantive response option, it is expected that at least some respondents use the midpoint option as a pseudo non-substantive response option. Table 6.11 shows how often the midpoint option was used in each variant.

Table 6.11: Use of the Neutral Response Category (%)

Question	Explicit DK Variants			Implicit DK Variants			4
	1A. Strong Filter	2A. Weak Filter	3A. No Filter	1B. Strong Filter	2B. Weak Filter	3B. No Filter	Forced Choice
Self-placement Income Differences	21	18	24	20	18	24	23
Self-placement Euthanasia	5	5	5	6	5	6	6
Self-placement Foreigners	18	19	16	20	18	17	19
Self-placement EU	14	13	14	15	13	17	17
What do you think should happen to the mortgage interest deduction?	54	54	50	57	53	52	52

Use of the Neutral Response Category is measured as a percentage of answers to the substantive opinion question. Respondents saying 'no' to the preceding filter question are excluded from the analysis.

Hypothesis H4b expected more use of a midpoint option in variants where no non-substantive response option is offered. The data do not show this pattern, however (see Table 6.11): the forced choice variant results in none of the five items with a neutral response category in a more frequent use of that midpoint option. Respondents who have the (explicit) DK option available do not use the neutral response category less than the forced choice respondents.

What also cannot be observed is that in the implicit DK variants (1B, 2B and 3B) the midpoint option is consistently used more often than in the explicit DK variants (1A, 2A and 3A). This was expected because although respondents may be able to skip the question, they are not visually alerted to that option. Like the respondents of the forced choice variant, the implicit DK respondents do not use the midpoint option more often; at least there is no clear and consistent trend.

So the overall conclusion about the use of the midpoint category can only be

that it is not used more often by respondents who do not have other non-substantive response options available (in the forced choice variant) or by respondents who do not have explicit non-substantive response options (in the implicit DK variants). This is a falsification of hypothesis H4b; offering a nonresponse (i.e. DK) option does not result in less use of the neutral response category.

6.5.6   Question Design and Break-Offs

Partial nonresponse becomes a threat to data quality when it results in data not missing at random, i.e. a nonresponse bias. In this study, only the amount of partial nonresponse was registered; these data reveal how much missing data (break-offs) result from question design choices. The specific research question is: how do non-substantive response options, i.e. filter questions and the DK option, influence partial item nonresponse? The expectation (hypothesis H6) is that since non-substantive response options provide an easy way out and therefore require less time and effort, the respondents of variants with explicit non-substantive response options are less likely to drop out. When respondents are unable to see that they do not have or are not willing to give an opinion, this results in frustration and eventual break-offs.
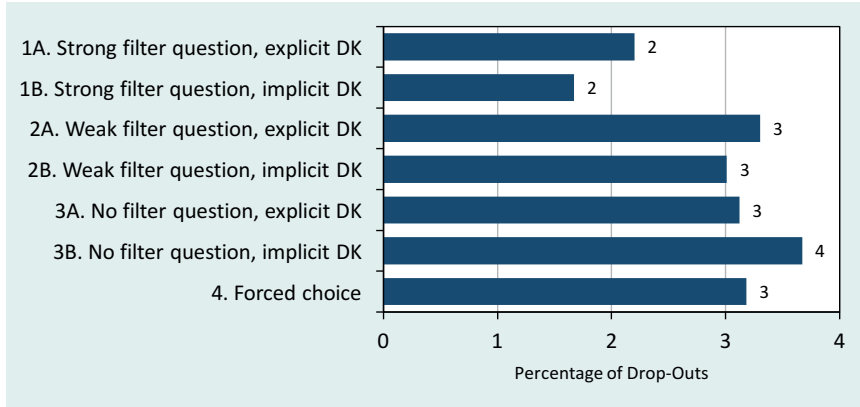
Figure 6.13: Drop-Outs (%) per Survey Questionnaire Variant

Figure 6.13 displays how many respondents dropped out of the survey after having started. The variants with an explicit DK option (1A, 2A and 3A) should result in fewer drop-outs. Indeed, for the strong filter variants the percentage of drop-outs is slightly lower in the explicit DK variant than in the implicit DK variants, but the same cannot be concluded about the weak filter and no filter variants. And while

the forced choice variant should lead to most frustration and consequently most drop-outs, it is one of the two variants with the least break-offs (together with implicit DK variant 3B). The number of break-offs in variants with an implicit DK option or no non-substantive response option at all (forced choice) is not higher than in the variants with an explicit DK option and/or a filter question. Hypothesis 6 is rejected.

It could be argued, however, that the filter variants (1A, 1B, 2A and 2B) result in more drop-outs because of questionnaire length and the increased survey burden. If longer questionnaires result in more break-offs, one would expect relatively high numbers towards the end of the survey. Figure 6.14 and Figure 6.15 show when the respondents of the seven variants dropped out[38]. However, most respondents already drop out soon after starting the survey, i.e. during the first three opinion questions (or the first six questions including filter questions), regardless of the variant they were assigned to. This is not in line with the expectation that questionnaire length is an alternative explanatory factor for the number of break-offs. Respondents simply decide early whether they want to proceed with the survey or not, regardless of design aspects.
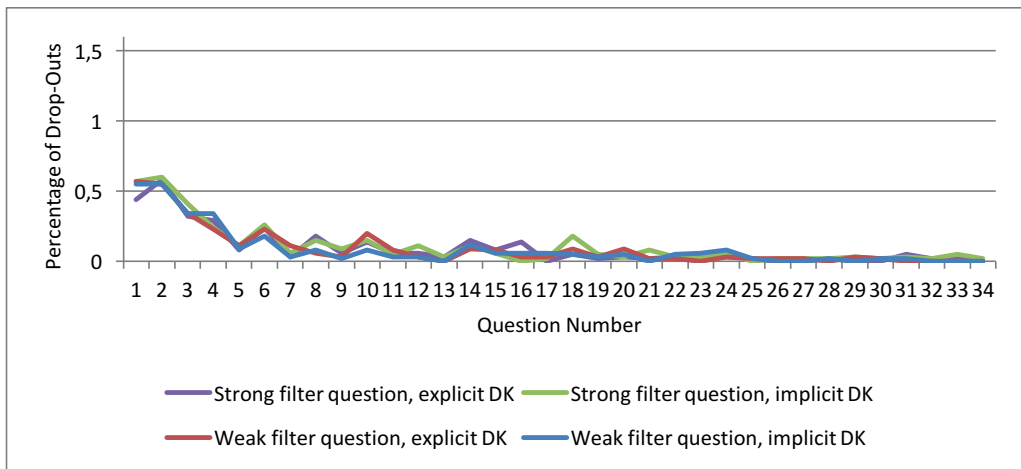
Figure 6.14: Drop-Outs (%) Filter Question Variants

---

38 It is confusing to display all variants at the same time, because some variants do not include filter questions while others do. The drop-out rates of the seven questionnaire variants are therefore split into two figures, one for the filter question variants and one for the other (no filter question and forced choice) variants.
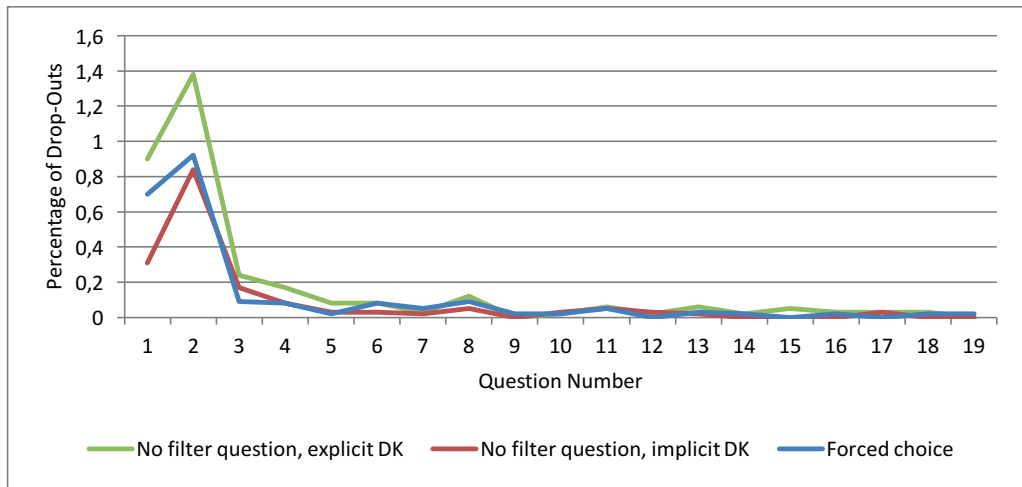
Figure 6.15: Drop-Outs (%) No Filter and Forced Choice Variants

6.6   Conclusion

In this chapter, the results of a survey experiment on the effects of a filter question were presented. The effects were assessed in terms of the collection and aggregation of substantive and non-substantive answers. The expectation was that using a filter question would result in an item nonresponse of about 20 percent regardless of question content, as suggested in the literature. Different types of questions were examined in terms of the number of response categories and the subject of the survey question. Furthermore, the number of drop-outs was examined. The experiment was carried out with the EenVandaag Opiniepanel.

The effect of a filter question can be assessed in two ways: 1) item nonresponse resulting from a 'no' to a filter question, which reveals only what part of the public has an opinion; and 2) the distribution of substantive answers, which shows what overall public opinion looks like. Posing a filter question before an opinion question has a strong effect on item nonresponse, although 20 percent proved to be an overestimation: about 10 to 11 percent of the respondents said 'no' to the filter question, depending on the wording of the filter question and regardless of question content. These findings support hypothesis H1a: explicit non-substantive response options result in more item nonresponse. Hypothesis H1b is also supported: questionnaire variants with filter questions rendered more item nonresponse than an explicit DK option. Nevertheless, the influence of filter questions on item nonresponse is not much stronger than the influence of an explicit DK option; only when a variant

offers both non-substantive response options the item nonresponse is substantially higher than compared variants with an explicit DK option.

The effect of a filter question on the distribution of opinions depends on whether item nonresponse is treated as missing data and as such excluded from the analysis, or included in the distribution. The emerging overall picture of public opinion does not change if a filter question is added and item nonresponse is excluded. If item nonresponse is considered valuable information about for instance public ignorance, the picture of public opinion changes in various ways for some items. Majorities become pluralities, and majorities become less pronounced – 'making' the public more divided about the particular issue. So despite the fact that no nonresponse bias is discovered and hypothesis H2b is rejected since missing data are missing at random (which supports hypothesis H2a), the picture of public opinion can be affected by question design and particularly by the use of a filter question. The distribution of opinions, however, and the 'substantive proportions' are not substantially affected and suggest a limited effect of using a filter question on survey results.

The finding that the picture of public opinion does not change when a filter question (and other non-substantive response options like the explicit DK option) is used, is rather surprising. Missing data appear to be randomly distributed, which seems counterintuitive. More item nonresponse resulting from filter questions definitively leaves more room for a bias and even though previous studies noted that 'the filtered distribution of opinions sometimes differs from the unfiltered (standard) distribution and sometimes does not' (McClendon & Alwin, 1993, p. 439), it was expected that the filtered distribution would differ from the unfiltered distribution for at least some issues. The levels of item nonresponse resulting from filter questions do suggest a less involved public than is actually the case. A politician trying to be responsive to the public could be acting upon what he thinks to be a majority which is actually a plurality or even a minority. Filter questions are therefore a valuable tool to reveal the complete picture of public opinion, including that part of the public that does not have or does not want to express an opinion.

Besides the effect of filter questions on item nonresponse and the distribution of opinions, other factors were considered: question content and response categories. The hypotheses regarding question content (H3a and H3b) could not be confirmed; questions related to a major political dimension in Dutch politics did not consistently result in less item nonresponse and foreign policy issues did not generate more item nonresponse. The same holds true for hypothesis H4b: the midpoint or 'neutral' response category was not used more often by respondents who were not offered non-substantive response options; hypothesis H4b was falsified. Finally, the expectation that more respondents would drop out of the survey in the implicit DK

6

and forced choice variants, because of frustration due to the lack of non-substantive response options, was also not supported by the data.

The main conclusion here is that design choices concerning questionnaires or, in other words, the way the questions are asked can and in practice do affect the results significantly. Furthermore, small changes may have consequences due to shifts in majorities or pluralities.

CHAPTER 7   The Follow-Up Question

7.1   Introduction

Suppose that public opinion about nuclear energy is assessed with a survey, because it is a hotly debated topic. Respondents are asked whether they want a ban on nuclear energy or not. 30 percent of the respondents is against a ban and wants to permit the use and production of nuclear energy; 60 percent supports the ban. The remaining 10 percent did not give an opinion. It seems clear what the public wants: a majority of 60 percent wants to ban nuclear energy. This picture becomes even clearer when item nonresponse is excluded as missing data; in that case 67 percent supports the ban.

Suppose that the respondents were asked in a follow-up question whether they would be upset if their opinion did not prevail and 25 percent of the (total number of) respondents who previously indicated they supported the ban, say they would not be upset (i.e. gave a permissive opinion). The other 35 percent (of the initial majority of 60 percent) would be upset if nuclear energy was not banned. The public now seems more divided: 30 percent of the respondents does not want a ban on nuclear energy and wants to see it transformed into a policy, 35 percent does want a ban and wants this preference executed, and the remaining 35 percent either gave a nonresponse or said they would not be upset if their opinion did not prevail. The picture of public opinion has suddenly become much more blurred.

This may seem a far-fetched example, but earlier applications of this follow-up question in the US by Gallup (see Table 7.1) show that the problem is real. At least for some issues the public turned out to be more uncaring than suggested by a standard survey without a follow-up question. And if respondents do not care about what happens to their opinion, why should others, i.e. pollsters and politicians, do?

Despite the fact that 'public policy polls should be a boon in a governmental system where the people elect their representatives', due to various issues this democratic potential often is not utilized (Moore, 2011). Moore (2008, pp. 145-146; Moore, 2011) summarizes three main problems: 1) non-opinions are ignored and not measured by pollsters; 2) the 'intensity' of opinion is not assessed; and 3) no differentiation is made between so-called hypothetical and actual opinions.

This chapter is the third part in a series of survey experiments of a project about the effects of question design on survey outcomes and the relevance of these outcomes for the public debate and democratic process. In this third experiment opinions are taken as a given, but the way they may affect political decision-making is central, i.e. the extent to which people want politicians to be responsive to their opinion. The salience, 'weight' or 'intensity' of an opinion is taken into account. Rather than determining whether an opinion is present, which is an assessment of non-opinions analyzed in the previous chapters, here the respondent is asked

7

whether he or she cares about what happens with his or her opinion. The respondent did give an opinion, but so what? Note that this variant is strictly speaking not a measurement of item nonresponse, but primarily an indicator of the extent to which the respondent thinks that his or her opinion matters or should matter.

The experiment presented is inspired by Moore (2008) and in particular his follow-up question: 'Would you be upset if the previously expressed opinion did not prevail when the issue was ultimately decided?'. The follow-up question is supposed to measure the '*intensity of opinion*' (Moore, 2008, pp. 145-146). This question allows more variation in the spectrum of non-attitudes to 'real' attitudes and an assessment of 'real' politically relevant public opinion: the focus is on opinions that the people themselves want to see carried out. Whereas high levels of item nonresponse raise doubts about the existence of public opinion concerning a certain subject, the 'so what' follow-up question renders information about the weight and importance of individual opinions and public opinion as a whole.

The research question of the third survey experiment and this chapter is: *to what extent do people want their individual opinions to matter and/or to be translated into policy?* To what extent do respondents 'care about' their opinion? Is it a 'directive opinion', in which case they expect it to be executed (Traugott, 2009, p. 432)?

7.2   Theoretical Reflection

The question whether public opinion measured via polls is as valid and robust as often assumed, is not only relevant for item nonresponse. 'My view is that at important policy matters, pollsters should measure at least three dimensions of public opinion: 1) direction of support (from support to opposition), including the magnitude; 2) intensity of views; and 3) the absence of a meaningful view on the matter, or non-opinion' (Moore, 2009). The first and third aspect were discussed in previous chapters; in this chapter the second aspect is central. Moore labels this as *intensity,* although it arguably is not about the strength of the opinion as such but whether and how the individual opinion should be included in the decision-making process. The focus is not primarily on 'the strength of an individual's feelings about an attitude object', which would be a way to measure attitude intensity (Krosnick & Abelson, 1994, p. 179; see also Schuman & Presser, 1996), but on the preferred implications of this individual opinion.

Follow-up questions may gauge to what extent an expressed opinion is '*directive*' instead of '*permissive*' (Moore, 2008, pp. 6-7). According to Lyons (2004) directive opinions reveal a psychological investment; such an 'opinion can in some way be considered an order or a directive for public officials'. Respondents with directive

opinions want to see their preference transformed into policies. Permissive opinions, on the other hand, permit the decision-making to go either way. The respondent does not really or deeply care how the issue is ultimately decided. From this perspective, 'so what'-follow-up questions collect additional information about the importance and desired impact of individual opinions and public opinion as a whole. It should be noted that giving a 'not upset' answer to a follow-up question does not necessarily mean that the respondent does not care about how the issue is decided. A permissive opinion might also reflect a sense of reality, because politicians do not (always) listen to the public, or a sense of how democracy works when a majority holds a different view than the respondent's.

The main difference between directive and permissive opinions is in their intended effects: should policies reflect public opinion? 'Permissive opinions "permit" the country's leaders to do whatever the leaders deem best' whereas citizens with directive opinions want to see their opinions executed (Moore, 2004). In the latter case, the citizen likely has strong beliefs about the way an issue should be solved and the expressed opinion can be considered an 'order' which should be carried out (Lyons, 2004). Listening to what people want is at the core of representative democracy; surveys are the means to assess the public's wants and needs. The more qualified idea is, however, that responsive politicians do not have to listen to opinions if the people themselves do not care, even when they do express an opinion. In such a situation the decision-makers would have 'a great deal of latitude' (Moore, 2004).

So besides the question whether a large part of the public *has* an opinion about a subject, another but related important question is whether those citizens who are giving their opinion also want it to practically *matter*. The answer to this question is not obvious or clear. Moore (2008) presents examples of follow-up questions in which the respondent was asked whether s/he would be upset if the previously expressed opinion did not prevail when the issue was ultimately decided; other examples can be found on www.gallup.com. To illustrate the rather innovative follow-up question and provide context for the findings in this chapter, an overview of some of these American examples is presented in Table 7.1.

One example in Table 7.1 is the Pledge of Allegiance, which refers to a decision in 2004 by the US Supreme Court on whether the words 'under God' should be removed from the Pledge, as the pledge was originally worded before 1954. In the preceding months, Gallup polled what the decision should be. The outcome was clear: 91 percent thought the words 'under God' should remain part of the pledge (the percentages of 78 plus 13 percent 'Keep' in Table 7.1. This is a very clear majority preference. The majority preference remained when a distinction was made between permissive and directive opinions: a large majority of 78 percent had

Table 7.1: Previous Application of Follow-up Questions by Gallup

Subject and Data Collection	Question wording	Directive	Permissive	
			No Opinion or Unsure	*Not upset*
War with Iraq *17-19 February 2003*	'Would you favor or oppose sending American ground troops to the Persian Gulf in an attempt to remove Saddam Hussein from power in Iraq?'	59% 29% Favor; 30% Oppose	3% No Opinion	38% 30% Favor; 8% Oppose
Pledge of allegiance *26-28 March 2004*	'As you may know, the Supreme Court is considering whether the Pledge of Allegiance should continue to include the words "under God" as part of the pledge. Which would you prefer?'	81% 78% Keep; 3% Remove		19% 13% Keep; 5% Remove
Anti-missile defense system *22-24 April 2002*	'Do you think the government should or should not spend the money that would be necessary to build [a defense system against nuclear missiles]?'	42% 29% Favor; 13% Oppose	6% Unsure	53% 35% Favor; 17% Oppose
Privatization of social security *8-11 April 2002*	'A proposal has been made that would allow people to cut a portion of their Social Security payroll taxes into personal retirement accounts that would be invested in private stocks and bonds. Do you favor or oppose this proposal?'	47% 27% Favor; 20% Oppose	4% No Opinion	49% 36% Favor; 13% Oppose
Defense spending: development missile defense shield *4-6 February 2002*	'... Do you think we are spending too little, about the right amount, or too much?'	41% 28% Favor; 13% Oppose	33% No Opinion	26% 19% Favor; 7% Oppose
Follow-up defense spending: missile defense shield *3-6 February 2003*	'Should the United States spend the money to develop a missile defense system?'	44% 28% Favor; 16% Oppose	33% Unsure	23% 18% Favor; 5% Oppose

Table 7.1: Continued

Subject and Data Collection	Question wording	Directive	Permissive	
			No Opinion or Unsure	Not upset
Kyoto agreement *8-11 March 2004*	'... Should the US agree to abide by the provisions of the Kyoto agreement on global warming?'	40% 30% Favor; 10% Oppose	36% No Opinion	24% 12% Favor; 12% Oppose
Oil drilling in the Arctic National Wildlife Refuge *7-10 March 2005*	'Do you think the Arctic National Wildlife Refuge in Alaska should or should not be opened up for oil exploration?'	64% 19% Should 45% Should Not	5% No Opinion	31% 23% Should; 8% Should Not
Closing the prison at Guantanamo Bay *6-8 July 2007*	'Do you think the US should – or should not – close the prison at the Guantanamo Bay military base in Cuba?'	47% 19% Favor 28% Oppose	13% DK or Unsure	39% 14% Favor; 25% Oppose

Sources: (Carlson, 2002; Lyons, 2004; Moore, 2002, 2003, 2004, 2005, 2008).

Note: a *directive* opinion is one where respondents explicitly indicated that they 'would be upset if the issue was not resolved to their liking' (Carlson, 2002). The column first summarizes the total number of directive opinions for each subject without their actual preference. After that, the percentage of directive opinions in favor and opposing the statement are indicated (in the same cell). A *permissive* opinion is one where respondents indicate they are not (too) upset or did not give an opinion at all. Thus, the item nonresponse combined with non-upset respondents are grouped into the permissive category.

a directive opinion of keeping the words. Furthermore, only 19 percent said they would not be upset if the issue was decided otherwise or did not give an opinion, which is a low percentage of permissive opinions. Consequently, whichever way these opinions were measured, the resulting overall picture of public opinion remained the same.

The experiment reported here will examine whether the outcome of a survey, executed with an internet panel in the Netherlands, is affected by the use of a follow-up question. A follow-up question could affect the outcome in various ways. First, no majority may support either side strongly, because many respondents gave a permissive opinion; the majority favoring sending American ground troops to the Persian Gulf disappears when the distinction between permissive and directive opinions is introduced (see Table 7.1). Secondly, the substantive outcome may change;

a plurality opposing the sending of ground troops with a directive opinion, when at first sight it seemed like a majority was in favor of it when the distinction between permissive and directive opinions was not included. This second consequence means that a different overall picture of public opinion emerges.

7.3   Expectations

The hypotheses used for the other two experiments referred specifically to non-substantive answers and the overall distribution of opinions. This chapter is about a slightly different topic, which is why new expectations will be developed which are tailored to the measurement of directive opinions with a follow-up question. They focus on two main points: 1) the number of permissive and directive opinions, both in general and for specific subjects; and 2) the resulting picture of public opinion, both including and excluding permissive opinions. Note, however, that these expectations have less ground in the literature, because the follow-up question has not been conceptualized and analysed as much as non-substantive response options. So only a number of more general expectations have been developed to guide the discussion and analysis of the effect of a follow-up question.

The first expectation makes a distinction between nonresponse and 'not upset'. The expectation is that for issues generating relatively much item nonresponse, the number of directive opinions is relatively low; item nonresponse and the 'directiveness' of opinions are expected to be negatively correlated since both the number of 'not upset' opinions and item nonresponse are assumed to be indicators of whether respondents care about an issue. Item nonresponse and 'not upset' answers are both treated as permissive opinions, but the expectation differentiates between these two permissive opinion categories.

The expectation (E1a and E1b) is that survey questions with higher item nonresponse rates have a relatively low proportion of directive opinions. When item nonresponse is included as permissive opinions, this expectation (E1a) is rather obvious or even tautological: questions with more item nonresponse by definition result in less directive opinions. Respondents using a non-substantive response option are not asked the follow-up question and therefore fewer people can indicate that they have a directive opinion. If 30 percent of the respondents use a non-substantive response option, only the remaining 70 percent can say whether their opinion is permissive or directive.

Even with item nonresponse excluded, however, the expectation is that items with relatively high item nonresponse rates also have fewer directive and subsequently more permissive opinions. Item nonresponse may be one indicator of public

ignorance and whether people care about the issue; the 'upsetness' or 'directiveness' of opinions may be considered another indicator of whether respondents care about the issue. The expectation is therefore that both including and excluding nonresponse (as part of the permissive opinions), more item nonresponse correlates with less directive opinions.

E1a: The more item nonresponse is measured, fewer directive opinions are measured (with item nonresponse included).

E1b: The more item nonresponse is measured, fewer directive opinions are measured (without item nonresponse).

Next, three expectations are developed about directive opinions and question content. Building on the idea that people do not have opinions about everything, but that there may be some core dimensions which organize their attitudes and are central to public and political debate (McClosky & Zaller, 1984; Wittkopf, 1990), the expectation is that issues related to core dimensions (in Dutch politics) result in more directive opinions. Furthermore, it is expected that in abstract or technical issues like foreign policy, opinions are less directive, since they are supposed to be less coherent, less stable and less informed (Alvarez & Brehm, 2002, p. 214; Everts, 2008, pp. 8-14).

E2a: If the topic of a survey question is related to a major political dimension, the number of directive opinions is higher compared to a survey question that is not related to such a dimension.

E2b: The number of directive opinions for questions about foreign policy issues is lower than for questions about issues related to the core dimensions.

Expectations E1a, E1b, E2a and E2b focus on the number of directive and permissive opinions while ignoring the substantive opinions. For the overall picture of public opinion, however, it is crucial to look at what the public wants in terms of the majority or plurality preference and not only whether respondents care about their opinion. The question is whether majorities or pluralities change or disappear when the distinction between directive and permissive opinions is taken into account.

E3: The use of a follow-up question affects the substantive distribution of opinions.

7

7.4   Data and Methods

The experiment is an internet survey experiment with respondents of the Team Vier internet panel. It was executed with a split sample design. Six subgroups were randomly selected from the panel and subjected to one particular question design. Respondents were randomly assigned to a group, so it can be assumed that the groups are similar; differences in response patterns result from question form, i.e. the possibilities to register non-substantive answers. The use of a follow-up question was not varied: this question was asked *after* a substantive answer was given to the opinion question and does not affect the substantive answer. The analysis presented here focuses on the impact of the follow-up question.

The general instruction of the block of questions read that respondents were asked to give their opinion on various issues. The order of the blocks of questions was randomized to avoid question-order effects. Scales for self-placement were placed horizontally on the screen while the answer categories of the other options were ranked vertically, which was consistent with the way they were originally offered. The experiment was fielded between July 24th and August 7th, 2012; respondents had two weeks to complete the questionnaire. Five general themes were addressed and each theme included three questions of which at least one question came from existing research – see section 4.4 for more information about issue selection.

Table 7.2 shows some descriptive statistics of the sample and subgroups. The sample is not representative of the population. Women are, for example, overrepresented: on average, 63.7 percent of the sample was female. The Team Vier internet panel is a nonprobability online panel, which is not a randomly drawn sample. Generalization is not possible. It is, however, possible to produce internally valid findings and explore causal mechanisms, due to the between-subjects-design and the resulting similarity of the subgroups. No significant differences were found between the variants in terms of gender, age and region of the respondent. These results confirm that the subgroups are similar and that any differences in the outcome can be attributed to the variant to which the respondent was assigned.

Table 7.2: Descriptive Statistics of Demographic Characteristics

		1A	1B	1C	2A	2B	2C	Total
Gender*	Male	35.1%	37.8%	39.2%	35.5%	35.2%	34.9%	36.3%
	Female	64.9%	62.2%	60.8%	64.5%	64.8%	65.1%	63.7%
Age (years)*	Mean	53.8	52.4	53.9	54.8	54.6	54.2	53.9
	SD	12.4	12.5	12.8	12.2	11.1	12.8	12.3
Region*	3Gem	19.5%	10.4%	14.8%	13.1%	13.2%	15.5%	14.4%
	Noord	14.3%	10.4%	10.8%	13.1%	12.8%	9.9%	11.9%
	Oost	19.9%	19.7%	17.6%	18.7%	19.6%	19.8%	19.2%
	West	25.5%	32.9%	30.8%	25.1%	34.4%	30.2%	29.8%
	Zuid	20.7%	26.5%	26.0%	29.9%	20.0%	24.6%	24.6%
N		251	249	250	251	250	252	1503

*No statistical significant differences between the subgroups at the .05 level

Table 7.3 displays the response rates of the respondents of the Team Vier internet panel participating in the survey experiment. The unit response rate is unknown, since the survey was closed after the target of 250 respondents (for each subgroup) completing a variant was reached.

Table 7.3: Response Rates

Variant	Number of Respondents	Number of Items in Questionnaire			Average Item Nonresponse		
		Filter	Opinion	Follow-up	Filter	DK	Total
1A. No filter question, explicit DK	251		14	14	---	8.9%	8.9%
1B. No filter question, implicit DK	249		14	14	---	1.2%	1.2%
1C. No filter question, forced choice	250		14	14	---	---	---
2A. Filter question, explicit DK	251	13	14	14	16.3%	1.3%	18.4%
2B. Filter question, implicit DK	250	13	14	14	17.4%	0.3%	18.7%
2C. Filter question, forced choice	252	13	14	14	18.9%	---	18.9%
Total	1503						

Average Item Nonresponse Filter indicates the use of the 13 filter questions; Average Item Nonresponse DK shows the average use of the DK option; and Average Item Nonresponse Total combines both categories for item nonresponse.

The average item nonresponse rates[39] (see Table 7.3) indicate that it matters how non-substantive response options are offered. A 10-percentage point difference shows between all filter variants and the explicit DK variant in the level of average item nonresponse. Including a filter question clearly and significantly increases item nonresponse. Secondly, almost no use is made of the implicit DK option where respondents can skip the question if they do not want to or cannot answer. Thirdly, the average item nonresponse shows that the explicit DK option was rarely used after a filter question was asked: in variant 2A the percentage of item nonresponse registered by the DK option is 1.3 percent. There are only small differences between the variants using a filter question. Furthermore, the difference in total average item nonresponse between the filter variants 2A, 2B and 2C is statistically non-significant in a t-test.

As said, the focus here is on the follow-up or 'so what' question. How 'upset' are citizens if their opinion is ignored or does not prevail when the particular issue is decided? Furthermore, the use of a filter question and/or an (explicit) DK option for some versions shows whether and how item nonresponse is related to these differences. Table 7.4 shows the six versions of the questions/questionnaire (see Appendix AIII for the complete questionnaires).

Table 7.4: Experimental Conditions and Safety Nets for Item Nonresponse

	Explicit DK	Implicit DK	Forced choice
No filter	1A: answer DK	1B: skip question	1C: no safety net, answer obligatory
Weak filter	2A: 'no' to filter question or answer DK	2B: 'no' to filter question or skip question	2C: 'no' to filter question

All respondents who gave a substantive answer were asked a follow-up question: how upset would you be if the previously expressed opinion did not prevail?
In Dutch: 'Hoe erg zou u het vinden als uw mening over [beschrijving onderwerp voorgaande vraag] niet door de politiek in beleid wordt omgezet?'

---

39   The item nonresponse and distribution of opinions of this survey experiment are examined further in the next (comparative) chapter. Here, the aim is simply to set some parameters for the discussion in the current chapter.

An example of a particular question from this survey is given in Figure 7.1 (for the euthanasia self-placement item). The parts in italics vary according to the variant the respondent is subjected to. (For a complete list of the questions and variations, see Appendix AIII).

> Some people think that euthanasia should be forbidden by law. Others feel that a doctor should always be allowed to end a life, if the patient makes that request. Of course, there are people whose opinions lie somewhere in between. Suppose that the people (and parties) who think that euthanasia should be forbidden are at the beginning of this line (at number 1), and the people (and parties) who feel that a doctor should always be allowed to end a life upon a patient's request are at the end of the line (at number 7).
> *Do you have an opinion on this or not?*
> Where would you place yourself on a line from 1 to 7, where 1 means that euthanasia should be forbidden and 7 that a doctor should always be allowed to end a life upon a patient's request?
>
> Euthanasia should be forbidden.
> 2 - 6
> 7. A doctor should always be allowed to end a life upon a patient's request.
> *99 Don't Know*
> How upset would you be if the previously expressed opinion did not prevail when the issue was ultimately decided?

Figure 7.1: Survey Question Variations

PUBLIC OPINION WITHOUT OPINIONS?

7.5   Results

To what extent want survey respondents their opinion to matter and/or translated into policy? One indirect and implicit way to answer this question is to look at item nonresponse. Another, more direct approach is based on Moore's follow-up question. This follow-up question is intended to gauge or measure the intent of opinions: does the respondent want to see the opinion translated into policy? The opinions are distinguished in two categories: *directive*, in which case the respondent wishes or expects his/her given opinion to be executed, and *permissive* which "permits" the country's leaders to do with the information provided by the respondents whatever they deem best. Permissive opinions include item nonresponse; item nonresponse does not contribute to public opinion and gives the leaders the full discretion to decide themselves what should be done.

The follow-up question applied is reminiscent to the format that was used by Moore (2004, 2005) and Carlson (2002). Rather than asking 'Would you be upset…?' as was the case for most applications of the follow-up question, the question here is 'How upset would you be if the previously expressed opinion did not prevail when the issue was ultimately decided?'. The respondents could give an answer on a four-point scale without a non-substantive response option; very upset, upset, not very upset, not upset at all. The scale offers a wider range to answer the question instead of being forced to answer yes or no.

Two questions are addressed in the analysis: 1) What part of the public gives a directive opinion – in general and regarding specific subjects; and 2) How does this impact on the resulting picture of public opinion? The first question looks at the number of non-opinions resulting from non-substantive response options, i.e. the filter question and the DK option, and the number of 'not upset' permissive opinions. The second question looks at the overall distribution of opinions, both excluding the distinction between directive and permissive opinions and including such information in the full distribution.

7.5.1   Permissive and Directive Opinions

How many answers to survey questions are directive, i.e. opinions the respondents want to see translated into policies? And how is item nonresponse resulting from the various non-substantive response options related to the number of permissive and directive opinions? Previous research by Moore (2008) found that at least 40 percent of the respondents gave a non-directive (permissive) response, i.e. either a nonresponse or an answer indicating that the respondent would not be upset if the issue was decided otherwise. The aggregated results (in Figure 7.2) show a lower number: 61 to 68 percent of the respondents gave a directive opinion ('upset' or
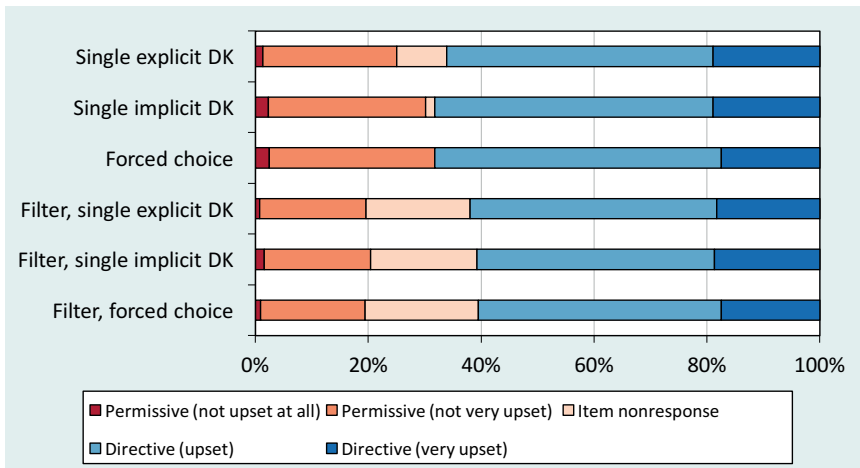
Figure 7.2: Average Rate Directive and Permissive Opinions (%)

'very upset') and the remaining 39 to 32 percent expressed a permissive opinion. The average rate of permissive opinions never exceeds 40 percent.

The first expectation differentiates between the two types of permissive opinions: item nonresponse and 'not (very) upset' answers to the follow-up question. It was expected that questions with more item nonresponse would result in less directive opinions, both including (E1a) and excluding (E1b) item nonresponse. When item nonresponse is included, the range of permissive opinions is 32 to 39 percent. This variation can primarily be attributed to item nonresponse, which varies between 0 (in non-filter forced choice variant 1c) and 20 percent (in filter forced choice variant 2c). The variants with the lowest item nonresponse (single implicit DK and forced choice) are also the variants with the lowest total percentage of permissive opinions (including item nonresponse) and consequently the highest percentage of directive opinions. The same is true for variants with high item nonresponse rates[40]: filter variants 2a, 2b and 2c result in the least directive opinions. The expectation that items with a higher item nonresponse rate result in less directive opinions when item nonresponse is included in the outcome (as formulated in expectation E1a) is supported.

---

40 Surprisingly, the variant with a filter question and an explicit DK option does not result in the highest item nonresponse rate. Variant 2c, which only offers a filter question as a nonresponse option, renders 20 percent item nonresponse (as compared to 18 percent in variant 2a). A possible explanation is that respondents of variant 2c found out that a filter question was their only way of not giving a substantive answer and therefore used the filter question sooner for subsequent survey questions (without looking at the actual opinion question).

Even without item nonresponse, the expectation was (E1b) that items with relatively high item nonresponse results in fewer '(very) upset' answers to the follow-up question. The average number of directive (upset) and permissive (not upset) opinions, excluding item nonresponse as missing data, is shown in Figure 7.3.
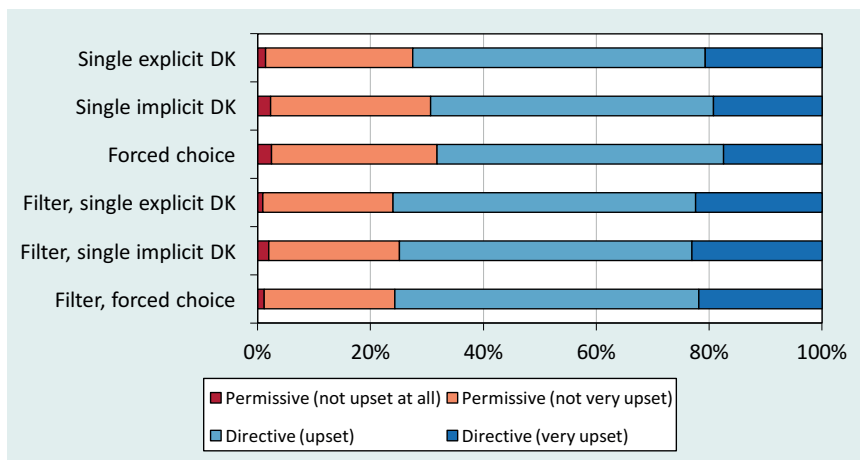

Figure 7.3: Average Rate Directive and Permissive Opinions (%) – Excluding Item Nonresponse

The question is: do variants with a higher average item nonresponse rate result in less directive opinions (as expected in E1b)? The answer is no. The pattern goes in the opposite direction: variants with the lowest average item nonresponse rates (forced choice and single implicit DK) result in 68 and 69 percent directive opinions whereas variants with the highest item nonresponse rates (the forced choice variants) result in 75 to 76 percent directive opinions (see Figure 7.3). The variants with a higher average item nonresponse rate thus result in *more* directive opinions; expectation E1b is rejected.

Table 7.5 shows the percentage of directive opinions for all individual survey questions[41], to move beyond general relations into an in-depth examination of whether respondents care about the answers for various topics they give in surveys.

---

41   The percentage of directive opinions consists of the 'very upset' and 'upset' answers to the question. Together with item nonresponse and other permissive opinions, i.e. 'not upset at all' or 'not very upset', these answers make up 100 percent. An additional analysis excluding item nonresponse can be found in , in order to examine the relation between item nonresponse and the number of directive opinions.

Table 7.5: Directive Opinions (%) Individual Survey Questions

Issue		1A: single explicit DK	1B: single implicit DK	1C: forced choice	2A: filter single explicit	2B: filter single implicit	2C: filter forced choice
Self-Placement Income Differences	Very Upset	16	13	15	15	16	17
	Upset	56	58	52	54	48	44
The welfare benefits should be lowered in order to stimulate people to work	Very Upset	19	20	16	18	19	15
	Upset	51	51	57	46	46	46
Self-Placement Euthanasia	Very Upset	28	27	31	31	34	32
	Upset	57	57	52	50	48	47
Adoption by same-sex couples should be possible	Very Upset	16	17	17	17	19	17
	Upset	48	48	48	42	38	37
Self-Placement Foreigners	Very Upset	19	20	14	18	16	15
	Upset	57	55	60	55	55	55
There are too many people of a non-Dutch nationality living in the Netherlands	Very Upset	12	15	14	16	15	16
	Upset	50	50	50	46	46	48
Self-Placement EU	Very Upset	12	15	14	16	15	16
	Upset	50	50	50	46	46	48
The Netherlands should spend more money on developmental aid	Very Upset	18	20	18	22	22	18
	Upset	47	44	52	46	44	52
The use of softdrugs should be completely prohibited	Very Upset	19	18	14	14	12	12
	Upset	39	47	47	36	38	33
'Establishing a 'Weed Permit' is a good idea'	Very Upset	14	15	10	10	11	12
	Upset	34	44	46	32	31	33
Powers EU	Very Upset	15	13	12	10	12	12
	Upset	45	53	57	45	40	48
'The Nederlands should abolish the mortgage interest deduction completely'	Very Upset	31	32	29	31	33	25
	Upset	44	41	47	42	40	44
'I want the Burqa Ban to proceed'	Very Upset	27	26	27	26	26	25
	Upset	44	40	38	38	36	37
'NATO should intervene in Syria'	Very Upset	13	12	14	12	8	13
	Upset	39	49	51	31	31	33
N		251	249	250	251	250	252

7

As said, it was found previously that the number of permissive opinions would amount to at least 40 percent, the remaining 60 percent or less of the respondents giving a directive opinion. This proves to be incorrect for the individual replicated scientific questions: the lowest number of directive opinions was 62 percent (variant 1A; Nationality), 65 percent in 1B and 64 percent in 1C. In the variants where a filter question was used (2A, 2B and 2C) the numbers are slightly lower, with the lowest number of directive opinions between 54 and 59 percent. All other survey questions generate over 60 percent 'upset' and 'very upset' directive opinions.

The results for the current affairs questions are more mixed. 73 percent of the respondents would on average be (very) upset if their opinion about mortgage interest deduction did not prevail when the issue was to be decided. The weed permit question, on the other hand, addresses an issue for which only a minority (in variants 1A, 2A, 2B and 2C) would be (very) upset if the outcome was different than their personal preferred option. The weed permit was an attempt (in 2012) to regulate the admission to 'coffeeshops' in the Netherlands to prevent problems caused by drug tourism; only people with a Dutch passport who had applied for a permit would be allowed admission. The weed permit was abolished after a couple of months when problems with illegal trade were discovered (NOS, 2012; Willems, 2012). It turns out that for four variants this item results in relatively few directive opinions. In the other two variants, the issue with the least directive opinions is related to NATO intervention in Syria.

These results reveal some differences in the percentage of directive opinions for individual items; the influence of the content of the questions on the percentage of directive opinions is examined below. The results in Table 7.5 do show that at least for some issues a majority does not express a directive opinion. This may have consequences for the overall picture of public opinion (as examined in the next section). A final point to notice is that the 'very upset' response is usually used by less than 20 percent of the respondents; only the questions about euthanasia, mortgage interest deduction and a Burqa ban show 25 to 32 percent 'very upset' directive opinions. Whether this can be attributed to question content will be discussed in the next section.

7.5.2   Directive Opinions and Question Content

Does the content of the individual survey questions matter? Several expectations were formulated based on whether a survey question appeals to the core dimensions of Dutch politics, in which case the number of directive opinions was expected to be relatively high, and whether an issue was related to foreign policy, in which case the number of directive opinions was expected to be lower. Table 7.6 shows the percentage of directive opinions for all questions grouped by content or theme.

Table 7.6: Directive Opinions (%) and Question Content

Theme	Question subject	1A: single explicit DK	1B: single implicit DK	1C: forced choice	2A: filter single explicit	2B: filter single implicit	2C: filter forced choice
Socio-economic	Self-placement income	72	71	67	69	64	61
	Welfare benefits	70	71	73	64	65	61
	Mortgage interest deduction	75	73	76	73	73	69
Ethical	Self-placement euthanasia	85	84	83	81	82	79
	Adoption same-sex	64	65	65	59	57	54
	Softdrugs	58	65	61	50	50	45
	Weed permit	48	59	56	42	42	45
Multicultural	Self-placement foreigners	76	75	74	73	71	70
	Nationality	62	65	64	62	61	64
	Burqa ban	71	66	65	64	62	62
Foreign policy	Self-placement EU	62	65	64	62	61	64
	Development aid	65	64	70	68	66	70
	Powers EU	60	66	69	55	52	60
	NATO in Syria	52	61	65	43	39	46

For most questions related to a major political dimension, the number of directive opinions is relatively high. The abolishment of the mortgage interest deduction (in the socio-economic sphere), for instance, has 69 to 76 percent of the respondents saying they would be (very) upset if their opinion did not prevail. With respect to euthanasia, 79 to 85 percent (depending on the questionnaire variant) express a directive opinion. These results support expectation E2a, but not all items related to a major political dimension show an equal high rate of directive opinions. The 'weed permit' is an item with relatively few directive opinions; in some variants only a minority of the respondents indicate they would be (very) upset if their opinion did not prevail.

The second expectation regarding question content (E2b) was that questions about foreign policy issues would render a relatively low number of directive opinions, because opinions about such issues are supposed to be less coherent, less stable and less informed. The four foreign policy questions (about the EU, development aid, the powers of the EU and the role of NATO in Syria) do result in relatively few directive opinions. These questions are all among the 7 (out of 14) questions with the lowest number of directive opinions. NATO's role in Syria is the foreign policy item with the least directive opinions: 39 to 65 percent, depending on

the questionnaire variant, which is the lowest percentage of directive opinions in two variants. So the general trend suggests a relation between issue content and the number of directive opinions, with questions related to a major political dimension resulting in more and questions related to foreign policy resulting in less directive opinions, as expected in E2a and E2b.

How does question design affect the number of directive opinions? Since item nonresponse is considered a permissive opinion and item nonresponse is encouraged by offering explicit non-substantive response options, it is to be expected that variants with explicit non-substantive response options - i.e. a filter question or an explicit DK category – result in more permissive opinions. The number of permissive opinions is consistently higher for the variants including a filter question as compared to non-filter variants. Including an explicit DK option as a response category (in variant 1A), however, renders a mixed effect on permissive opinions; questions about foreign policy issues generally do result in more permissive opinions when an explicit DK option is offered as compared to the implicit DK and forced choice variants, but the same cannot be said for the other items. Furthermore, this effect for the foreign policy items does not show in the filter variants. So the use of a filter question does affect the number of permissive opinions for all issue themes, but no definite conclusions can be drawn about the effect of a DK option on the number of permissive opinions. The likely explanation for this conclusion is that the variant with an explicit DK option (1A) resulted in an average item nonresponse rate of 9 percent, whereas the filter variants (2A, 2B and 2C) resulted in 18 to 20 percent item nonresponse. Only the respondents with a substantive answer were asked the follow-up question and this means that less respondents of the filter variants answered the follow-up questions.

7.5.3   Distribution of Opinions – Towards Public Opinion

If a politician looks at a poll, what would s/he make of it? Acknowledging the possibility that some respondents do not report a directive opinion, which majority or plurality emerges? The question is whether majorities or pluralities change or disappear when a follow-up question is used by making a distinction between directive and permissive opinions.

The follow-up question could affect the distribution of opinions in two ways: 1) the substantive outcome preferred by a majority or plurality may change when only directive opinions are considered, resulting in a different overall picture of public opinion; and 2) the size of the majority changes and the public seems more divided (but the preferred outcome does not change). In fact, both effects occur, in varying degrees. All distributions of opinions can be found in Appendix C.

To start with the first effect: do majorities or pluralities change? For some items

in some variants the response option preferred by a majority changes to a different response option preferred by a plurality when permissive opinions are excluded. Figure 7.4 illustrates this point with the distribution of opinions for NATO in Syria. If no distinction is made between permissive and directive opinions, a majority of 54 percent of the forced choice respondents disagrees with the statement that NATO should intervene in Syria (displayed in Figure 7.4 as light blue and dark blue segments). When the permissive opinions are excluded, however, more respondents agree with the statement; 35 percent gave a directive 'agree' answer compared to 30 percent directive *disagreeing* opinions. This is an example of a changing picture of public opinion: the substantive outcome preferred by most respondents – either a majority or a plurality – changes when only directive opinions are considered as the relevant public opinion.
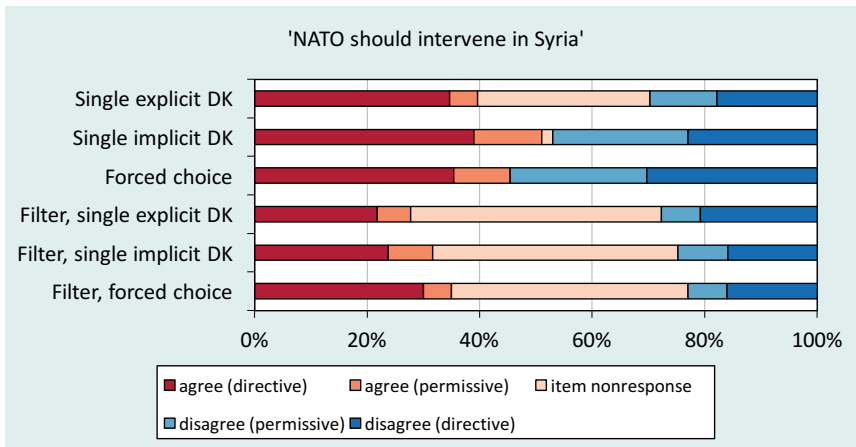

Figure 7.4: Distribution of Opinions NATO in Syria

The changing majority or plurality is arguably the effect of a follow-up question which has the most consequences for the outcome, because the majority preference is different when the distinction between permissive and directive opinions is made. The change from an option preferred by a majority to a different option preferred by a plurality happens on several occasions: once for 'lowering welfare benefits' (variant 1c), twice for 'softdrugs' (variant 1b and 1c), twice for the 'powers of the EU' (variant 1a and 2b) and once for the NATO in Syria question (see Figure 7.4). Most distributions, however, do not show a different outcome of what the public wants.

    The second effect of the follow-up question can be a disappearing majority; more than 50 percent of the respondents used a particular response option, but

after excluding the permissive opinions only a minority remains. The preferred outcome, however, does not change; only the size of the majority diminishes. This happens fairly often: for most items the majority agreeing or disagreeing with a statement becomes a plurality when ('not upset') permissive opinions are excluded. From a total of ten survey questions (in six questionnaire variants), only three show a majority expressing a particular directive preferred outcome: 'developmental aid' (in all variants), 'mortgage interest deduction' (in five variants) and 'burqa ban' (in all variants). Figure 7.5 shows that a majority of the respondents of all subgroups disagreed with the statement that there are too many people of a non-Dutch nationality living in the Netherlands, but the distinction between permissive and directive opinions results in a plurality giving a directive disagreeing answer. The opposite but less common effect is present in Figure 7.6: even when permissive answers are classified as non-substantive answers, a majority remains supporting the burqa ban.
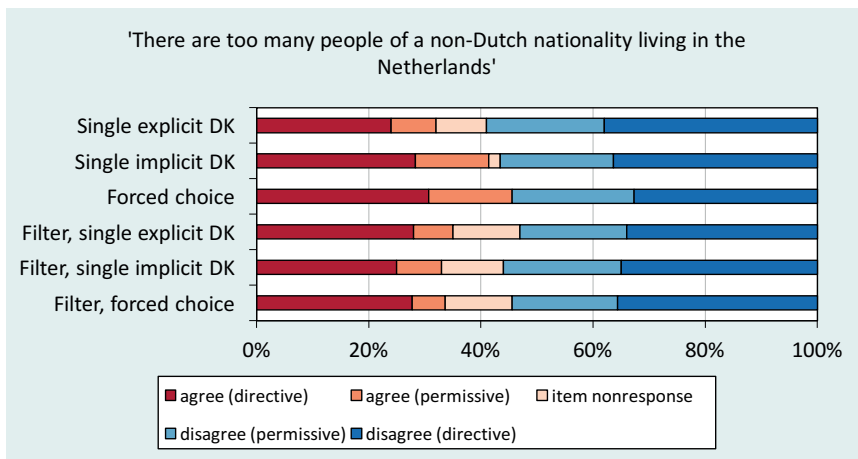

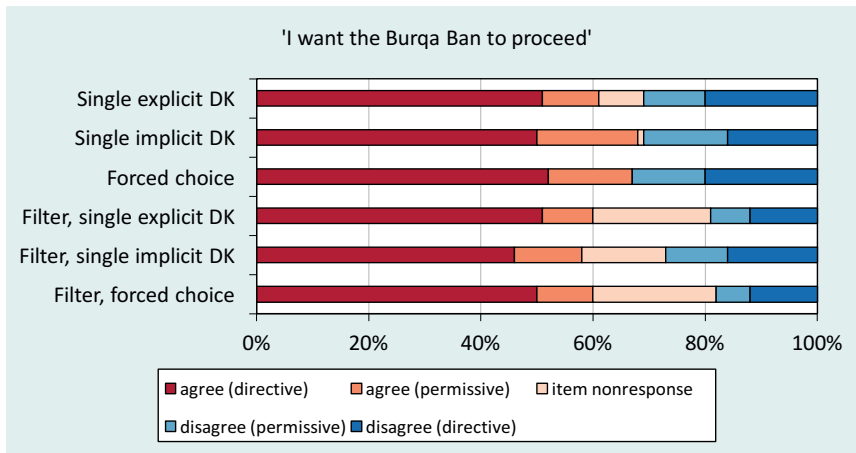*Figure 7.5: Distribution of Opinions Non-Dutch Nationality*

Figure 7.6: Distribution of Opinions Burqa Ban

The disappearing majority resulting from the follow-up question suggests that in many cases the overall picture of public opinion is less clear, or at least that the preferred option is not strongly supported by over half of the public. In some cases, the majority may not even have a directive opinion. The main point here is regarding the full distribution. Often no majority remains when permissive opinions are excluded as substantive answers; remaining majorities are at best small.

To summarize: the use of a follow-up question affects the substantive distributions of opinions in several ways. The majority or plurality supporting a certain outcome changes, disappears or diminishes substantially. The use of a follow-up question adds information about what individual citizens want to happen with their opinion; the answer is that they often do not really care. And if they do not care, why should others?

7.6   Conclusion

In this chapter, the focus was on the way reported substantive answers may affect political decision-making. More specially, it was about the extent to which citizens want politicians to be responsive. The analysis was structured by the question whether respondents considered their opinions to be directive or not. The experiment was carried out by the Team Vier internet panel.

The first part of the analysis focused on directive opinions, i.e. opinions that the respondent would be upset about if the issue was decided otherwise. Almost

all items had a majority reporting a directive opinion, with the exception of some current affairs questions, and the percentage often exceeded 60 percent. This finding is different from Moore, who concluded for the US that most issues rendered at least 40 percent permissive opinions and consequently at most 60 percent directive opinions. Overall our Dutch respondents indicated far more often that they would be (very) upset if the policy outcome did not reflect their preferred position. The results did, however, show a relation between item nonresponse and directive opinions (supporting expectation E1a): the average percentage of directive opinions was strongly correlated to the average percentage of respondents saying they would not be upset if their opinion did not prevail.

The expectation (E2a) that questions about issues related to a major political dimension – socio-economic, ethical and multicultural – would result in more directive opinions compared to questions not related to these dimensions was supported for most items. Also, survey questions relating to foreign policy issues in general resulted in less directive opinions, as expected (E2b).

Since surveys and polls are often used by politicians as a means to assess what the general public wants, their outcome can be a relevant or even crucial part of the public and political debate. This is what makes the inclusion of the directive – permissive distinction important: if a large majority or plurality does not really care either way, should politicians follow them? And although Dutch citizens expressed less permissive opinions than Moore's Americans, a substantial part of the Dutch public (about 35 percent) did not give an answer to a survey question indicating that they actually wanted to see their opinion carried out in the decision-making. The analysis of the distributions of opinions strongly suggested that the overall picture that emerges may indeed be different in terms of majorities and pluralities. Most importantly: often only a small minority holds a strong, directive position.

The use of a follow-up question affected the outcome in a number of ways. First, for some items the outcome preferred by a majority or plurality changed, resulting in a different overall picture of public opinion. Secondly, various items did not result in a response option preferred by a majority when all permissive opinions, i.e. item nonresponse and 'not upset' answers to the follow-up question, were excluded. And finally, for some items the size of the majority changed and the public seemed more divided. The second effect of disappearing majorities was prevalent, but all three effects strongly suggest that the use of a follow-up question results in a different picture of public opinion, with a substantial part of the public that does not deeply seem to care about the particular policy outcome.

Painting a valid picture of public opinion is complex and even more complex than simply including non-substantive response options. It can be argued that only a small part of the public at large has an opinion that should be considered as

part of the decision-making process. Simply measuring whether people agree or disagree with a statement is not enough to use the opinions of individual citizens in the decision-making process and treat their opinions as serious expressions of a preferred policy outcome.

7

Comparing Survey Experiments (Conclusions I)

8.1   Introduction

The previous chapters showed that non-substantive response options – i.e. the explicit DK option, the filter question and the follow-up question – affect the picture of public opinion emerging from surveys by increasing item nonresponse rates (the DK option and filter question) or making a distinction between respondents with a directive and permissive opinion (the follow-up question). Surprisingly, the substantive overall distributions of opinions were rarely altered by question design choices, suggesting that item nonresponse is missing at random.

What has not been addressed so far is whether the results of the separate experiments are consistent with one another. The aim of this chapter is to compare the three experiments on item nonresponse and the overall distribution of opinions. The main question is what the impact of a Don't Know option or filter question is on the outcome of a survey and whether this impact is consistent in all experiments. Some question design elements – i.e. explicit and implicit DK options and the filter question – were used in multiple experiments, enabling such a comparative analysis. If this analysis shows similar effects, more robust conclusions can be drawn about how non-substantive response options affects survey outcomes.

8.2   Theoretical Reflection

A number of factors play a role in influencing the results of a survey, including question design. This process is graphically illustrated in Figure 8.1.
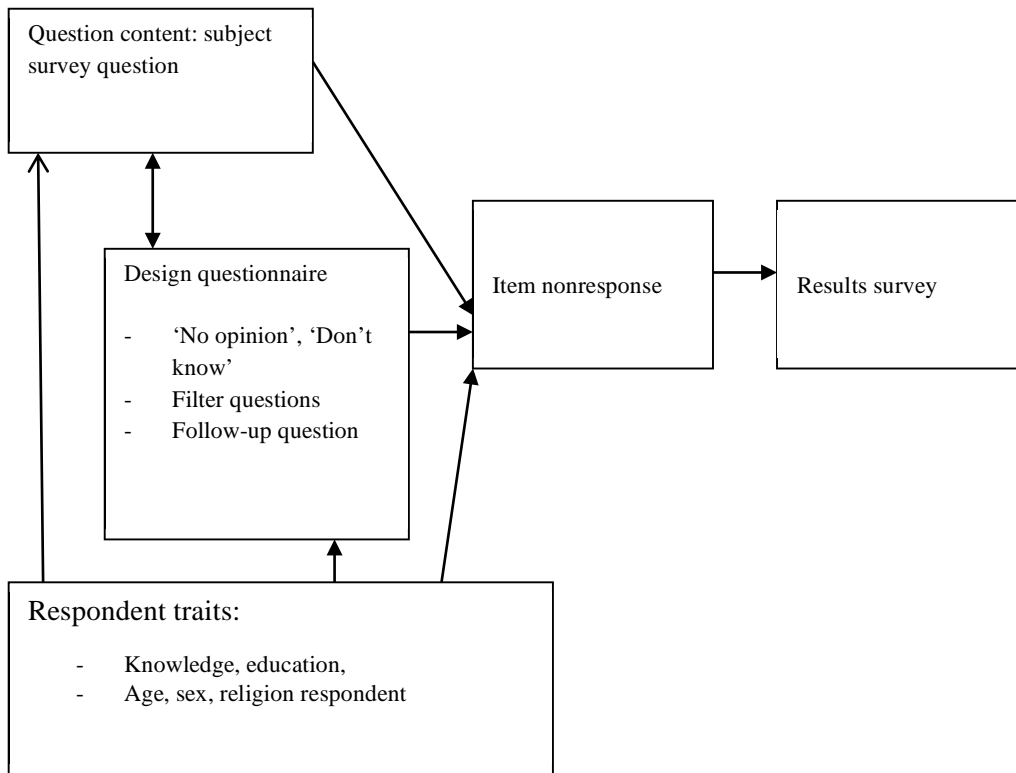
8

Figure 8.1: Model for Non-Substantive Response Options and Survey Outcomes

The model positions the analysis from the previous chapters and illustrates which variables and effects were central in the discussion so far. It does not contain an exhaustive account of all methodological aspects affecting survey results; the intention is merely to illustrate how the elements examined in this chapter (i.e. non-substantive response options and survey results) may interact. The model shows that both the design of the questionnaire and the subject of the survey question(s) affect item nonresponse. At the same time, an interaction effect was expected: the content of a question affects the susceptibility of a respondent to design effects.

8.3   Hypotheses

The hypotheses for the comparative analyses focus on three aspects: 1) item nonresponse; 2) the distribution of opinions; and 3) question content and its interaction with the other two aspects. The hypotheses are summarized in Table 8.1. See chapter 3 for a more extensive discussion.

Table 8.1: Hypotheses for Comparative Analyses

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

8.4 Data and Methods

The design of the survey experiments was discussed in Chapter 4 (Data and Methods). Details on the execution of the individual survey experiments can be found in Chapter 5, 6 and 7. Three internet panels were used to collect data: the probability-based LISS Panel and two volunteer or convenience panels (EenVandaag Opiniepanel and Team Vier's internet panel). Table 8.2 gives an overview of the questionnaire variants and how many respondents completed that particular variant in the three internet panels.

Table 8.2: Overview of Comparison Questionnaire Variants and Panels

		Number of respondents
Explicit Don't Know Option		
LISS Panel	Variant 2 (Single Explicit DK)	1464
EenVandaag Opiniepanel	Variant 3a (No Filter, Explicit DK)	3931
Team Vier internet panel	Variant 1a (No Filter, Explicit DK)	251
Implicit Don't Know Option – Possibility to Skip Question		
LISS Panel	Variant 3 (Single Implicit DK)	1375
EenVandaag Opiniepanel	Variant 3b (No Filter, Implicit DK)	3591
Team Vier internet panel	Variant 1b (No Filter, Implicit DK)	249
(Weak) Filter Question, Explicit Don't Know Option		
EenVandaag Opiniepanel	Variant 2a (Weak Filter, Explicit DK)	4327
Team Vier internet panel	Variant 2a (Weak Filter, Explicit DK)	251
(Weak) Filter Question, Implicit Don't Know Option		
EenVandaag Opiniepanel	Variant 2b (Weak Filter, Implicit DK)	4283
Team Vier internet panel	Variant 2b (Weak Filter, Implicit DK)	250
Forced Choice – No Non-substantive response options Offered		
LISS Panel	Variant 4 (No DK / Forced Choice)	1421
EenVandaag Opiniepanel	Variant 4 (Forced Choice)	4460
Team Vier internet panel	Variant 1c (Forced Choice)	252

The variants differ in the number of respondents, sample composition and their timing, but otherwise they are identical or at least equivalent; for example, the three surveys with the Explicit Don't Know Option used the same introduction, question order and answer categories. The questions that allow for comparison are the eight replicated scientific questions that were included in all three experiments. These questions cover four policy fields or themes: socio-economic, ethical, multicultural and foreign affairs. In the first experiment, executed with

the LISS Panel, four additional scientific questions were included. These questions were asked in between the other eight questions and another example of the four issue themes. Furthermore, in the third experiment (executed with the Team Vier internet panel) the order of the questionnaire was randomized. Question order is not held completely constant.

The aim of the survey experiments was to identify causal relations. The between-subjects-design with random assignment to subgroups ensured that question design was the only factor affecting survey results. The experiments have a high internal validity.

It should be noted, however, that the external validity of the two volunteer or convenience samples is very limited. Generalization to the population is problematic. The ability to compare the findings of the three survey experiments is impaired by the unrepresentative nature of the EenVandaag Opiniepanel and Team Vier's internet panel. To give an example: 31 percent of the EenVandaag Opiniepanel respondents and 64 percent of the Team Vier internet panel respondents was female. A comparison of the three experiments may reinforce the findings of the individual chapters, but differences between the experiments and panels cannot be ruled out beforehand. Since the differences between the panels cannot be excluded as a potential explanation, any conclusions about differences between the three individual survey experiments need to be drawn very tentatively.

8.5   Results

8.5.1   Item Nonresponse
The first main question is how non-substantive response options affect item nonresponse. The expectation (hypothesis H1a) was that more explicit non-substantive response options result in more item nonresponse. Figure 8.2 displays the average item nonresponse for the variants in the three survey experiments[42]. The item nonresponse results from a *single* non-substantive response option, i.e. a DK option or a filter question, and not from the combination of both a filter question and a DK option in one variant.

---

42   The average item nonresponse was computed by counting either the DK answers or the 'no's' to the filter question, and dividing this sum by the total number of questions in the questionnaire. See   for an additional analysis of the total item nonresponse resulting from all non-substantive response options offered in one variant. Since no DK option was offered in the Forced Choice variants, these are excluded from this analysis.
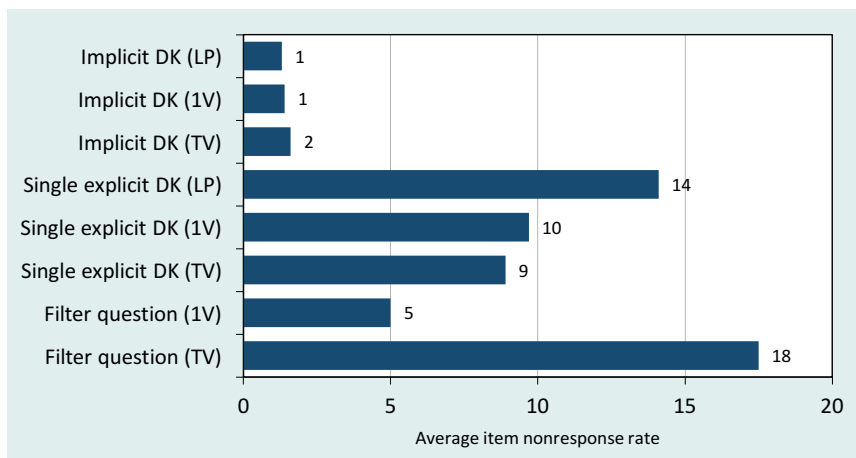
Figure 8.2: Average Item Nonresponse Rate (%) of Individual Non-Substantive Response Options

The first hypothesis (H1a) stated that more explicit non-substantive response options result in more item nonresponse. The least explicit non-substantive response option in this analysis is the implicit DK option, the most explicit option is the filter question[43]. Figure 8.2 shows mixed results: the filter question in the experiment with the Team Vier internet panel resulted in the highest item nonresponse rate (17 percent) and the implicit DK options indeed yielded the least item nonresponse (1 to 2 percent), but there are exceptions to the rule that more explicit non-substantive response options result in more item nonresponse. The single explicit DK option in the experiment with the LISS panel, for example, resulted in 14 percent of the respondents using the DK option and this was substantially higher than the 5 percent saying 'no' to the more explicit filter question in the experiment in the EenVandaag Opiniepanel. In fact, all explicit DK options resulted in more item nonresponse than the filter question in this experiment.

A possible explanation of these mixed results is that respondents may have learned during the survey that there were two non-substantive response options available: when they said 'yes' to the filter question, they were still able to use the DK option. An additional analysis of the average item nonresponse rate including both non-substantive response options does indeed show that filter variants, i.e. questionnaire variants including both a filter question and a DK option, have the highest total item nonresponse of all questionnaire variants except for one (the explicit DK variant executed with the LISS panel in the first experiment – see Figure C.62 in the Appendix.

---

43   Only variants that were used at least twice by two different panels are included.

Looking at the average item nonresponse rates in Figure 8.2, another thing catches the eye: the differences within the single explicit DK and within the filter question variants. Depending on the survey experiment and panel, the item nonresponse for the single explicit DK variants varied between 9 and 14 percent; the item nonresponse for the filter question variants varied between 5 and 18 percent. This range is rather large.

A number of potential explanations can be given for the differences between the three experiments: panel composition, the standard practice of the panels and incentives. The first explanation, panel composition, stems from the fact that both the EenVandaag Opiniepanel and Team Vier's internet panel are non-probability based samples resulting from self-selection. Only the LISS Panel is probability-based. As a result, the outcome of the second (filter question) and third (follow-up question) experiment in this study cannot be generalized to the population. Furthermore, comparisons between the three panels are difficult because specific characteristics of the panel respondents may have contributed to the survey outcomes. Due to the unrepresentative nature of two panels and their subsequent incomparability with the third (LISS) panel, no conclusions can be drawn about the differences in item nonresponse rate within the explicit DK and within the filter question variants.

The second potential explanation concerns the panels' standard practice. Do they offer non-substantive response options to the respondents and if so, how? The respondents of the EenVandaag Opiniepanel usually get an explicit DK option and may as such be more used to the possibility to use it than the respondents of the LISS Panel. This standard practice could explain why the respondents of the former used the explicit DK option less than respondents of the latter. It does not explain, however, why Team Vier's respondents said 'no' to the filter question more often than EenVandaag Opiniepanel's respondents.

The third potential explanation is the incentive offered to the respondents for participating in the survey. Respondents of the LISS Panel and Team Vier's internet panel are awarded money for finishing the survey, whereas EenVandaag Opiniepanel's respondents 'only' get the satisfaction of seeing the results of the survey in EenVandaag's television programme. It might be the case that Team Vier's respondents used the filter question more often than EenVandaag Opiniepanel's respondents, because they wanted to finish the survey and collect their reward. Likewise, respondents of the LISS Panel may have used the explicit DK option more to avoid having to think long about a question and finish the survey quickly. It is, however, mere speculation which factor explains the differences in survey outcomes, since the factors cannot be isolated. EenVandaag's Opiniepanel differs on all factors from the LISS Panel, for example, which means that it is impossible to say which factor explains the differences in item nonresponse best.

The results in Figure 8.2 do confirm that more explicit non-substantive response options on average resulted in more item nonresponse, which supports expectation H1a. Also, the use of a filter question resulted in more item nonresponse than the use of a DK option in one of the two experiments. The nonresponse is not as high as found in previous research on filter questions (Bishop et al., 1983; Schuman & Presser, 1979), but the results partially support hypothesis H1b that filter questions result in more item nonresponse than an explicit DK option when the effect of a single non-substantive response option is considered; when all non-substantive response options are included the total item nonresponse is consistently higher in the filter variants. Offering a non-substantive response option more explicitly results in more item nonresponse; the size of this effect varies.

8.5.2 Distribution of Opinions

A higher nonresponse rate 'provides greater opportunities for bias' (Lynn, 2014, p. 319), but whether this actually happens is dependent on whether the data are missing at random or not. If the data is missing at random, the distribution of opinions should be similar in variants with varying item nonresponse. If the data are not missing at random, the distribution of opinions should be significantly different when nonresponse is excluded. Table 8.3 and Table 8.4 show some characteristics of the distribution of opinions for the self-placement items. Two things should be assessed on the basis of these data: 1) whether the data are missing at random; and 2) whether the panels and variants differ in their outcome for the same survey question.

Table 8.3: Mean, Mode and Standard Deviation of Self-Placement Items in the Non-Filter Variants

Question		Explicit DK			Implicit DK		
		LP	**1V**	**TV**	**LP**	**1V**	**TV**
Income Differences	Mean	5.3	5.1	5.2	5.0	5.1	5.1
	Mode	5	7	4	5	7	4
	SD	1.3	1.5	1.4	1.3	1.6	1.3
Euthanasia	Mean	5.8	6.0	5.9	5.7	6.0	6.0
	Mode	7	7	7	7	7	7
	SD	1.5	1.5	1.5	1.5	1.4	1.4
Foreigners	Mean	5.1	5.0	4.9	5.0	4.9	4.9
	Mode	5	7	5	5	5	5
	SD	1.4	1.6	1.3	1.4	1.6	1.5
European Unification	Mean	4.8	4.4	4.5	4.6	4.4	4.6
	Mode	4	7	4	4	7	4
	SD	1.7	2.2	1.8	1.6	2.1	1.7

LP = LISS Panel; 1V = EenVandaag Opiniepanel; TV = Team Vier's internet panel

Table 8.4: Mean, Mode and Standard Deviation of Self-Placement Items in the Filter Variants

Question		Weak Filter, Explicit DK		Weak Filter, Implicit DK		Forced Choice		
		1V	TV	1V	TV	LP	1V	TV
Income Differences	Mean	5.2	5.2	5.2	5.0	5.1	5.1	5.1
	Mode	7	5	7	4	4	7	4
	SD	1.6	1.4	1.5	1.5	1.3	1.5	1.3
Euthanasia	Mean	6.1	6.2	6.0	6.0	5.7	6.0	5.9
	Mode	7	7	7	7	7	7	7
	SD	1.4	1.2	1.4	1.5	1.5	1.4	1.5
Foreigners	Mean	5.0	5.0	4.9	4.9	5.1	4.9	4.9
	Mode	5	5	5	5	6	5	5
	SD	1.6	1.4	1.6	1.4	1.3	1.6	1.4
European Unification	Mean	4.4	4.5	4.1	4.7	4.7	4.4	4.5
	Mode	7	7	7	5	4	7	4
	SD	2.1	1.9	2.2	1.8	1.6	2.1	1.8

LP = LISS Panel; 1V = EenVandaag Opiniepanel; TV = Team Vier's internet panel

To examine whether the missing data are missing at random, both means and modes are considered. The means do not show much range; the biggest difference between the lowest and highest means is for European Unification with an average of 4.1 in the weak filter, implicit DK EenVandaag Opiniepanel variant and 4.8 in the explicit DK LISS Panel variant. Looking at the modes, the most stable item is self-placement euthanasia; all variants and panels show a mode of '7', i.e. the most liberal stance on euthanasia. The foreigners self-placement item is also rather stable with two exceptions from the general mode of '5'. The other two self-placement items, income differences and European Unification, show the most dissimilarity with '4' and '7' as the answer reported most. Overall, however, the differences are very small and when the complete distributions of opinions are regarded (as displayed in Chapter 5, 6 and 7), the picture of public opinion hardly changes when a different (or no) non-substantive response option is offered. This suggests that the data are missing at random, which supports hypothesis H2a.

Item nonresponse for the self-placement items was relatively low, lowering the probability for a major change in the emerging overall picture of public opinion. The other four items varied more in terms of item nonresponse; hence the possibility for data not missing at random, significantly affecting the distribution of opinions. Table 8.5 depicts the modes of the four non-self-placement items.

Table 8.5: Modes of the Non-Self-Placement Items

Question	Explicit DK			Implicit DK			Filter, Explicit DK		Filter, Implicit DK		Forced Choice		
	LP	1V	TV	LP	1V	TV	1V	TV	1V	TV	LP	1V	TV
Welfare benefits should be lowered in order to stimulate people to work	3	3	3	3	3	3	3	3	3	2+3	3	3	3
Adoption by same-sex couples should be possible	2	2	2	2	2	2	2	2	2	1	2	2	2
There are too many people of a non-Dutch nationality living in the Netherlands	3	3	3	2	3	3	3	3	3	3	2	3	3
The Netherlands should spend more money on development aid	3	4	3	3	4	3	4	3	4	3	3	4	3

LP = LISS Panel; 1V = EenVandaag Opiniepanel; TV = Team Vier's internet panel

The items referred to in Table 8.5 included four response categories – completely agree, agree, disagree and completely disagree – which makes them unsuitable for comparing means. Based on the modes, one item ('lowering welfare benefits' in the socio-economic theme) shows no variation at all. Another item, same-sex adoption, has one 'exceptional' mode; the nationality item within the multicultural theme has two exceptions. So overall, it does not really matter which panel was used or how the questions were asked: the overall outcome would be quite similar. This conclusion is supported by the complete distributions of opinions (as displayed in previous chapters). When item nonresponse is excluded as missing data, the picture of public opinion remains the same. This suggests that the data are missing at random (supporting hypothesis H2a).

8.5.3 Differences in Question Content and Panels

How do non-substantive response options affect the outcome of individual questions? The goal of this section is to go beyond the aggregate effect of question design and to look at differences in question content and internet panels. The analysis focuses on the effects on item nonresponse, because it turned out (above and in previous chapters) that non-substantive response options do not have a substantial impact on the outcome in terms of the overall distribution of opinions (when nonresponse is excluded as missing data). To examine in more detail how question design affects item nonresponse, how this is related to question content and whether the panels differ in their outcome, the levels of item nonresponse for the individual eight items are shown in Table 8.6 – Table 8.9.

Table 8.6: Item Nonresponse (%) in the Implicit DK Variants

Question	LP	1V	TV
Self-placement [1-7] Income Differences	0	1	0
Welfare benefits should be lowered in order to stimulate people to work	1	1	1
Self-placement [1-7] Euthanasia	1	0	1
Adoption by same-sex couples should be possible	1	1	1
Self-placement [1-7] Foreigners	0	0	0
There are too many people of a non-Dutch nationality living in the Netherlands	1	1	2
Self-placement [1-7] European Unification	1	0	2
The Netherlands should spend more money on development aid	1	1	2

LP = LISS Panel; 1V = EenVandaag Opiniepanel; TV = Team Vier's internet panel

Table 8.7: Item Nonresponse (%) in the Explicit DK Variants

Question	LP	1V	TV
Self-placement [1-7] Income Differences	5	1	2
Welfare benefits should be lowered in order to stimulate people to work	11	4	8
Self-placement [1-7] Euthanasia	4	1	1
Adoption by same-sex couples should be possible	11	5	4
Self-placement [1-7] Foreigners	2	0	1
There are too many people of a non-Dutch nationality living in the Netherlands	12	4	9
Self-placement [1-7] European Unification	10	2	4
The Netherlands should spend more money on development aid	12	5	6

LP = LISS Panel; 1V = EenVandaag Opiniepanel; TV = Team Vier's internet panel

Table 8.8: Item Nonresponse (%) in the Weak Filter, Implicit DK Variants – Total Item Nonresponse versus Filtered Out

Question	Total Item Nonresponse		Filtered Out	
	1V	TV	1V	TV
Self-placement [1-7] Income Differences	10	16	8	16
Welfare benefits should be lowered in order to stimulate people to work	8	15	7	14
Self-placement [1-7] Euthanasia	3	5	2	5
Adoption by same-sex couples should be possible	8	16	7	16
Self-placement [1-7] Foreigners	2	9	2	9
There are too many people of a non-Dutch nationality living in the Netherlands	3	11	2	11
Self-placement [1-7] European Unification	5	18	4	18
The Netherlands should spend more money on development aid	4	10	3	9

LP = LISS Panel; 1V = EenVandaag Opiniepanel; TV = Team Vier's internet panel

Total Item Nonresponse is measured as a percentage of the total number of respondents not responding to a certain opinion question by saying 'no' to the preceding filter question or by using the implicit DK option and skipping the question. The Filtered Out category only includes the respondents saying 'no' to the filter question, which consequently excludes them from the following opinion question. This category does not include DK answers.

Table 8.9: Item Nonresponse (%) in the Weak Filter, Explicit DK Variants – Total Item Nonresponse versus Filtered Out

Question	Total Item Nonresponse		Filtered Out	
	1V	TV	1V	TV
Self-placement [1-7] Income Differences	8	12	8	12
Welfare benefits should be lowered in order to stimulate people to work	10	16	7	15
Self-placement [1-7] Euthanasia	2	6	2	6
Adoption by same-sex couples should be possible	8	13	7	12
Self-placement [1-7] Foreigners	1	8	1	8
There are too many people of a non-Dutch nationality living in the Netherlands	4	12	2	10
Self-placement [1-7] European Unification	4	17	3	16
The Netherlands should spend more money on development aid	5	9	2	9

LP = LISS Panel; 1V = EenVandaag Opiniepanel; TV = Team Vier's internet panel

Total Item Nonresponse is measured as a percentage of the total number of respondents not responding to a certain opinion question by saying 'no' to the preceding filter question or by using the explicit DK option. The Filtered Out category only includes the respondents saying 'no' to the filter question, which consequently excludes them from the following opinion question. This category does not include DK answers.

The implicit DK option is rarely used: for all survey questions in all experiments included in the analysis (see Table 8.6), between 0 and 2 percent used the option to skip a question. This leaves little room for variation between internet panels or, more importantly, between questions with a different content. The implicit DK option is therefore excluded from the discussion below about the susceptibility of individual survey questions to design effects; no relation between offering an implicit DK option and question content can be discerned.

For the explicit DK option (Table 8.7), the main finding is that self-placement items result in less item nonresponse than the other items related to the same theme. A possible explanation is that self-placement items contain more substantive response categories (7 compared to 4 for non-self-placement questions), including a neutral response category, and that it is easier for respondents to pick a response category. Furthermore, the question related to foreign policy (EU) generates more nonresponse – consistent with hypothesis H3b. The other questions, however, do not reveal such a pattern: 'developmental aid' (in the foreign policy domain) yields 5 to 12 percent nonresponse as compared to 4 to 11 percent for 'lowering welfare benefits'

(socio-economic domain) and 'adoption by same-sex couples' (ethical domain). So it would be incorrect to conclude that questions related to foreign policy consistently result in more nonresponse than items related to the socio-economic, ethical and multicultural issues; hypothesis H3a and H3b are not supported for the explicit DK option.

Are some questions, related to certain issues, more susceptible to the use of a filter question? The answer depends on the specific experiment and panel. Looking at either total nonresponse or only at the number of no's to a filter question, i.e. 'filtered out' in Table 8.8 and Table 8.9, the respondents of the filter variants in the third experiment (Team Vier) gave a non-substantive answer most often for the foreign policy self-placement item on the EU: 16 to 18 percent of the Team Vier respondents said 'no' to the filter question or used the (implicit or explicit) DK option, which is the highest item nonresponse for all issues included in Table 8.8 and Table 8.9. The respondents in the second (EenVandaag Opiniepanel) experiment, however, used a non-substantive response option most for the first two items in the questionnaire: self-placement on income differences and 'lowering welfare benefits'. Whether this is related to issue content (socio-economic) or the placement of the questions at the start of the questionnaire, when respondents were not yet used to the unfamiliar filter question, is unknown. Nevertheless, the conclusion that should be drawn is that both hypotheses H3a and H3b are not supported; questions related to a major political dimension did not result in the least nonresponse nor did questions related to foreign policy consistently result in the most item nonresponse.

8.6   Conclusion

The goal of this chapter was to compare the three survey experiments to find out whether non-substantive response options consistently affected the results of a survey. In doing so, the findings would be cross validated if similar patterns were discovered. The answer to the question how non-substantive response options affect survey results depends on two things: the type of option and the type of results. Three non-substantive response options were considered: the implicit (DK) option to skip questions, the explicit DK option as a response category and the filter question.

The main effect of the implicit DK option is not much effect at all. On average a maximum of 2 percent used this option. The relatively low item nonresponse rate is, however, consistent with the expectation (hypothesis H1a) that more explicit options would result in more item nonresponse and hence less explicit options – e.g. an implicit DK option – in less item nonresponse.

The explicit DK variants yielded more DK answers than the implicit DK variants and the filter question variants rendered more 'no's' than when respondents used an explicit DK option, at least for one application of the filter question (the Team Vier experiment). So it can be concluded that the expectation (hypothesis H1a) that more explicit non-substantive response options result in more nonresponse and the expectation (hypothesis H1b) that filter questions render more nonresponse than an explicit DK option is correct, although the second statement is tentative because filter questions did not have the strongest effect on item nonresponse in all experiments – possibly because of the way non-substantive response options are normally offered to the respondents of the three panels or because of differences in panel composition and offered incentives.

The finding that more explicit non-substantive response options result in more item nonresponse is not very surprising. The effect of non-substantive response options on the overall distribution of opinions is, however, much less expected. The data strongly support hypothesis H2a; item nonresponse data are missing at random and therefore do not result in nonresponse bias. Consequently, the picture of public opinion hardly changes for any of the questions or variants included in the survey experiments when item nonresponse is excluded as missing data. The outcome seems robust when item nonresponse is disregarded.

No systematic patterns could be discerned regarding question design effects and question content. Survey questions related to the main dimensions in Dutch politics did not consistently result in less item nonresponse (as expected in hypothesis H3a) and questions about foreign policy issues did not consistently result in more item nonresponse (as expected in hypothesis H3b), at least for the items included in the comparative analysis.

8

CHAPTER 9

Discussion and Implications (Conclusions II)

'The argument today is not about developing a new conception of public opinion in which polls have little or no role. It's about having polls remain in their central role by making sure they tell the whole story about what the public is thinking - to include not just what preferences it has, but also what proportion of the public has no preferences at all' (Moore, 2008, p. 158).

Moore nicely summarizes the goal of this study: to reveal a more complete picture of public opinion as measured by surveys and polls, including that part of the public that does not express an opinion or does not care about the outcome. Methodological aspects like item nonresponse and question design effects may be of interest to academics and survey methodologists, but they are relevant more generally when they affect the picture of public opinion. What impression of public opinion do we get when the image emerging from surveys is different when a different question design is used? Can we expect politicians to be responsive to public opinion if it is dependent on methodological choices like question wording? Sniderman *et al* argue that 'if it is true that citizens are just making it up as they go along, then political leaders may even have an obligation to discount what the public thinks that it thinks, since if the question had varied even slightly, the answers could well have varied markedly' (Sniderman et al., 2001, p. 256). If the picture of public opinion varies depending on question design, politicians may feel allowed or even obliged to dismiss public opinion. Methodological issues, and specifically question design effects, are therefore crucial for the debate on public opinion in a democracy and politicians' responsiveness to public opinion surveys.

The general research question in this study was: *How does question design regarding non-substantive response options affect survey outcomes?* More specifically, the effect of using non-substantive response options – the Don't Know option, filter question and follow-up question – on item nonresponse and the substantive overall distribution of opinions was examined. To find out whether and how question design elements, and specifically non-substantive response options, affect results, three survey experiments were conducted with three Dutch internet panels (the LISS panel, the EenVandaag Opiniepanel and the Team Vier internet panel). The general aim was to investigate the impact of various ways of offering a non-substantive response option on 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 aim was to look at various ways to register non-substantive answers and their impact on survey outcomes and specifically on substantive answers to survey questions.

9

9.1 Non-Substantive Response Options and Survey Results: A Summary of Findings

How do non-substantive response options affect item nonresponse? Four options were explored: no non-substantive response option (forced choice), the implicit (DK) option to skip questions, an explicit DK option (both single explicit as a response category and double explicit with a reference to the DK option in the question) and the filter question (in a weak and strong worded version). These were used as single non-substantive response options or in a combination of a filter question followed by a substantive opinion question with a DK option. The main expectation was that the more explicit a non-substantive response option was presented, the more item nonresponse would result. This expectation is supported by the data: the explicit DK option renders more item nonresponse than the implicit DK option and the filter question renders more item nonresponse than the explicit DK option in one of the two applications of the filter question. Offering a non-substantive response option more explicitly 'encourages' respondents to give a non-substantive answer and the effect of a filter question preceding the opinion question is stronger than the effect of a DK option, although the size of the effect depends on the specific panel and experiment and the filter question did not have the strongest effect on item nonresponse in all experiments. The filter question rendered about 5 to 6 percent (in EenVandaag's Opiniepanel) or about 18 percent nonresponse (in Team Vier's internet panel), which was much lower than suggested in the literature (Bishop, 2005; Bishop et al., 1983; Eckman et al., 2014; Schuman & Presser, 1979). Nevertheless, both the explicitness and type of non-substantive response option clearly affected the number of non-substantive answers.

A second question was whether and how the overall substantive distribution of opinions changed with question design. A higher nonresponse rate provides greater opportunities for bias, but the occurrence of a bias depends on whether the data are missing at random or not. However, the picture of public opinion proved to be barely affected by question design. The overall outcome was robust and, with few exceptions, does not suggest that we should be overly concerned about whether and how the respondent could express having no opinion, at least for gauging what the general public wants or thinks. Including non-substantive response options may still be preferable to establish what part of the public holds an opinion, but for a valid impression of the policy option preferred in society at large it does not really make a difference whether a non-substantive response option is offered or not. Overall, the effect of offering a non-substantive response option is that item nonresponse changes, but not the resulting picture of public opinion in terms of majorities or pluralities.

The effect of a non-substantive response option for substantively different issues was expected to vary per the relation to a major political dimension (resulting in less item nonresponse) or to foreign policy issues (resulting in more nonresponse). The results were mixed: both expectations were supported in two of the three experiments. More research is needed to differentiate between question content.

Strictly speaking not being a non-substantive response option, the follow-up question, in which respondents were asked after a substantive answer to the opinion question, whether they would be upset if their opinion did not prevail, makes a distinction between permissive and directive opinions. Permissive opinions, i.e. 'not upset' answers and item nonresponse, arguably cannot be considered as clear or strong directives for politicians. Permissive opinions can therefore be construed or at least considered to be a non-substantive answer; part of the public does not care which policy is pursued, even after expressing a particular opinion. About two-third of the respondents expressed a directive opinion, which was more than expected (see Moore, 2008). Furthermore, the expectations that issues related to a major political dimension would result in more directive opinions and foreign policy issues in relatively fewer directive opinions were both supported.

The effect of non-substantive response options on survey results were the core of this study, but some other methodological issues were addressed. Questions with more substantive response categories yielded less item nonresponse and respondents of variants without a non-substantive response option, or a less explicit option, used the neutral or midpoint response category relatively often. All of these conclusions are, however, only tentative; more systematic analyses are needed.

To summarize: Bogart (1972) is right in saying that 'the question of what people think about public issues is really secondary to the question of whether they think about them at all' and this is captured by non-substantive response options. Public opinion is, however, quite robust as regards the substantive outcome.

9.2 Implications and Recommendations

What does it mean that the various non-substantive response options affect item nonresponse but not the overall substantive distribution of opinions? One could conclude that for the substantive overall outcome it does not matter how and whether non-substantive response options are offered: the picture of public opinion does not differ. If journalists or politicians want to know what policy option or position is preferred by the public at large, they can look at any questionnaire variant and as a rule the same preference is measured. But that is not the whole story: item nonresponse should be treated as an indicator of whether the public

9

thinks about an issue at all and this information should not be ignored. Should 'we' listen to what the public wants if a large part of the public does not have an opinion or does not care about the issue?

Respondents should be given the option to give a non-substantive answer in polls and survey questions for several reasons. First, it is a service to respondents who take the time to participate and may feel discouraged when they are unable to leave a question unanswered. Secondly, item nonresponse reflects how many people are unable or unwilling to give a substantive answer. This is relevant in and of itself, and can be considered as an indicator of whether the general public cares or thinks about a particular issue and subsequently whether public opinion on this issue should be taken seriously and into account. The notions 'non-substantive answers' and 'missing data', which were used frequently in this study, should not detract from the fact that valuable information is gathered. Offering an explicit DK option and/or filter question is a means to get a complete and nuanced picture of public opinion.

An innovative non-substantive response option is the follow-up question. The resulting distinction between permissive and directive opinions provides information as to how individual opinions and how 'serious' the answers to opinion questions should be regarded. When this distinction is taken into account, oftentimes only a plurality remains which really supports and wants to stick to a certain policy. Surveys may provide valuable information about what the public wants (in a democracy), and the use of the follow-up question shows that we can investigate whether the public itself wants the opinions to matter. Even when the general public has and expresses an opinion about an issue, their thoughts and preferences do not always have to guide politicians in their decision-making, especially if part of the public does not care about what happens to their thoughts and preferences. Listening to the largest group when part of the public does not really care about the outcome means that a relatively small group with strong opinions dictates policy, which is very important to keep in mind for politicians who want to take the public's voice into account.

Based on this study, a number of recommendations can be made. From a methodological perspective, I recommend a) to focus more on the question whether the *complete* picture of public opinion changes and not just item nonresponse, by looking at both the substantive answers and the number of non-opinions expressed; b) to include non-substantive response options to reveal public ignorance and non-opinions and discourage nonattitudes, i.e. to view non-substantive answers as valid and relevant information; and c) to make more use of the follow-up question and apply it more systematically to examine the respondent's intent and to see whether the given answer should be considered as directive for policymaking. For politicians,

journalists and anyone interested in the picture of public opinion, I recommend a) to always view survey results critically and find out whether respondents were offered a non-substantive response option; b) to keep in mind how much of the public is ignorant about the issue at hand; and c) to consider the distinction between directive and permissive opinions. Not all survey results are equally guiding.

My main recommendation, however, is to start by including the most basic non-substantive response option, the explicit Don't Know option, as a response category for opinion questions. And this recommendation (to offer an explicit DK option) is even more applicable for technical and abstract issues, for issues the respondents may not have (as much) personal experience with or for new 'emerging' issues. Knowledge is power and non-substantive answers provide valuable information about public opinion. Respondents should not be 'pushed' into using substantive response options. The DK option does not lengthen the survey (thereby increasing respondent burden) or alter response behaviour substantially, as shown in this study, but it does give us more information about public opinion as measured with surveys. Furthermore, by making the DK option a standard practice, we may find out more about when respondents are unable or unwilling to express an opinion.

9.3   Limitations of the Study

This study has three main limitations, related to 1) external validity; 2) the internal validity of the findings about other methodological elements than the non-substantive response options, for example the number of substantive response categories and panel characteristics; and 3) the choice for the Netherlands as a case to collect data.

The first limitation is about the ability to generalize the findings of this study to the (Dutch) population. Only one of the three internet panels used in this study is probability-based (the LISS panel). The other two panels are volunteer or convenience samples and result from self-selection, which threatens their representativeness of the population. Only probability-based, random samples allow for generalization to the population. Nonprobability-based panels may suffer from problems with non-coverage and selection bias, e.g. an overrepresentation of Internet users, people interested in research or substantive issues, or people wanting to earn money by doing surveys. Generalizing to the population is therefore problematic. Regarding external validity, the main point is that one should be careful in generalizing the findings of this study to the population. The American Association for Public Opinion Research concluded in a report on online panels that 'claims of "representativeness" should be avoided (AAPOR, 2010, p. 5).

At the same time, the problems with external validity should not be exaggerated.

9

First of all, the outcome of a survey conducted with an internet panel is how public opinion in practice is often gauged and this representation of what the public wants is subsequently used in the public and political debate. So even in case of the sample being unrepresentative of the population, the outcome is typical for what the picture of public opinion very often looks like.

The second reason why the limited external validity is not a major limitation in this study, is that the main aim is to establish causal validity. Survey experiments are very suitable for that purpose. AAPOR concludes that a nonprobability online panel, like the EenVandaag Opiniepanel and Team Vier's internet panel, can be 'an acceptable alternative to traditional probability-based methods' when generalizing to the population is not the goal of the study (AAPOR, 2010, p. 5). The between-subjects experimental design of this study allowed for a comparison of various randomly drawn subgroups within the panels to establish internal validity. Subgroups were randomly assigned to a question design, i.e. the treatment variable non-substantive response options. The main goal was to examine the effect of non-substantive response options on survey results under various treatment conditions, not necessarily to generalize to the population. All three internet panels used in this study were suitable for such internet survey experiments. In other words: despite the limited external validity, the findings are robust.

The second limitation, i.e. the internal validity of other methodological elements than non-substantive response options, directly follows from the experimental design. The variables manipulated were non-substantive response options; other factors - including number of response categories, the use of a midpoint option, and panel characteristics – were held constant as much as possible. The number of response categories and inclusion of a midpoint option varied, because the response categories of the original questions varied, but as such they were not subject to experimental manipulation. All elements were held constant across subgroups and questionnaire variants, which produced robust findings that non-substantive response options affect survey outcomes.

What was not possible, however, was a comparison of the findings between the three experiments. The internet panels used to collect data differed substantially from one another, e.g. in sample composition, incentives and number of respondents. Furthermore, the timing of the three experiments was not the same. Some tentative trends among the three experiments and panels were described in chapter 8, but the main strength in terms of internal validity lies in the between-subjects-design set-up of the three individual experiments. How panel characteristics affected the outcome of the internet survey experiments is unknown.

The third and final main limitation of this study is the choice for the Netherlands as a case to conduct data collection. Whether the findings travel to other countries

is unknown, although previous research in for example the United States and Germany suggests that the findings of this study are not deviant. Nevertheless, the conclusions should not unthinkingly be applied to other settings.

9.4 Avenues for Future Research

In this study, I looked at a small but important part of the puzzle regarding the way public opinion is measured with surveys: non-substantive response options and their effect on survey outcomes. It has been demonstrated that survey questions do not simply collect existing opinions; they affect the answers given and therefore partially create the picture of public opinion. The first avenue for future research concerns the limitations discussed and points to a more systematic analysis of the number of response categories, and the use of a midpoint or 'neutral' response category. The general suggestion is that the scope should be broadened to encompass other elements of survey methodology to see how the way we ask survey questions affects the outcome.

The focus in this study was on respondents' answers to survey questions and the picture of public opinion which is created by aggregating such individual answers. Question design may, however, not affect all respondents equally and in the same way. More research should be done to see which respondents are more susceptible to certain design choices. This study has not distinguished between data missing completely at random (MCAR) and data missing at random given covariates (MAR). Question design aspects may, however, affect certain respondents differently. By looking at respondent characteristics and the differences between respondents and non-respondents at the individual level, a more complete picture can be created of the ways in which question design affects non-substantive and substantive answers to survey questions.

Another way of differentiating between the susceptibility of individual respondents to question design is by looking at their level of information. A more informed citizenry may result in a different picture of public opinion, which can be gauged by deliberative polling. Deliberative polling is according to some researchers the solution for the two major problems that citizens seem to suffer from: 'rational ignorance' and 'nonattitudes' (Fishkin, 1996, pp. 133-134). When using the deliberative polling method, the question is what the public's opinion would be if everyone was well and equally informed. By giving respondents information and discussing the issue at hand in small groups, according to some it is established 'what the public would think if it had a better chance to think about the questions at issue' (Fishkin, 1996, p. 134; see also Althaus, 1996; Fishkin, 1991; Steiner, 2010; Sturgis, 2003). Even

9

though it can be argued that feeding respondents information might 'contaminate' the sample of citizens (Moore, 2008, p. 34) and the process is time consuming and costly, deliberative polling provides the opportunity to see what public opinion would look like if citizens were better informed.

While this study has ignored what device respondents used to complete the web or internet survey, there is some evidence that response patterns are not the same for all devices (see e.g. De Bruijne & Wijnant, 2014; Mavletova, 2013; Mavletova & Couper, 2014; Toninelli, Pinter, & Pedraza, 2015). Initial response to a survey invitation, completion time, drop-out rates and the length of answers to open-ended questions are only a few examples of challenges introduced by doing surveys on tablets and mobile phones. The differences between web surveys completed by mobile phone or tablet rather than laptop or computer are even less well-known when item nonresponse rates and substantive response patterns are considered.

A final avenue for future research is to move beyond public opinion as measured by surveys and also look at other means to assess public opinion, like media content or letters (to officials) (Herbst, 1998, pp. 182-185) or social media. These representations of public opinion are based on a different conceptualization of public opinion, but they also affect the citizens who are asked about their opinion in surveys. Does public opinion look different in these alternative representations? How does the process of public opinion formation work in these alternative representations? And how do they affect public opinion as measured by surveys?

The main point to take away from this study is that public opinion consists of answers to individual opinion *questions*, not necessarily individual *opinions*. These answers are at least to some extent affected by the way the questions are asked.

Bibliography
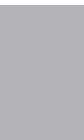
Aarts, K., & Thomassen, J.J.A. (2008). Dutch Voters and the Changing Party Space 1989–2006. Acta Politica, 43(2), 203-234.

AD (2011, 22 October). Vader Maxima niet welkom bij kroning. *AD*. Retrieved from http://www.ad.nl/ad/nl/1012/Nederland/article/detail/2982585/2011/10/22/Vader-Maxima-niet-welkom-bij-kroning.dhtml

Agiesta, J. (2016, March 25). Support for SCOTUS hearings remains strong, CNN/ORC poll finds. *CNN Politics*. Retrieved from http://edition.cnn.com/2016/03/25/politics/merrick-garland-supreme-court-nominee/

Aldrich, J. H., Gelpi, C., Feaver, P., Reifler, J., & Sharp, K. T. (2006). Foreign Policy and the Electoral Connection. *Annual Review of Political Science, 9*(1), 477-502.

Althaus, S. (1996). Opinion Polls, Information Effects, and Political Equality: Exploring Ideological Biases in Collective Opinion. *Political Communication, 13*(1), 3-21.

Alvarez, M. R., & Brehm, J. (2002). *Hard Choices, Easy Answers*. New Jersey: Princeton University Press.

Alwin, D. F. (1997). Feeling Thermometers Versus 7-Point Scales: Which are Better? *Sociological Methods & Research, 25*(3), 318-340.

Andeweg, R. B. (2012). A Least Likely Case: Parliament and Citizens in the Netherlands. *The Journal of Legislative Studies, 18*(3-4), 368-383.

Andeweg, R. B. (2014). Roles in Legislatures. In S. Martin, T. Saalfeld, & K. W. Strom (Eds.), *The Oxford Handbook of Legislative Studies* (pp. 267-282). Oxford: Oxford University Press.

Andeweg, R. B., & Thomassen, J. J. A. (2005). Modes of Political Representation: Toward a New Typology. *Legislative Studies Quarterly, 30*(4), 507-528.

ANES. (2014). ANES History. Retrieved 15 October, 2014, from www.electionstudies.org

Anstead, N. & O'Loughlin, B. (2014). Social Media Analysis and Public Opinion: The 2010 UK General Election. *Journal of Computer-Mediated Communication 20*(2), 204-220.

Arceneaux, K. (2010). The Benefits of Experimental Methods for the Study of Campaign Effects. *Political Communication, 27*(2), 199-215.

Arnold, C., & Franklin, M.N. (2012). Introduction: Issue Congruence and Political Responsiveness. *West European Politics, 35* (6), 1217-1225.

Ayidaya, S. A., & McClendon, M. J. (1990). Response Effects in Mail Surveys. *Public Opinion Quarterly, 54*(2), 229-247.

Barabas, J., & Jerit, J. (2010). Are Survey Experiments Externally Valid? *American Political Science Review, 104*(2), 226-242.

Bassili, J. N., & Fletcher, J. F. (1991). Response Time Measurement in Survey Research: A Method for CATI and a New Look at Nonattitudes. *Public Opinion Quarterly, 55*(3), 331-346.

Beatty, P., Hermmann, D., Puskar, C., & Kerwin, J. (1998). 'Don't Know' Responses in Surveys: Is What I Know What You Want to Know and Do I Want You to Know It? *Memory 6*(4), 407-426.

Beniger, J. R. (1983). Comments. *Public Opinion Quarterly, 47*(4), 479-489.

Berinsky, A. J. (2004). *Silent Voices. Public Opinion and Political Participation in America*. Princeton: Princeton University Press.

Bethlehem, J. (2010). Selection Bias in Web Surveys. *International Statistical Review, 78*(2), 161-188.

Bethlehem, J. (2013). *De Kwaliteit van Internetpeilingen [Oratie]*. Retrieved from https://peilingpraktijken.nl/wp-content/uploads/2014/06/Oratie-Bethlehem.pdf

Bethlehem, J., & Biffignandi, S. (2011). *Handbook of Web Surveys*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Biemer, P. P. (2010a). Overview of Design Issues: Total Survey Error. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of Survey Research* (pp. 27-57). Bingley, UK: Emerald Publishing Group Limited.

Biemer, P. P. (2010b). Total Survey Error: Design, Implementation, and Evaluation. *Public Opinion Quarterly* 74(5), 817-848.

Biemer, P. P. (2011). *Latent Class Analysis of Survey Error*. Hoboken, New Jersey: John Wiley & Sons.

Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to Survey Quality*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Bishop, G. F. (2005). *The Illusion of Public Opinion: Fact and Artifact in American Public Opinion Polls*. Lanham, etc.: Rowman & Littlefield Publishers.

Bishop, G. F. (2015, November 30). Polls Can Create an Illusion of Public Opinion. *The New York Times*. Retrieved from http://www.nytimes.com/roomfordebate/2015/11/30/does-polling-undermine-democracy/polls-can-create-an-illusion-of-public-opinion

Bishop, G. F., Oldendick, R. W., & Tuchfarber, A. J. (1983). Effects of Filter Questions in Public Opinion Surveys. *Public Opinion Quarterly, 47*(4), 528-546.

Bishop, G. F., Oldendick, R. W., Tuchfarber, A. J., & Bennett, S. E. (1980). Pseudo-Opinions on Public Affairs. *Public Opinion Quarterly, 44*(2), 198-209.

Blumer, H. (1946). Collective Behavior. In A. M. Lee (Ed.), *New Outlines of the Principles of Sociology* (pp. 167-222). New York: Barnes and Noble.

Blumer, H. (1948). Public Opinion and Public Opinion Polling. *American Sociological Review, 13*(5), 542-549.

Bogart, L. (1972). *Silent Politics: Polls and the Awareness of Public Opinion*. New York: Wiley.

Bosnjak, M., & Tuten, T. L. (2001). Classifying Response Bbehaviors in Web-Based Surveys. *Journal of Computer-Mediated Communication, 6*(3), 1–17.

Bourdieu, P. (1973). *L'Opinion Publique n'Existe Pas*. Retrieved from http://www.acrimed.org/IMG/article_PDF/article_a3938.pdf

Bovens, M. A. P., & Wille, A. C. (2008). Politiek Vertrouwen Langs de Meetlat. *Socialisme en Democratie, 65*(10), 32-43.

Brace, P., & Boyea, B. D. (2007). Judicial Selection Methods and Capital Punishment in the American States. In M. J. Streb (Ed.), *Running for Judge: The Rising Political, Financial, and Legal Stakes of Judicial Elections* (pp. 186-203). New York: New York University Press.

Bradburn, N. M., Sudman, S., Blair, E., & Stocking, C. (1978). Question Threat and Response Bias. *Public Opinion Quarterly, 42*(2), 221-234.

Bradburn, N. M., Sudman, S., & Wansink, B. (2004). *Asking Questions. The Definitive Guide to Questionnaire Design - For Market Research, Political Polls, and Social and Health Questionnaires*. San Francisco: Jossey-Bass.

Brettschneider, F. (1997). The Press and the Polls in Germany, 1980–1994 Poll Coverage as an Essential Part of Election Campaign Reporting. *International Journal of Public Opinion Research, 9*(3), 248-265.

British Election Studies (2014). British Election Studies Information System. Retrieved 15 October, 2014, from www.besis.org

Bryman, A. (2012). *Social Research Methods* (4th ed.). Oxford: Oxford University Press.

Burchell, B., & Marsh, C. (1992). The Effect of Questionnaire Length on Survey Response. *Quality and Quantity, 26*(3), 233-244.

Caeyers, B., Chalmers, N., & De Weerdt, J. (2012). Improving Consumption Measurement and Other Survey Data Through CAPI: Evidence from a Randomized Experiment. *Journal of Development Economics, 98*(1), 19-33.

Callegaro, M., Baker, R., Bethlehem, J., Göritz, A. S., Krosnick, J. A., & Lavrakas, P. J. (2014). Online Panel Research: History, Concepts, Applications and a Look at the Future. In M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online Panel Research: A Data Quality Perspective* (pp. 1-22). Hoboken, New Jersey: John Wiley & Sons, Inc.

Callegaro, M., Yang, Y., Bhola, D. S., Dillman, D. A., & Chin, T.-Y. (2009). Response Latency as an Indicator of Optimizing in Online Questionnaires. *Bulletin de Méthodologie Sociologique, 103*(1), 5-25.

Campbell, A., Converse, P. E., Miller, W. E., & Stokes, D. E. (1960). *The American Voter*. New York: John Wiley & Sons, Inc.

Carlson, D. K. (2002). Privatization Paradox: Americans and Social Security. Retrieved from http://news.gallup.com/poll/6091/Privatization-Paradox-Americans-Social-Security.aspx

Carpini, M. X. D., & Keeter, S. (1993). Measuring Political Knowledge: Putting First Things First. *American Journal of Political Science, 37*(4), 1179-1206.

Carpini, M. X. D., & Keeter, S. (1997). *What Americans know about politics and why it matters*. New Haven: Yale University Press.

Champagne, P. (1990). *Faire l'Opinion: Le Nouveau Jeu Politique*. Paris: Les Éditions de Minuit.

Champagne, P. (2004). Making the People Speak: The Use of Public Opinion Polls in Democracy. *Constellations 11*(1), 61-75.

CNN. (2017, April 10). Operation Iraqi Freedom and Operation New Dawn Fast Facts. Retrieved May 30, 2017, from http://edition.cnn.com/2013/10/30/world/meast/operation-iraqi-freedom-and-operation-new-dawn-fast-facts/

Conrad, F. G., Couper, M. P., Tourangeau, R., & Peytchev, A. (2010). The Impact of Progress Indicators on Task Completion. *Interacting with Computers, 22*(5), 417-427.

Converse, J. M. (1987). *Survey Research in the United States: Roots and Emergence, 1890-1960*. Berkeley: University of California Press.

Converse, P. E. (1964). The Nature of Belief Systems in Mass Publics. In D. E. Apter (Ed.), *Ideology and Discontent*. New York: The Free Press.

Converse, P. E. (1987). Changing Conceptions of Public Opinion in the Political Process. *Public Opinion Quarterly, 51*(2), S12-S24.

Converse, P. E. (2000). Assessing the Capacity of Mass Electorates. *Annual Review of Political Science, 3*, 331-353.

Couper, M. P. (2000). Review: Web Surveys: A Review of Issues and Approaches. *Public Opinion Quarterly, 64*(4), 464-494.

Couper, M. P., & Miller, P. V. (2008). Web Survey Methods: Introduction. *Public Opinion Quarterly, 72*(5), 831-835.

Couper, M. P., Tourangeau, R., Conrad, F. G., & Zhang, C. (2013). The Design of Grids in Web Surveys. *Social Science Computer Review, 31*(3), 322-345.

Couper, M. P., Traugott, M. W., & Lamias, M. J. (2001). Web Survey Design and Administration. *Public Opinion Quarterly, 65*(2), 230-253.

Crawford, S. D., Couper, M. P., & Lamias, M. J. (2001). Web Surveys: Perceptions of Burden. *Social Science Computer Review, 19*(2), 146-162.

Crosby, S. L. (2016, May 17). EU referendum: 51 per cent back Remain as Sir Lynton Crosby warns Leave camp must focus more on voters and less on infighting. *The Telegraph*. Retrieved from http://www.telegraph.co.uk/news/2016/05/16/eu-referendum-51-per-cent-back-remain-as-sir-lynton-crosby-warns/

Dalton, R. J. (2013). *Citizen Politics: Public Opinion and Political Parties in Advanced Industrial Democracies*. Washington: CQ Press.

De Bruijne, M., & Wijnant, A. (2014). Improving Response Rates and Questionnaire Design for Mobile Web Surveys. *Public Opinion Quarterly, 78*(4), 951-962.

De Graaf, P. (2010, 26 January). Peiling is geen snelle enquête, ook geen steekproef. *De Volkskrant*.

de Leeuw, E. D. (2001). Reducing Missing Data in Surveys: An Overview of Methods. *Quality and Quantity, 35*(2), 147-160.

De Leeuw, E. D. (2008). The Effect of Computer-Assisted Interviewing on Data Quality: A Review of the Evidence. In J. Blasius, J. Hox, E. D. De Leeuw, & P. Schmidt (Eds.), *Social Science Methodology in the New Millennium: Proceedings of the Fifth International Conference on Logic and Methodology*. Retrieved from https://edithl.home.xs4all.nl/surveyhandbook/deleeuw-cai-overview-updated.pdf

De Leeuw, E. D., & Hox, J. (2011). Internet Surveys as part of a Mixed-Mode Design. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and Behavioral Research and the Internet* (pp. 45-76). New York: Routledge.

De Leeuw, E. D., Hox, J., & Huisman, M. (2003). Prevention and Treatment of Item Nonresponse. *Journal of Official Statistics, 19*(2), 153-176.

De Leeuw, E. D., Hox, J. J., & Dillman, D. A. (2008). The Cornerstones of Survey Research. In E. D. De Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International Handbook of Survey Methodology* (pp. 1-17). New York: Lawrence Erlbaum.

De Leeuw, E. D., Hox, J. J., & Scherpenzeel, A. (2010). *Emulating Interviewers in an Online Survey: Experimental Manipulation of 'Do-Not-Know' over the Phone and on the Web*. Paper presented at the AAPOR annual conference, Chicago, Illanois.

DeMaio, T. J. (1980). Refusals: Who, Where and Why. *Public Opinion Quarterly, 44*(2), 223-233.

Derouvray, C., & Couper, M. P. (2002). Designing a Strategy for Reducing "No Opinion" Responses in Web-Based Surveys. *Social Science Computer Review, 20*(1), 3-9.

Dewey, J. (1954). *The Public and Its Problems*. Chicago: The Swallow Press.

Dillman, D. A. (2007). *Mail and Internet Surveys: The Tailored Design Method*. Hoboken: Wiley.

Dillman, D. A., & Bowker, D. K. (2002). The Web Questionnaire Challenge to Survey Methodologists. In B. Batinic, U.-D. Reips, M. Bosnjak, & A. Werner (Eds.), *Online Social Sciences* (pp. 53-71). Seattle, Toronto, Bern, Göttingen: Hogrefe & Huber Publishing.

Dillman, D. A., Smyth, J. D., & Christian L. M. (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. New York: Wiley.

Dixhoorn, A. v. (2006). *De stem des volks. Publieke opinie, opinieonderzoek en democratie*. Den Haag: SCP.

Druckman, J. N., Green, D. P., Kuklinski, J. H., & Lupia, A. (2006). The Growth and Development of Experimental Research in Political Science. *American Political Science Review, 100*(04), 627-635.

Eckman, S., Kreuter, F., Kirchner, A., Jäckle, A., Tourangeau, R., & Presser, S. (2014). Assessing the Mechanisms of Misreporting to Filter Questions in Surveys. *Public Opinion Quarterly, 78*(3), 721-733

EenVandaagOpiniepanel. (2016). Opiniepanel uitleg. Retrieved May 27th, 2016, from http://opiniepanel.eenvandaag.nl/uitleg

Erikson, R. S., Mackuen, M. B., & Stimson, J. A. (2002). Public Opinion and Policy: Casual Flow in a Macro System Model. In J. Manza, F. L. Cook, & B. I. Page (Eds.), *Navigating Public Opinion: Polls, Policy and the Future of American Democracy* (pp. 33-53). Oxford: Oxford University Press.

Erikson, R. S., & Tedin, K. L. (2015). *American Public Opinion: Its Origins, Content, and Impact* (9th ed.). New York: Routledge.

Eulau, H. (1962). The Legislator as Representative: Representational Roles. In J. C. Wahlke, H. Eulau, W. Buchanan, & L. C. Ferguson (Eds.), *The Legislative System: Explorations in Legislative Behavior* (pp. 267-286). New York: Wiley.

Eulau, H., Wahlke, J. C., & Abramowitz, A. I. (1978). *The Politics of Representation: Continuities in Theory and Research*. Beverly Hills: Sage Publications (CA).

Eulau, H., Wahlke, J. C., Buchanan, W., & Ferguson, L. C. (1959). The Role of the Representative: Some Empirical Observations on the Theory of Edmund Burke. *American Political Science Review, 53*(3), 742-756.

Eurobarometer. (2014). Eurobarometer. Retrieved 15 October, 2014, from http://ec.europa.eu/public_opinion/index_en.htm

Everts, P. (2008). *De Nederlanders en de Wereld: Publieke Opinies na de Koude Oorlog*. Assen: Van Gorcum.

Everts, P., & Isernia, P. (2015). *Public Opinion, Transatlantic Relations and the Use of Force*. Basingstoke, Hampshire (UK): Palgrave Macmillan.

EVS. (2014). About EVS. Retrieved 15 October, 2014, from www.europeanvaluesstudy.eu

Feldman, S. (1988). Structure and Consistency in Public Opinion: the Role of Core Beliefs and Values. *American Journal of Political Science, 32*(2), 416-440.

Fink, A. (2009). *How To Conduct Surveys: A Step-by-Step Guide* (4th ed.). Thousand Oaks, California: SAGE Publications, Inc.

Fiorina, M. P. (1978). Economic Retrospective Voting in American National Elections: A Micro-Analysis. *American Journal of Political Science, 22*(2), 426-443.

Fiorina, M. P. (1981). *Retrospective Voting in American National Elections*. New Haven: Yale University Press.

Fishkin, J. S. (1991). *Democracy and Deliberation: New Directions for Democratic Reform*. New Haven: Yale University Press.

Fishkin, J. S. (1996). The Televised Deliberative Poll: An Experiment in Democracy. *Annals of the American Academy of Political and Social Science, 546*, 132-140.

Fowler, F. J. (2014). *Survey Research Methods* (5th ed.). Thousand Oaks: SAGE Publications, Inc.

Fowler, F. J., & Cosenza, C. (2008a). Design and Evaluation of Survey Questions. In L. Bickman & D. J. Rog (Eds.), *The SAGE Handbook of Applied Social Research Methods* (pp. 375-412). Newbury Park: SAGE Publications.

Fowler, F. J., & Cosenza, C. (2008b). Writing Effective Questions. In E. D. De Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International Handbook of Survey Methodology* (pp. 136-160). London: Lawrence Erlbaum Associates.

Fowler, F. J., & Mangione, T. W. (1990). *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. Newbury Park: Sage Publications.

Fricker, S., Galesic, M., Tourangeau, R., & Yan, T. (2005). An Experimental Comparison of Web and Telephone Surveys. *Public Opinion Quarterly, 69*(3), 370-392.

Fricker, S., & Tourangeau, R. (2010). Examining the Relationship Between Nonresponse Propensity and Data Quality in Two National Household Surveys. *Public Opinion Quarterly, 74*(5), 934-955.

Gaines, B. J., Kuklinski, J. H., & Quirk, P. J. (2007). The Logic of the Survey Experiment Reexamined. *Political Analysis, 15*(1), 1-20.

Galesic, M., & Bosnjak, M. (2009). Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. *Public Opinion Quarterly, 73*(2), 349-360.

Gallup (2003). Permissive Public Opinion vs. a Public Mandate. Retrieved from http://news.gallup.

com/poll/8701/permissive-public-opinion-vs-public-mandate.aspx

Gallup, G. (1947). The Quintamensional Plan of Question Design. *Public Opinion Quarterly, 11*(3), 385-393.

Gallup, G., & Rae, S. F. (1940/1968). *The Pulse of Democracy. The Public-Opinion Poll and How it Works.* New York: Greenwood Press.

Ganassali, S. (2008). The Influence of the Design of Web Survey Questionnaires on the Quality of Responses. *Survey Research Methods, 2*(1), 21-32.

Geer, J. (1996). *From Tea Leaves To Opinion Polls. A Theory of Democratic Leadership.* New York: Greenwood Press.

Genovese, M. A., & Streb, M. J. (2004). Polling and the Dilemmas of Democracy. In M. A. Genovese & M. J. Streb (Eds.), *Polls and Politics: The Dilemmas of Democracy* (pp. 1-14). Albany: State University of New York Press.

Gilljam, M., & Granberg, D. (1993). Should We Take Don't Know for an Answer? *Public Opinion Quarterly, 57*(3), 348-357.

Ginsberg, B. (1986). *The Captive Public: How Mass Opinion Promotes State Power.* New York: Basic Books.

Glynn, C. J., Herbst, S., O'Keefe, G., Shapiro, R. Y., & Lindeman, M. (2004). *Public Opinion.* Boulder: Westview Press.

Greenhouse, E. (2015, May 12). Harriet Tubman Wins Survey for First Woman on the $20 Bill. *BloombergPolitics.* Retrieved from http://www.bloomberg.com/politics/articles/2015-05-12/harriet-tubman-wins-poll-for-first-woman-on-the-20-bill

Greszki, R., Meyer, M., & Schoen, H. (2014). The Impact of Speeding on Data Quality in Nonprobability and Freshly Recruited Probability-based Online Panels. In M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online Panel Research: A Data Quality Perspective* (pp. 238-262). Chichester: Wiley.

Greszki, R., Meyer, M., & Schoen, H. (2015). Exploring the Effects of Removing "Too Fast" Responses and Respondents from Web Surveys. *Public Opinion Quarterly, 79*(2), 471-503.

Gripper, A. (2016, May 17). EU referendum poll tracker: Is Britain heading for Brexit and what does the UK think of Europe? *The Daily Mirror.* Retrieved from http://www.mirror.co.uk/news/uk-news/eu-referendum-poll-tracker-uk-7699714

Groves, R. M. (2004). *Survey Errors and Survey Costs.* Hoboken, New Jersey: John Wiley & Sons, Inc.

Groves, R. M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly Special Issue, 70*(5), 646-675.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology.* New York: Wiley.

Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding The Decision to Participate in a Survey. *Public Opinion Quarterly, 56*(4), 475-495.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in Household Interview Surveys.* New York: Wiley.

Groves, R. M., & Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. *Public Opinion Quarterly, 72*(2), 167-189.

The Guardian View on Opinion Polling: Quality Before Quantity. (2015, November 13). *The Guardian.* Retrieved from http://www.theguardian.com/commentisfree/2015/nov/13/the-guardian-view-on-opinion-polling-quality-before-quantity

Haladyna, T. M., & Downing, S. M. (1993). How Many Options is Enough for a Multiple-Choice Test Item? *Educational and Psychological Measurement, 53*(4), 999-1010.

Heath, A., Fisher, S., & Smith, S. (2005). The Globalization of Public Opinion Research. *Annual Review of Political Science, 8,* 297-333.

Heerwegh, D. (2009). Mode Differences Between Face-to-Face and Web Surveys: An Experimental Investigation of Data Quality and Social Desirability Effects. *International Journal of Public Opinion Research, 21*(1), 111-121.

Heerwegh, D., & Loosveldt, G. (2008). Face-to-Face versus Web Surveying in a High-Internet-Coverage Population. *Public Opinion Quarterly, 72*(5), 836-846.

Heith, D. J. (2015, November 30). Politicians Use Polls to Adjust Their Message. *The New York Times.* Retrieved from http://www.nytimes.com/roomfordebate/2015/11/30/does-polling-undermine-democracy/politicians-use-polls-to-adjust-their-message

Hellemann, A., Hollstein, M., Peters, D., Pfeffer, S., & Niehus, M. (2016, April 16). Zwei Drittel der Deutschen gegen Merkels Entscheidung. *Bild.* Retrieved from http://www.bild.de/politik/inland/angela-merkel/nicht-mal-jeder-vierte-gibt-ihrrecht-45406420.bild.html

Herbst, S. (1993a). The Meaning of Public Opinion: Citizens' Constructions of Political Reality. *Media Culture and Society, 15*(3), 437-437.

Herbst, S. (1993b). *Numbered Voices: How Opinion Polling Has Shaped American Politics.* Chicago: University of Chicago Press.

Herbst, S. (1998). *Reading Public Opinion: How Political Actors View the Democratic Process.* Chicago: The University of Chicago Press.

Hippler, H. J., & Schwarz, N. (1989). No Opinion-Filters: A Cognitive Perspective *International Journal of Public Opinion Research, 1*(1), 77-87.

Holsti, O. R. (1992). Public Opinion and Foreign Policy: Challenges to the Almond-Lippmann Consensus Mershon Series: Research Programs and Debates. *International Studies Quarterly, 36*(4), 439-466.

Hoogendoorn, A., & Daalmans, J. (2009). Nonresponse in the Recruitment of an Internet Panel Based on Probability Sampling. *Survey Research Methods, 3*(2), 59-72.

Hox, J. J., & De Leeuw, E. D. (1994). A Comparison of Nonresponse in Mail, Telephone, and Face-to-Face Surveys. *Quality and Quantity, 28*(4), 329-344.

Huisman, M., & van Der Zouwen, J. (1998). *Item Nonresponse in Scale Data from Surveys: Types, Determinants and Measures.* Groningen: Department of Statistics & Measurement Theory: University of Groningen.

Inglehart, R., & Klingemann, H. (1976). The Changing Structure of Political Cleavages in Western Society. In R. J. Dalton, S. C. Flanagan, & P. A. Beck (Eds.), *Electoral Change in Advanced Industrial Societies: Realignment or Dealignment?* (pp. 25-69). Princeton, NJ: Princeton University Press.

IPSOS. (2014). Onze Geschiedenis. Retrieved 15 October, 2014, from www.ipsos-nederland.nl

Jackson, M., & Cox, D. R. (2013). The Principles of Experimental Design and Their Application in Sociology. *Annual Review of Sociology, 39*(1), 27-49.

Jacobs, L. R., & Shapiro, R. Y. (2002). Politics and Policymaking in the Real World: Crafted Talk and the Loss of Democratic Responsiveness. In J. Manza, F. L. Cook, & B. I. Page (Eds.), *Navigating Public Opinion: Polls, Policy and the Future of American Democracy* (pp. 54-75). Oxford: Oxford University Press.

Josten, M., & Trappmann, M. (2016). Interviewer Effects on a Network-Size Filter Question. *Journal of Official Statistics, 32*(2), 349-373.

Kanne, P. (2016). De Schizofrene Relatie tussen Opiniepeilers, Media en Wetenschappers. *Stuk*

Rood Vlees. Retrieved from http://stukroodvlees.nl/schizofrene-relatie-opiniepeilers-media-en-wetenschappers/

Kaplowitz, M. D., Hadlock, T. D., & Levine, R. (2004). A Comparison of Web and Mail Survey Response Rates. *Public Opinion Quarterly, 68*(1), 94-101.

Katz, D., & Cantril, H. (1937). Public Opinion Polls. *Sociometry, 1*(1/2), 155-179.

Key, V. O. (1961). *Public Opinion and American Democracy*. New York: Alfred A. Knopf.

Key, V. O. (1966). *The Responsible Electorate: Rationality in Presidential Voting, 1936-1960*. Harvard: Belknap Press.

Kinder, D. R. (1998). Opinion and Action in the Realm of Politics. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of Social Psychology* (pp. 778-867). London: Oxford University Press.

Kirk, A., & Wilkinson, M. (2016, May 18). EU Referendum Poll Tracker and Odds. *The Telegraph*. Retrieved from http://www.telegraph.co.uk/news/2016/03/23/eu-referendum-poll-tracker-and-odds/

Knauper, B. (1998). Filter Questions and Question Interpretation: Presuppositions at Work. *Public Opinion Quarterly, 62*(1), 70-78.

KnowledgePanel. (2011). Overview. Retrieved November 3rd 2011, from http://www.knowledgenetworks.com/knpanel/index.html

Kohler, U. (2007). Surveys from inside: An Assessment of Unit Nonresponse Bias with Internal Criteria. *Survey Research Methods, 1*(2), 55-67.

Kohut, A. (2009, October 14). But What Do the Polls Show? How Public Opinion Surveys Came to Play a Major Role in Policymaking and Politics. from http://www.pewresearch.org/2009/10/14/but-what-do-the-polls-show/

Koop, C., & Van Holsteyn, J. (2008). Burke Leeft en Woont in Nederland: Over Volksvertegenwoordigers en de Invloed van Publieke Opinie. *Res Publica, 50*(3), 275-299.

Koopmans, R. & Erbe, J. (2004). Towards a European Public Sphere? *Innovation: The European Journal of Social Science Research*, 17 (2), 97-118.

Kosyakova, Y., Skopek, J., & Eckman, S. (2015). Do Interviewers Manipulate Responses to Filter Questions? Evidence from a Multilevel Approach. *International Journal of Public Opinion Research, 27*(3), 417-431.

Kranenburg, L., & Weimar, W. (2008). Incentives for Living Kidney Donation: What Does the Public Think? *Transplantation, 86*(4), 500-501.

Kreuter, F. (Ed.). (2013). *Improving Surveys with Paradata: Analytic Uses of Process Information*. Hoboken, New Jersey: John Wiley & Sons.

Kreuter, F., McCulloch, S., Presser, S., & Tourangeau, R. (2011). The Effects of Asking Filter Questions in Interleafed Versus Grouped Format. *Sociological Methods & Research, 40*(1), 88-104.

Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity. *Public Opinion Quarterly, 72*(5), 847-865.

Kroh, M. (2007). Measuring Left–Right Political Orientation: The Choice of Response Format. *Public Opinion Quarterly, 71*(2), 204-220.

Krosnick, J. A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology, 5*(3), 213-236.

Krosnick, J. A. (1999). Survey Research. *Annual Review of Psychology, 50*(1), 537-567.

Krosnick, J. A. (2002). The Causes of No-Opinion Responses to Attitude Measures in Surveys: They Are Rarely What They Appear to Be. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A.

Little (Eds.), *Survey Nonresponse* (pp. 87-100). New York: John Wiley & Sons, Inc.

Krosnick, J. A., & Abelson, R. P. (1994). The Case for Measuring Attitude Strength in Surveys. In J. M. Tanur (Ed.), *Questions About Questions: Inquiries into the Cognitive Bases of Surveys* (pp. 177-198). New York: Russell Sage Foundation.

Krosnick, J. A., & Fabrigar, L. R. (1997). Designing Rating Scales for Effective Measurement in Surveys *Survey Measurement and Process Quality* (pp. 141-164). New York: John Wiley & Sons, Inc.

Krosnick, J. A., Holbrook, A. L., Berent, M. K., Carson, R. T., Hanemann, M. W., Kopp, R. J., Mitchell, R. C., Presser, S., Ruud, P. A., Smith, V. K., Moody, W. R., Green, M. C. & Conaway, M. (2002). The Impact of "No Opinion" Response Options on Data Quality: Non-Attitude Reduction or an Invitation to Satisfice? *Public Opinion Quarterly, 66*(3), 371-403.

Krosnick, J. A., & Presser, S. (2010). Question and Questionnaire Design. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of Survey Research* (pp. 263-314). Bingley, UK: Emerald Group Publishing Limited.

Kuijlaars, A. M. D. A. (1999). *Het Huis der Getallen: De Institutionele Geschiedenis van het Centraal Bureau voor de Statistiek (CBS) en de Centrale Commissie voor de Statistiek (CCS), 1899-1996.* (PhD), Erasmus Universiteit, Rotterdam. Retrieved from file://vuw/Personal$/Homes/M/maatjvande/Downloads/huisdergetallen.pdf

Lambert, R. D. (1983). Question Design, Response Set and the Measurement of Left/Right Thinking in Survey Research. *Canadian Journal of Political Science / Revue canadienne de science politique, 16*(1), 135-144.

Lane, C. (2004, June 15). Justices Keep 'Under God' in Pledge: Atheist Father Lacked Standing to Sue on Behalf of Daughter, Court Rules. *Washington Post,* p. A01. Retrieved from http://www.washingtonpost.com/wp-dyn/articles/A41802-2004Jun14.html

Laswell, H. D. (1941). *Democracy Through Public Opinion*. Menasha: George Banta.

Leigh, J. H., & Martin, C. R. J. (1987). "Don't Know" Item Nonresponse in a Telephone Survey: Effects of Question Form and Respondent Characteristics. *Journal of Marketing Research, 24*(4), 418-424.

Lepore, J. (2015). Politics and the New Machine: What the Turn from Polls to Data Science Means for Democracy. *The New Yorker*. Retrieved from https://www.newyorker.com/magazine/2015/11/16/politics-and-the-new-machine

Levy, M. R. (1983). The Methodology and Performance of Election Day Polls. *Public Opinion Quarterly, 47*(1), 54-67.

Lindhout, S. (2016, April 18). Merkel stuit op onbegrip bij meeste Duitsers. *De Volkskrant,* p. 14.

Lippmann, W. (1927). *The Phantom Public*. New York: The MacMillan Company.

LISSPanel. (2016). LISS Panel - Listening to People. Retrieved May 27th, 2016, from http://www.lissdata.nl/lissdata/Home

Loosveldt, G., Pickery, J., & Billiet, J. (2002). Item Nonresponse as a Predictor of Unit Nonresponse in a Panel Survey. *Journal of Official Statistics, 18*(4), 545-557.

Lupia, A. (2015, November 30). Polls Can Give People a Stronger Voice. *The New York Times*. Retrieved from http://www.nytimes.com/roomfordebate/2015/11/30/does-polling-undermine-democracy/polls-can-give-people-a-stronger-voice

Lusinchi, D. (2015). Straw Poll Journalism and Quantitative Data. *Journalism Studies, 16*(3), 417-432.

Luskin, R. C., & Bullock, J. G. (2011). "Don't Know" Means "Don't Know": DK Responses and the Public's Level of Political Knowledge. *The Journal of Politics, 73*(2), 547-557.

Lynn, P. (2008). The Problem of Nonresponse. In E. D. De Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International Handbook of Survey Methodology* (pp. 35-55). London: Lawrence Erlbaum Publishers.

Lynn, P. (2014). Coping with Nonresponse: Overview and Introduction. In U. Engel, B. Jann, P. Lynn, A. Scherpenzeel, & P. Sturgis (Eds.), *Improving Survey Methods: Lessons from Recent Research* (pp. 319-321). London: Routledge.

Lyons, L. (2004). Americans Indivisible on Pledge of Allegiance. Retrieved from http://news.gallup.com/poll/11551/Americans-Indivisible-Pledge-Allegiance.aspx

Malhotra, N. (2008). Completion Time and Response Order Effects in Web Surveys. *Public Opinion Quarterly, 72*(5), 914-934.

Manheim, J. B., Rich, R. C., Willnat, L., Brians, C. L., & Babb, J. (2012). *Empirical Political Analysis*. Essex, England: Pearson Education Limited.

Manin, B. (1997). *The Principles of Representative Government*. Cambridge: Cambridge University Press.

Manza, J., Cook, F. L., & Page, B. I. (2002, Eds.). *Navigating Public Opinion: Polls, Policy and the Future of American Democracy*. Oxford: Oxford University Press.

Marquis, L., & Sciarini, P. (1999). Opinion Formation in Foreign Policy: the Swiss Experience. *Electoral Studies, 18*(4), 453-471.

Mason, R., Lesser, V., & Traugott, M. W. (2002). Effect of Item Nonresponse on Nonresponse Error and Inference. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey Nonresponse* (pp. 149-161). New York: John Wiley & Sons, Inc.

Mavletova, A. (2013). Data Quality in PC and Mobile Web Surveys. *Social Science Computer Review, 31*(6), 725-743.

Mavletova, A., & Couper, M. P. (2014). Mobile Web Survey Design: Scrolling versus Paging, SMS versus E-mail Invitations. *Journal of Survey Statistics and Methodology, 2*(4), 498-518.

McClendon, M. J. (1986). Unanticipated Effects of No Opinion Filters on Attitudes and Attitude Strength. *Sociological Perspectives, 29*(3), 379-395.

McClendon, M. J., & Alwin, D. F. (1993). No Opinion Filters and Attitude Measurement Reliability. *Sociological Methods & Research, 21*(4), 438-464.

McClosky, H., & Zaller, J. (1984). *The American Ethos: Public Attitudes towards Capitalism and Democracy*. Cambridge, MA: Harvard University Press.

McDermott, R. (2002). Experimental Methods in Political Science. *Annual Review of Political Science, 5*(1), 31-61.

McDermott, R. (2011). Internal and External Validity. In J. N. Druckman, D. P. Green, J. H. Kuklinski, & A. Lupia (Eds.), *Cambridge Handbook of Experimental Political Science* (pp. 27-40). Cambridge: Cambridge University Press.

McFarland, S. G. (1981). Effects of Question Order on Survey Response. *Public Opinion Quarterly, 45*(2), 208-215.

Messer, B. L., Edwards, M. L., & Dillman, D. A. (2012). Determinants of Item Nonresponse to Web and Mail Respondents in Three Address-Based Mixed-Mode Surveys of the General Public. *Survey Practice, 5*(2). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.673.8650&rep=rep1&type=pdf

Mikkers, R. (2013, 3 December). Toestroom beu: grens bereikt met Oost-Europeanen. *De Telegraaf*.

Miller, W. E., & Stokes, D. E. (1963). Constituency Influence in Congress. *American Political Science Review, 57*(1), 45-56.

Moore, D. W. (2002). Measuring New Types of Question-Order Effects: Additive and Subtractive.

Public Opinion Quarterly, 66(1), 80-91.

Moore, D. W. (2003). Americans a Little Less Hawkish on Defense Than Last Year. Retrieved from http://news.gallup.com/poll/7825/americans-little-less-hawkish-defense-than-last-year.aspx

Moore, D. W. (2004). Americans Tepid on Global Warming Accord. Retrieved from http://news.gallup.com/poll/11287/Americans-Tepid-Global-Warming-Accord.aspx

Moore, D. W. (2005). Permissive Opinion on ANWR Leads to Potential Volatility. Retrieved from http://news.gallup.com/poll/15355/Permissive-Opinion-ANWR-Leads-Potential-Volatility.aspx

Moore, D. W. (2008). *The Opinion Makers: An Insider Exposed the Truth Behind the Polls*. Boston: Beacon Press.

Moore, D. W. (2009). "Manipulating" Public Opinion. Retrieved from https://www.huffingtonpost.com/david-moore/manipulating_public_opinion_b_725989.html

Moore, D. W. (2011). Contemporary Issues with Public Policy Polls. *Survey Practice, 4*(4). Retrieved from https://surveypractice.scholasticahq.com/article/3035-contemporary-issues-with-public-policy-polls

Morton, R. B., & Williams, K. C. (2010). *Experimental Political Science and the Study of Causality: From Nature to the Lab*. Cambridge: Cambridge University Press.

Mutz, D. C. (2011). *Population-Based Survey Experiments*. Princeton: Princeton University Press.

Neijens, P., Ridder, J. A. d., & Saris, W. E. (1992). An Instrument for Collecting Informed Opinions. *Quality and Quantity, 26*(3), 245-258.

Newport, F. (2004). *Polling Matters: Why Leaders Must Listen to the Wisdom of the People*. New York: Warner Books.

NKO. (2014). Geschiedenis. Retrieved 15 October, 2014, from http://dpes.nl/nl/nationaal-kiezersonderzoek/geschiedenis

Noelle-Neumann, E. (1974). The Spiral of Silence: A Theory of Public Opinion. *The Journal of Communication, 24* (2), 43-51.

NORC. (2014). Our History and Future. Retrieved 14 October, 2014, from http://www.norc.org/About/Pages/our-history.aspx

NOS. (2012). Wietpas per direct afgeschaft. Retrieved June 30, 2015, from http://nos.nl/artikel/442412-wietpas-per-direct-afgeschaft.html

Opiniepanel, E. (2015). Opiniepanel uitleg. Retrieved July 30th, 2015, from http://opiniepanel.eenvandaag.nl/uitleg

Osborne, T., & Rose, N. (1999). Do the Social Sciences Create Phenomena? The Example of Public Opinion Research. *British Journal of Sociology, 50*(3), 367-396.

Owen, S. V., & Froman, R. D. (1987). What's Wrong with Three-Option Multiple Choice Items? *Educational and Psychological Measurement, 47*(2), 513-522.

Parool, A. R. H. (2011, 16 oktober 2011). 'Vader Maxima welkom bij huldiging Willem-Alexander'. *Het Parool*. Retrieved from http://www.parool.nl/parool/nl/224/BINNENLAND/article/detail/2969801/2011/10/16/Vader-Maxima-welkom-bij-huldiging-Willem-Alexander.dhtml

Payne, S. L. B. (1951). *The Art of Asking Questions*. Princeton: Princeton University Press.

Pellikaan, H. (2010). The Impact of Religion on the Space of Competition: The Dutch Case. *Politics and Religion, 3*(3), 469-494.

Pellikaan, H., de Lange, S. L., & Van der Meer, T. (2007). Fortuyn's legacy: party system change in the Netherlands. *Comparative European Politics, 5*(3), 282-302.

Peytchev, A. (2009). Survey Breakoff. *Public Opinion Quarterly, 73*(1), 74-97.

Peytchev, A. (2013). Consequences of Survey Nonresponse. *The ANNALS of the American Academy of Political and Social Science, 645*(1), 88-111.

Peytchev, A., Couper, M. P., McCabe, S. E., & Crawford, S. D. (2006). Web Survey Design: Paging versus Scrolling. *Public Opinion Quarterly, 70*(4), 596-607.

Poe, G. S., Seeman, I., McLaughlin, J., Mehl, E., & Dietz, M. (1988). Don't Know Boxes in Factual Questions in a Mail Questionnaire: Effects on Level and Quality of Response. *Public Opinion Quarterly, 52*(2), 212-222.

Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., & Singer, E. (2004). Methods for Testing and Evaluating Survey Questions. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for Testing and Evaluating Survey Questionnaires* (pp. 1-22). Hoboken, New Jersey: John Wiley & Sons, Inc.

Presser, S., & McCulloch, S. (2011). The Growth of Survey Research in the United States: Government-Sponsored Surveys, 1984-2004. *Social Science Research, 40*(4), 1019-1024.

Presser, S., & Schuman, H. (1980). The Measurement of a Middle Position in Attitude Surveys. *Public Opinion Quarterly, 44*(1), 70-85.

Price, V. (1992). *Public Opinion*. Newbury Park, London, New Delhi: Sage.

Price, V., & Neijens, P. (1997). Opinion Quality in Public Opinion Research. *International Journal of Public Opinion Research, 9*(4), 336-360.

Przeworski, A., Stokes, S. C., & Manin, B. (1999a). Introduction. In A. Przeworski, S. C. Stokes, & B. Manin (Eds.), *Democracy, Accountability, and Representation* (pp. 1-28). Cambridge: Cambridge University Press.

Przeworski, A., Stokes, S. C., & Manin, B. (1999b). Elections and Representation. In A. Przeworski, S. C. Stokes, & B. Manin (Eds.), *Democracy, Accountability, and Representation* (pp. 29-54). Cambridge: Cambridge University Press.

Quealy, K. (2016, March 22). What Changes Minds About the Senate and Judge Garland? *The New York Times*. Retrieved from http://www.nytimes.com/2016/03/22/upshot/what-changes-minds-about-the-senateandjudge-garland.html?_r=0

Raaijmakers, Q. A. W., van Hoof, A., 't Hart, H., Verbogt, T. F. M. A., & Vollebergh, W. A. M. (2000). Adolescents' midpoint responses on Likert-type scale items: Neutral or missing values? *International Journal for Public Opinion Research, 12*(2), 208-216.

Research Center for Public Opinion Research (2014). History of the Roper Center. Retrieved 15 October, 2014, from www.ropercenter.uconn.edu

Revilla, M. A., & Ochoa, C. (2015). What are the Links in a Web Survey Among Response Time, Quality, and Auto-Evaluation of the Efforts Done? *Social Science Computer Review, 33*(1), 97-114.

Revilla, M. A., Saris, W. E., & Krosnick, J. A. (2014). Choosing the Number of Categories in Agree-Disagree Scales. *Sociological Methods & Research, 43*(1), 73-97.

Robinson, C. E. (1937). Recent Developments in the Straw-Poll Field. *Public Opinion Quarterly, 1*(3), 45-56.

Saris, W. E., & Gallhofer, I. (2007). *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. New York City: John Wiley & Sons, Inc.

Saris, W. E., & Sniderman, P. M. (2004, Eds.). *Studies in Public Opinion: Attitudes, Nonattitudes, Measurement Error, and Change*. New Jersey: Princeton University Press.

Saßenroth, D. (2013). *The Impact of Personality on Participation Decisions in Surveys: A Contribution to the Discussion on Unit Nonresponse*. Heidelberg: Springer.

Scheufele, D. A. & Moy, P. (2000). Twenty-Five Years of the Spiral of Silence: A Conceptual Review and Empirical Outlook. *International Journal of Public Opinion Research 12*(1), 3-28.

Scheuren, F. (2004). *What is a Survey*. Retrieved from https://www.whatisasurvey.info/downloads/pamphlet_current.pdf

Schonlau, M., & Toepoel, V. (2015). Straightlining in Web Survey Panels Over Time. *Survey Research Methods, 9*(2), 125-137.

Schuman, H., & Presser, S. (1979). The Assessment of "No Opinion" in Attitude Surveys. *Sociological Methodology, 10*, 241-275.

Schuman, H., & Presser, S. (1996). *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context*. Thousand Oaks, London, New Delhi: Sage.

Schwarz, N. (2007). Cognitive Aspects of Survey Methodology. *Applied Cognitive Psychology, 21*(2), 277-287.

Schwarz, N., Knäuper, B., Oyserman, D., & Stich, C. (2008). The Psychology of Asking Questions. In E. D. De Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International Handbook of Survey Methodology* (pp. 18-34). London: Lawrence Erlbaum Associates.

Schwarz, N., & Sudman, S. (1996). *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*. San Francisco: Jossey-Bass Publishers.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2006). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, New York: Houghton Mifflin Company.

Shih, T.-H., & Fan, X. (2008). Comparing Response Rates from Web and Mail Surveys: A Meta-Analysis. *Field Methods, 20*(3), 249-271.

Shiraev, E., & Sobel, R. (2006). *People and Their Opinions: Thinking Critically About Public Opinion*. New York: Pearson Education.

Shirky, C. (2011). The Political Power of Social Media. *Foreign Affairs 90*(1), 28-41.

Shoemaker, P. J., Eichholz, M., & Skewes, E. A. (2002). Item Nonresponse: Distinguishing between Don't Know and Refuse. *International Journal for Public Opinion Research, 14*(2), 193-201.

Sigelaman, L. (1981). Question-Order Effects on Presidential Popularity. *Public Opinion Quarterly, 45*(2), 199-207.

Silvera, I. (2015, December 16). EU Referendum: Will Brexit Polls Prove More Accurate Than Derided General Election Ones? *International Business Times*. Retrieved from http://www.ibtimes.co.uk/eu-referendum-will-brexit-polls-prove-more-derided-general-election-ones-1533695

Smith, T. W. (1984). Nonattitudes: A review and evaluation. In C. F. Turner & E. Martin (Eds.), *Surveying subjective phenomena* (Vol. 2, pp. 215-255). New York: Russell Sage Foundation.

Smith, T. W. (1990). The first straw? A study of the origins of election polls. *Public Opinion Quarterly, 54*(1), 21-36.

Smith, T. W. (2011). Refining the Total Survey Error Perspective. *International Journal of Public Opinion Research 23*(4), 464-484.

Smyth, J. D., Dillman, D. A., Christian, L. M., & Stern, M. J. (2006). Comparing Check-All and Forced-Choice Question Formats in Web Surveys. *Public Opinion Quarterly, 70*(1), 66-77.

Smyth, J. D., Dillman, D. A., Christian, L. M., & Stern, M. J. (2006). Effects of Using Visual Design Principles to Group Response Options in Web Surveys. *International Journal of Internet Science, 1*(1), 6-16.

Sniderman, P. M., Tetlock, P. E., & Elms, L. (2001). Public Opinion and Democratic Politics: The Problem of Non-Attitudes and the Social Construction of Political Judgment. In J. H. Kuklinski

(Ed.), *Citizens and Politics: Perspectives from Political Psychology* (pp. 254-288). Cambridge: Cambridge University Press.

Steeh, C. G. (1981). Trends in Nonresponse Rates, 1952-1979. *Public Opinion Quarterly, 45*(1), 40-57.

Steinbrecher, M., Roßmann, J., & Blumenstiel, J. E. (2015). Why Do Respondents Break Off Web Surveys and Does It Matter? Results From Four Follow-up Surveys. *International Journal of Public Opinion Research, 27*(2), 289-302.

Steiner, J. (2010). *The Foundations of Deliberative Democracy: Empirical Research and Normative Implications.* Cambridge: Cambridge University Press.

Stern, M. J., Dillman, D. A., & Smyth, J. D. (2007). Visual Design, Order Effects, and Respondent Characteristics in a Self-Administered Survey. *Survey Research Methods, 1*(3), 121-138.

Stimson, J. A. (1995). Opinion and Representation. *American Political Science Review, 89*(1), 179-183.

Stimson, J. A., Mackuen, M. B., & Erikson, R. S. (1995). Dynamic Representation. *American Political Science Review, 89*(3), 543-565.

Stoop, I. A. L. (2005). *The Hunt for the Last Respondent: Nonresponse in Sample Surveys.* Den Haag: SCP.

Stoop, I. A. L. & Harrison, E. (2012). Classification of Surveys. In L. Gideon (Ed.), *Handbook of Survey Methodology for the Social Sciences* (pp. 7-21). Heidelberg: Springer.

Stoop, I. A. L., & Wittenberg, M. (2008, Eds.). *Access panels and online research, panacea or pitfall? Proceedings of the DANS symposium, Amsterdam, October 12th 2006.* Amsterdam: Uitgeverij Aksant.

Stöss, R. (2009). Die Bedeutung von Item-Nonresponse für die Messung von rechtsextremen Einstellungen. In H. Kaspar, H. Schoen, S. Schumann, & J. R. Winkler (Eds.), *Politik-Wissenschaft-Medien* (pp. 47-66). Wiesbaden: VS Verlag für Sozialwissenschaften.

Streb, M. J. (Ed.). (2007). *Running for Judge: The Rising Political, Financial, and Legal Stakes of Judicial Elections.* New York: New York University Press.

Sturgis, P. (2003). Knowledge and Collective Preferences: A Comparison of Two Approaches to Estimating the Opinions of a Better Informed Public. *Sociological Methods Research, 31*(4), 453-485.

Sudman, S., & Bradburn, N. M. (1974). *Response Effects in Surveys: A Review and a Synthesis.* Chicago: Aldine Publishing Co.

Sudman, S., & Bradburn, N. M. (1989). *Asking Questions: A Practical Guide to Questionnaire Design.* San Francisco: Jossey-Bass.

Sweers, B. M., & Lous, H. (1946). *Vrije Meeningen in een Vrij Land.* Amsterdam: Elsevier.

TeamVier. (2016). Team Vier Panel. Retrieved May 27th 2016, from http://www.teamvier.nl/nl/wie+zijn+we%3F/team+vier+panel

Thomassen, J. (1994). Introduction: The Intellectual History of Election Studies. *European Journal of Political Research, 25*(3), 239-245.

Thomassen, J., & Andeweg, R. B. (2004). Beyond Collective Representation: Individual Members of Parliament and Interest Representation in the Netherlands. *The Journal of Legislative Studies, 10*(4), 47-69.

Tiemeijer, W. L. (2006). *Het Geheim van de Burger: Over Staat en Opinieonderzoek.* Amsterdam: Aksant.

Tiemeijer, W. L. (2008). *Wat 93,7 Procent van de Nederlanders moet weten over Opiniepeilingen.* Amsterdam: Aksant.

TNS-NIPO. (2014). Over TNS-NIPO. Retrieved 15 October, 2014, from www.tns-nipo.com

Toepoel, V. D., Marcel, & Van Soest, A. (2009). Relating Question Type to Panel Conditioning: Comparing Trained and Fresh Respondents. *Survey Research Methods, 3*(2), 73-80.

Toninelli, D., Pinter, R., & de Pedraza, P. (2015). *Mobile Research Methods: Opportunities and Challenges of*

Mobile Research Methodologies. London, United Kingdom: Ubiquity Press.

Tourangeau, R., Conrad, F. G., & Couper, M. P. (2013). *The Science of Web Surveys*. Oxford: Oxford University Press.

Tourangeau, R., Couper, M. P., & Conrad, F. (2004). Spacing, Position, and Order. *Public Opinion Quarterly, 68*(3), 368-393.

Tourangeau, R., Couper, M. P., & Conrad, F. (2007). Color, Labels, and Interpretive Heuristics for Response Scales. *Public Opinion Quarterly, 71*(1), 91-112.

Tourangeau, R., Kreuter, F., & Eckman, S. (2015). Motivated Misreporting: Shaping Answers to Reduce Survey Burden. In U. Engel (Ed.), *Survey Measurements: Techniques, Data Quality and Sources of Error* (pp. 24-41). Frankfurt: Campus Verlag.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

Tourangeau, R., & Yan, T. (2007). Sensitive Questions in Surveys. *Psychological Bulletin, 133*(5), 859-883.

Traugott, M. W. (2004). Do Polls Give the Public a Voice in a Democracy? In M. A. Genovese & M. J. Streb (Eds.), *Polls and Politics: The Dilemmas of Democracy* (pp. 77-94). Albany: State University of New York Press.

Traugott, M. W. (2009). David W. Moore. The Opinion Makers: An Insider Exposes the Truth behind the Polls. *Public Opinion Quarterly, 73*(2), 432-436.

Traugott, M. W., & Lavrakas, P. L. (2007). *Voter's Guide to Election Polls*. Lanham, Maryland: Rowman & Littlefield Publishers, Inc.

Turner, C. F. (1981). Surveys of Subjective Phenomena: A Working Paper. In D. Johnston (Ed.), *Measurement of Subjective Phenomena*. (pp. 37-95). Washington: US Government Printing Office.

Turner, G., Sturgis, P., & Martin, D. (2015). Can Response Latencies Be Used to Detect Survey Satisficing on Cognitively Demanding Questions? *Journal of Survey Statistics and Methodology, 3*(1), 89-108.

van de Maat, J. (2009). *Designing What the Public Thinks?! A Research into the Effects of Filter Questions on the Outcome of Opinion Polls.* (MPhil thesis), Leiden, Leiden.

van der Eijk, C., & Niemöller, K. (1994). Election Studies in the Netherlands: Pluralism and Accommodation. *European Journal of Political Research, 25*(3), 323-342.

van der Meer, T. W. G. (2016). Betere Discussie over Peilingen is Hard Nodig. *Stuk Rood Vlees*. Retrieved from http://stukroodvlees.nl/noodzaak-voor-betere-dialoog-over-peilingen/

van der Meer, T. W. G., Lubbe, R., Van Elsas, E., Elff, M., & Van der Brug, W. (2012). Los Zand of Geëmancipeerde Burgers? Bewegingen op de Nederlandse Kiezersmarkt 2006-2010. In A. E. Bronner, P. Dekker, E. D. De Leeuw, L. J. Paas, K. de Ruyter, A. Smidts, & J. E. Wieringa (Eds.), *Ontwikkelingen in het Marktonderzoek: Jaarboek MarktOnderzoek Associatie deel 37* (pp. 211-228). Haarlem: Spaarenhout.

van der Meer, T. W. G., Van Elsas, E., Lubbe, R., & Van der Brug, W. (2012). Kieskeurige Kiezers: een Panelstudie naar de Veranderlijkheid van Partijvoorkeuren. *Beleid en Maatschappij, 39*(2), 153-178.

van der Veld, W., & Saris, W. E. (2004). Separation of Error, Method Effects, Instability, and Attitude Strength. In W. E. Saris & P. M. Sniderman (Eds.), *Studies in Public Opinion, Attitudes, Nonattitudes, Measurement Error, and Change* (pp. 37-59). New Jersey: Princeton University Press.

van Ewijk, R. (2004). Onderzoek via Telefoon en Internet: de Verschillen. *Clou, 14*, 38-40.

van Ginneken, J. (1993). *De Uitvinding van het Publiek: de Opkomst van het Opinie- en Marktonderzoek in Nederland*. Amsterdam: Otto Cramwinckel Uitgever.

van Ginneken, J. (1999). *Brein-bevingen: Snelle Omslagen in Opinie en Communicatie*. Amsterdam: Boom.

van Holsteyn, J., & Cupido, L. (2013a). *Maar Wat Bedoelt Ie Nou Precies...? Over Politieke Tekeningen en het Correcte Begrip van hun Boodschap*. EenVandaag Opiniepanel.

van Holsteyn, J., & Cupido, L. (2013b, February 28th 2013). Politieke cartoons vaak verkeerd begrepen. Retrieved August 15th, 2016, from http://opiniepanel.eenvandaag.nl/uitslagen/44995/politieke_cartoons_vaak_verkeerd_begrepen

van Keken, K. (2010a, 17 April). Kernenergie: Veel mogelijkheden, veel onzekerheden: Tijd voor een serieuze discussie. *De Volkskrant*.

van Keken, K. (2010b, 17 April). Tweederde van bevolking voor kernenergie. *De Volkskrant*.

van Maarseveen, J. G. S. J., & Schreijnders, R. (Eds.). (1999). *Welgeteld een Eeuw*. Voorburg: CBS.

van Vaerenbergh, Y., & Thomas, T. D. (2013). Response Styles in Survey Research: A Literature Review of Antecedents, Consequences, and Remedies. *International Journal for Public Opinion Research, 25*(3), 195-217.

van Vonno, C. M. C. (2012). Role-Switching in the Dutch Parliament: Reinvigorating Role Theory? *The Journal of Legislative Studies, 18*(2), 119-136.

van Vonno, C. M. C. (2016). *Achieving Party Unity: A Sequential Approach to Why MPs Act in Concert*. (PhD). Retrieved from https://openaccess.leidenuniv.nl/handle/1887/38275

Vannette, D. L., & Krosnick, J. A. (2014). Answering Questions: A Comparison of Survey Satisficing and Mindlessness. In A. Ie, C. T. Ngnouman, & E. J. Langer (Eds.), *The Wiley Blackwell Handbook of Mindlessness* (pp. 312-327): John Wiley & Sons, Ltd.

Verba, S., & Nie, N. H. (1987). *Participation in America: Political Democracy and Social Equality*. Chicago: University of Chicago Press.

Vermeulen, M. (2015, October 24). We worden kapot gepeild: Al dat enqueteren leidt vooral tot chagrijn. *De Volkskrant*. Retrieved from http://www.volkskrant.nl/wetenschap/we-worden-kapot-gepeild-a4169520/

Vicente, P., & Reis, E. (2010). Using Questionnaire Design to Fight Nonresponse Bias in Web Surveys. *Social Science Computer Review, 28*(2), 251-267.

Visser, P. S., Holbrook, A., & Krosnick, J. A. (2008). Knowledge and Attitudes. In W. Donsbach & M. W. Traugott (Eds.), *The SAGE Handbook of Public Opinion Research* (pp. 127-140). Los Angeles: Sage.

Volkskrant, R. d. (2011, 13 oktober 2011). PvdA: vader Maxima niet welkom bij kroning. *de Volkskrant*. Retrieved from http://www.volkskrant.nl/dossier-inhuldiging-willem-alexander/pvda-vader-maxima-niet-welkom-bij-inhuldiging-willem-alexander-a2965616/

Vonk, T., Ossenbruggen, R. v., & Willems, P. (2008). A Comparison Study Across 19 Online Panels. In I. Stoop & M. Wittenberg (Eds.), *Access Panels and Online Research, Panacea or Pitfall?* (pp. 53-78). Amsterdam: Aksant Academic Publishers.

Voogt, R. J., & Van Kempen, H. (2002). Nonresponse Bias and Stimulus Effects in the Dutch National Election Study. *Quality and Quantity, 36*(4), 325-345.

Wagner, J. (2010). The Fraction of Missing Information as a Tool for Monitoring the Quality of Survey Data. *Public Opinion Quarterly 74*(2), 223-243.

Weisberg, H. F. (2005). *The Total Survey Error Approach: A Guide To The New Science of Survey Research*. Chicago: The University of Chicago Press.

Weisberg, H. F. (2008). The Methodological Strengths and Weaknesses of Survey Research. In W. Donsbach & M. W. Traugott (Eds.), *The SAGE Handbook of Public Opinion Research* (pp. 223-232). London: SAGE Publications Ltd.

Willems, M. (2012, November 19). Wietpas per direct afgeschaft - 'Opstelten voert beleid uit waar hij eerder voor waarschuwde'. *NRC Handelsblad*. Retrieved from http://www.nrc.nl/nieuws/2012/11/19/wietpas-per-direct-afgeschaft/

Wittkopf, E. R. (1990). *Faces of Internationalism: Public Opinion and American Foreign Policy*. Durham, NC: Duke University Press.

Worcester, R. M. (1987). The Internationalization of Public Opinion Research. *Public Opinion Quarterly, 51*(2), S79-S85.

Yan, T., & Tourangeau, R. (2008). Fast Times and Easy Questions: The Effects of Age, Experience and Question Complexity on Web Survey Response Times. *Applied Cognitive Psychology, 22*, 51-68.

Yankelovich, D. (1991). *Coming to Public Judgement: Making Democracy Work in a Complex World*. Syracuse: Syracuse University Press.

Zaller, J. R. (1992). *The Nature and Origins of Mass Opinion* (13th ed.). Cambridge: Cambridge University Press.

Zaller, J. R., & Feldman, S. (1992). A Simple Theory of the Survey Response: Answering Questions versus Revealing Preferences. *American Journal of Political Science, 36*(3), 579-616.

Zhang, C., & Conrad, F. G. (2013). Speeding in Web Surveys: The Tendency to Answer Very Fast and its Association with Straightlining. *Survey Research Methods, 8*(2), 127-135.

APPENDIX A   Questionnaires
(in Dutch)
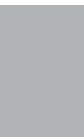
I   Nederlandse Vragenlijst Experiment 1 (DK / Weet Niet)

- Variant 1: 'Geen mening' <u>dubbel expliciet</u> in stam van de vraag en de antwoord-categorieën aangeboden.
- Variant 2: 'Geen mening' <u>enkel expliciet</u> in de antwoordcategorieën aangeboden.
- Variant 3: 'Geen mening' wordt <u>enkel impliciet</u> aangeboden: de respondent kan het antwoordvakje leeglaten.
- Variant 4: 'Geen mening' wordt <u>niet</u> aangeboden en de respondent moet een antwoord aanvinken.

Hieronder vindt u een aantal stellingen over uiteenlopende onderwerpen. Geef aan wat u van deze stellingen vindt door het vakje aan te vinken dat het beste bij uw mening past. *(U kunt ook aangeven dat u geen mening heeft over een bepaalde stelling).*

Sociaal-economisch

Sommige mensen en partijen vinden dat de verschillen in inkomens in ons land groter moeten worden. Anderen dat ze kleiner moeten worden. Natuurlijk zijn er ook mensen met een mening die daar tussenin ligt[1].

Waar zou u zichzelf plaatsen op een lijn van 1 tot en met 7, waarbij de 1 betekent dat inkomensverschillen groter moeten worden en de 7 dat ze kleiner moeten worden? *Als u helemaal niet weet waar u zichzelf zou plaatsen, geeft u dat dan gerust aan.*
- 1. De verschillen in inkomens in ons land moeten groter worden.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7. De verschillen in inkomens in ons land moeten kleiner worden.
- 8. *Weet niet.*

---

1   Nationaal Kiezersonderzoek 2010, vraagteksten en schema's 1e golf pagina 26.

De AOW-leeftijd moet gehandhaafd blijven op 65 jaar. Bent u het hiermee eens, mee oneens *of heeft u geen mening*?
- Volledig mee eens
- Mee eens
- Mee oneens
- Volledig mee oneens
- *Geen mening*

De bijstand moet verlaagd worden zodat mensen gestimuleerd worden om te gaan werken. Bent u het hiermee eens, mee oneens *of heeft u geen mening*?
- Volledig mee eens
- Mee eens
- Mee oneens
- Volledig mee oneens
- *Geen mening*

Ethisch / Religieus

Sommige mensen vinden dat euthanasie altijd verboden moet zijn. Anderen vinden dat euthanasie mogelijk moet zijn als de patiënt daarom vraagt. Natuurlijk zijn er ook mensen met een mening die daar tussenin ligt[2].

Waar zou u uzelf plaatsen op een lijn van 1 tot en met 7, waarbij de 1 betekent dat euthanasie moet worden verboden en de 7 dat euthanasie mogelijk moet zijn? *Als u helemaal niet weet waar u zichzelf zou plaatsen, geeft u dat dan gerust aan.*
- 1. Euthanasie moet verboden zijn.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7. De arts mag altijd euthanasie toepassen als de patiënt daarom vraagt.
- 8. Weet niet.

---

2   Idem.

Adoptie door homoseksuele paren moet mogelijk zijn[3]. Bent u het hier mee eens, mee oneens *of heeft u geen mening*?
- Volledig mee eens
- Mee eens
- Mee oneens
- Volledig mee oneens
- *Geen mening*

Het is goed dat vrouwen hun eicellen kunnen laten invriezen om op latere leeftijd nog kinderen te kunnen krijgen. Bent u het hier mee eens, mee oneens *of heeft u geen mening*[4]?
- Volledig mee eens
- Mee eens
- Mee oneens
- Volledig mee oneens
- *Geen mening*

Multicultureel / Homogeen

In Nederland vinden sommigen dat allochtonen hier moeten kunnen leven met behoud van de eigen cultuur. Anderen vinden dat zij zich geheel moeten aanpassen aan de Nederlandse cultuur[5].

Waar zou u uzelf plaatsen op een lijn van 1 tot en met 7, waarbij de 1 behoud van eigen cultuur voor allochtonen betekent en de 7 dat zij zich geheel moeten aanpassen? *Als u helemaal niet weet waar u zichzelf zou plaatsen, geeft u dat dan gerust aan.*
- 1. Etnische minderheden moeten in Nederland kunnen leven met behoud van alle gewoonten van de eigen cultuur.
- 2.
- 3.
- 4.
- 5.
- 6.

---

3   Nationaal Kiezersonderzoek 2010, vraagteksten en schema's 1e golf pagina 32.
4   This question is a replication of a question in the pilot experiment (van de Maat, 2009).
5   Nationaal Kiezersonderzoek 2010, vraagteksten en schema's 1e golf pagina 30.

- 7. Etnische minderheden moeten zich volledig aanpassen aan de Nederlandse cultuur.
- 8. *Weet niet.*

Er wonen teveel mensen van een andere nationaliteit in Nederland[6]. Bent u het hier mee eens, mee oneens *of heeft u geen mening?*

- Volledig mee eens
- Mee eens
- Mee oneens
- Volledig mee oneens
- *Geen mening*

Alle illegalen die al lange tijd in Nederland wonen, moeten hier kunnen blijven[7]. Bent u het hiermee eens, mee oneens *of heeft u geen mening?*

- Volledig mee eens
- Mee eens
- Mee oneens
- Volledig mee oneens
- *Geen mening*

Buitenlandse zaken

De Europese eenwording is in volle gang. De landen van de Europese Unie hebben besloten om steeds nauwer met elkaar te gaan samenwerken. Maar niet iedereen denkt hier hetzelfde over. Sommige mensen en partijen vinden dat de Europese eenwording nog verder zou moeten gaan. Anderen vinden dat de Europese eenwording al te ver is gegaan[8].

Stel dat de mensen en partijen die vinden dat de Europese eenwording nog verder zou moeten gaan aan het begin van de lijn staan (bij cijfer 1) en dat de mensen en partijen die vinden dat de eenwording al te ver is gegaan aan het einde van de lijn staan (bij cijfer 7). Waar zou u zichzelf plaatsen op deze lijn? *Als u helemaal niet weet waar u zichzelf zou plaatsen, geeft u dat dan gerust aan.*

---

6   Jaarrapport Integratie 2009: 259.
7   Nationaal Kiezersonderzoek 2010, vraagteksten en schema's 1e golf pagina 32.
8   Nationaal Kiezersonderzoek 1998, vraag 23g

- 1. De Europese eenwording zou nog verder moeten gaan.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7. De Europese eenwording is al veel te ver gegaan.
- 8. *Weet niet.*

Nederland moet meer geld uitgeven aan ontwikkelingshulp[9]. Bent u het hiermee eens, mee oneens *of heeft u geen mening*?
- Volledig mee eens
- Mee eens
- Mee oneens
- Volledig mee oneens
- *Geen mening*

De Verenigde Naties heeft te weinig macht. Bent u het hiermee eens, mee oneens *of heeft u geen mening*?
- Volledig mee eens
- Mee eens
- Mee oneens
- Volledig mee oneens
- *Geen mening*

Actueel

De Koningin mag tegenover journalisten slechts het regeringsbeleid uitdragen. Bent u het hiermee eens, mee oneens *of heeft u geen mening*?
- Mee eens
- Mee oneens
- Geen mening

Bent u voor of tegen draagmoederschap *of heeft u geen mening*?
- Voor
- Tegen
- *Geen mening*

---

9   Nationaal Kiezersonderzoek 2010, vraagteksten en schema's 1e golf pagina 32.

II   Nederlandse Vragenlijst Experiment 2 (Filtervraag)

Het gebruik van filtervragen en 'weet niet' opties varieert.

- **Variant 1a: filtervraag in sterke formulering met expliciet 'weet niet' antwoordalternatief**
- **Variant 1b: filtervraag in sterke formulering zonder expliciet 'weet niet' alternatief**
- **Variant 2a: filtervraag in minder sterke formulering met expliciet 'weet niet' antwoordalternatief**
- **Variant 2b: filtervraag in minder sterke formulering zonder expliciet 'weet niet' antwoordalternatief**
- **Variant 3a: geen filtervraag, met expliciet 'weet niet' antwoordalternatief**
- **Variant 3b: geen filtervraag, zonder expliciet 'weet niet' antwoordalternatief**
- **Variant 4: geen filtervraag en geen 'weet niet' antwoordalternatief, de respondent moet antwoorden**

Sociaal-economisch

Sommige mensen en partijen vinden dat de verschillen in inkomens in ons land groter moeten worden. Anderen dat ze kleiner moeten worden. Natuurlijk zijn er ook mensen met een mening die daar tussenin ligt[10].

Heeft u hier voldoende over gehoord, gelezen en nagedacht om hierover een mening te kunnen geven?
- Ja > *Ga verder met de inhoudelijke vraag.*
- Nee > *Sla de inhoudelijke vraag over en ga door met het volgende onderwerp.*

Waar zou u zichzelf plaatsen op een lijn van 1 tot en met 7, waarbij de 1 betekent dat inkomensverschillen groter moeten worden en de 7 dat ze kleiner moeten worden?
- 1. De verschillen in inkomens in ons land moeten groter worden.
- 2.
- 3.
- 4.
- 5.

---

10   Nationaal Kiezersonderzoek 2010, vraagteksten en schema's 1e golf pagina 26.

- 6.
- 7. De verschillen in inkomens in ons land moeten kleiner worden.
- 8. *Weet niet.*

Er wordt heel verschillend gedacht over de bijstand en de hoogte hiervan. Heeft u hier voldoende over gehoord, gelezen en nagedacht om hierover een mening te kunnen geven?
- Ja > *Ga verder met de inhoudelijke vraag.*
- Nee > *Sla de inhoudelijke vraag over en ga door met het volgende onderwerp.*

'De bijstand moet verlaagd worden zodat mensen gestimuleerd worden om te gaan werken'. Bent u het hiermee eens of mee oneens?
- Volledig mee eens
- Mee eens
- Mee oneens
- Volledig mee oneens
- *Weet niet*

Ethisch / Religieus

Sommige mensen vinden dat euthanasie altijd verboden moet zijn. Anderen vinden dat euthanasie mogelijk moet zijn als de patiënt daarom vraagt. Natuurlijk zijn er ook mensen met een mening die daar tussenin ligt[11].

Heeft u hier voldoende over gehoord, gelezen en nagedacht om hierover een mening te kunnen geven?
- Ja > *Ga verder met de inhoudelijke vraag.*
- Nee > *Sla de inhoudelijke vraag over en ga door met het volgende onderwerp.*

Waar zou u uzelf plaatsen op een lijn van 1 tot en met 7, waarbij de 1 betekent dat euthanasie moet worden verboden en de 7 dat euthanasie mogelijk moet zijn?
- 1. Euthanasie moet verboden zijn.
- 2.
- 3.
- 4.
- 5.

---

11   Idem.

- 6.
- 7. De arts mag altijd euthanasie toepassen als de patiënt daarom vraagt.
- 8. *Weet niet.*

Er wordt heel verschillend gedacht over de adoptie van kinderen door homoseksuele paren. Heeft u hier voldoende over gehoord, gelezen en nagedacht om hierover een mening te kunnen geven?
- Ja > *Ga verder met de inhoudelijke vraag.*
- Nee > *Sla de inhoudelijke vraag over en ga door met het volgende onderwerp.*

'Adoptie door homoseksuele paren moet mogelijk zijn'[12]. Bent u het hiermee eens of mee oneens?
- Volledig mee eens
- Mee eens
- Mee oneens
- Volledig mee oneens
- *Weet niet*

Multicultureel / Homogeen

In Nederland vinden sommigen dat allochtonen hier moeten kunnen leven met behoud van de eigen cultuur. Anderen vinden dat zij zich geheel moeten aanpassen aan de Nederlandse cultuur[13].

Heeft u hier voldoende over gehoord, gelezen en nagedacht om hierover een mening te kunnen geven?
- Ja > *Ga verder met de inhoudelijke vraag.*
- Nee > *Sla de inhoudelijke vraag over en ga door met het volgende onderwerp.*

Waar zou u uzelf plaatsen op een lijn van 1 tot en met 7, waarbij de 1 behoud van eigen cultuur voor allochtonen betekent en de 7 dat zij zich geheel moeten aanpassen?
- 1. Etnische minderheden moeten in Nederland kunnen leven met behoud van alle gewoonten van de eigen cultuur.
- 2.

---

12   Nationaal Kiezersonderzoek 2010, vraagteksten en schema's 1e golf pagina 32.
13   Nationaal Kiezersonderzoek 2010, vraagteksten en schema's 1e golf pagina 30.

- 3.
- 4.
- 5.
- 6.
- 7. Etnische minderheden moeten zich volledig aanpassen aan de Nederlandse cultuur.
- 8. *Weet niet.*

Er wordt heel verschillend gedacht over de aanwezigheid van andere nationaliteiten in Nederland. Heeft u hier voldoende over gehoord, gelezen en nagedacht om hierover een mening te kunnen geven?

- Ja > *Ga verder met de inhoudelijke vraag.*
- Nee > *Sla de inhoudelijke vraag over en ga door met het volgende onderwerp.*

'Er wonen teveel mensen van een andere nationaliteit in Nederland'[14]. Bent u het hiermee eens of mee oneens?

- Volledig mee eens
- Mee eens
- Mee oneens
- Volledig mee oneens
- *Weet niet*

Buitenlandse zaken

De Europese eenwording is in volle gang. De landen van de Europese Unie hebben besloten om steeds nauwer met elkaar te gaan samenwerken. Maar niet iedereen denkt hier hetzelfde over. Sommige mensen en partijen vinden dat de Europese eenwording nog verder zou moeten gaan. Anderen vinden dat de Europese eenwording al te ver is gegaan[15].

Heeft u hier voldoende over gehoord, gelezen en nagedacht om hierover een mening te kunnen geven?

- Ja > *Ga verder met de inhoudelijke vraag.*
- Nee > *Sla de inhoudelijke vraag over en ga door met het volgende onderwerp.*

---

14   Jaarrapport Integratie 2009: 259.
15   Nationaal Kiezersonderzoek 1998, vraag 23g

Stel dat de mensen en partijen die vinden dat de Europese eenwording nog verder zou moeten gaan aan het begin van de lijn staan (bij cijfer 1) en dat de mensen en partijen die vinden dat de eenwording al te ver is gegaan aan het einde van de lijn staan (bij cijfer 7). Waar zou u zichzelf plaatsen op deze lijn?

- 1. De Europese eenwording zou nog verder moeten gaan.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7. De Europese eenwording is al veel te ver gegaan.
- 8. *Weet niet.*

Er wordt heel verschillend gedacht over ontwikkelingshulp. Heeft u hier voldoende over gehoord, gelezen en nagedacht om hierover een mening te kunnen geven?

- Ja > *Ga verder met de inhoudelijke vraag.*
- Nee > *Sla de inhoudelijke vraag over en ga door met het volgende onderwerp.*

'Nederland moet meer geld uitgeven aan ontwikkelingshulp'[16]. Bent u het hiermee eens of mee oneens?

- Volledig mee eens
- Mee eens
- Mee oneens
- Volledig mee oneens
- *Weet niet*

Er wordt heel verschillend gedacht over de hypotheekrenteaftrek. Heeft u hier voldoende over gehoord, gelezen en nagedacht om hierover een mening te kunnen geven?

- Ja > *Ga verder met de inhoudelijke vraag.*
- Nee > *Sla de inhoudelijke vraag over en ga door met het volgende onderwerp.*

Wat moet er volgens u met de hypotheekrenteaftrek gebeuren?

- Die moet volledig gehandhaafd blijven
- Die moet op termijn worden beperkt
- Die moet op termijn volledig worden afgeschaft
- *Weet niet*

---

16   Nationaal Kiezersonderzoek 2010, vraagteksten en schema's 1e golf pagina 32.

Er wordt heel verschillend gedacht over de dreigende tekorten van pensioenfondsen. Heeft u hier voldoende over gehoord, gelezen en nagedacht om hierover een mening te kunnen geven?

- Ja > *Ga verder met de inhoudelijke vraag.*
- Nee > *Sla de inhoudelijke vraag over en ga door met het volgende onderwerp.*

Wat vindt u de beste oplossing voor de dreigende tekorten van pensioenfondsen?

- Korten op de pensioenuitkering van pensioengerechtigden
- Verhoging van de pensioenpremies van mensen die nu werken
- *Weet niet*

Er wordt heel verschillend gedacht over de euro. Heeft u hier voldoende over gehoord, gelezen en nagedacht om hierover een mening te kunnen geven?

- Ja > *Ga verder met de inhoudelijke vraag.*
- Nee > *Sla de inhoudelijke vraag over en ga door met het volgende onderwerp.*

'Nederland moet het komende jaar stoppen met de euro en terug naar de gulden'. Bent u het hiermee eens of mee oneens?

- Mee eens
- Mee oneens
- *Weet niet*

Er wordt heel verschillend gedacht over de invloed van de koning(in). Heeft u hier voldoende over gehoord, gelezen en nagedacht om hierover een mening te kunnen geven?

- Ja > *Ga verder met de inhoudelijke vraag.*
- Nee > *Sla de inhoudelijke vraag over en ga door met het volgende onderwerp.*

Vindt u dat de koning(in) van Nederland ook politieke invloed moet hebben, of moet hij/zij zich beperken tot ceremoniële functies?

- De koning(in) moet ook politieke invloed hebben
- De koning(in) moet zich beperken tot ceremoniële functies
- *Weet niet*

Er wordt heel verschillend gedacht over de aanwezigheid van Maxima's vader bij de kroning. Heeft u hier voldoende over gehoord, gelezen en nagedacht om hierover een mening te kunnen geven?

- Ja > *Ga verder met de inhoudelijke vraag.*
- Nee > *Sla de inhoudelijke vraag over en ga door met het volgende onderwerp.*

Vindt u dat de vader van Maxima wel of niet aanwezig mag zijn bij de kroning?
- Wel
- Niet
- *Weet niet*

Er wordt heel verschillend gedacht over Libië. Heeft u hier voldoende over gehoord, gelezen en nagedacht om hierover een mening te kunnen geven[17]?
- Ja > *Ga verder met de inhoudelijke vraag.*
- Nee > *Sla de inhoudelijke vraag over en ga door met het volgende onderwerp.*
- 'Ik denk dat Libië op den duur een normaal democratisch land zal worden'.

Bent u het hiermee eens of mee oneens?
- Mee eens
- Mee oneens
- *Weet niet*

'Nederland moet actief helpen om in Libië een democratisch bewind te vestigen'. Bent u het hiermee eens of mee oneens?
- Mee eens
- Mee oneens
- *Weet niet*

Er wordt heel verschillend gedacht over hoe lang het kabinet blijft zitten. Heeft u hier voldoende over gehoord, gelezen en nagedacht om hierover een mening te kunnen geven[18]?

Ja > *Ga verder met de inhoudelijke vraag.*

Nee > *Sla de inhoudelijke vraag over en ga door met het volgende onderwerp.*

Hoe lang denkt u dat dit kabinet blijft zitten?
- Valt dit jaar
- Valt in 2012
- Valt in 2013 of 2014
- Zit de rit uit
- *Weet niet*

---

17 Deze filtervraag wordt gevolgd door twee opinievragen.
18 Deze filtervraag wordt gevolgd door twee opinievragen.

Stel dat volgend jaar nog eens 5 miljard euro extra bezuinigd moet worden, denkt u dan dat de PVV het kabinet laat vallen?

- Ja
- Nee
- *Weet niet*

III   Nederlandse Vragenlijst Experiment 3 (Vervolgvraag)

- Variant 1a: geen filtervraag, wel vervolgvraag, expliciet 'weet niet' antwoordalternatief
- Variant 1b: geen filtervraag, wel vervolgvraag, impliciet 'weet niet' antwoordalternatief
- Variant 1c: geen filtervraag, wel vervolgvraag, geen 'weet niet' antwoordalternatief
- Variant 2a: filtervraag, wel vervolgvraag, expliciet 'weet niet' antwoordalternatief
- Variant 2b: filtervraag, wel vervolgvraag, impliciet 'weet niet' antwoordalternatief
- Variant 2c: filtervraag, wel vervolgvraag, geen 'weet niet' antwoordalternatief

Sociaal-economisch

Sommige mensen en partijen vinden dat de verschillen in inkomens in ons land groter moeten worden. Anderen dat ze kleiner moeten worden. Natuurlijk zijn er ook mensen met een mening die daar tussenin ligt[19].

Waar zou u zichzelf plaatsen op een lijn van 1 tot en met 7, waarbij de 1 betekent dat inkomensverschillen groter moeten worden en de 7 dat ze kleiner moeten worden?
- 1. De verschillen in inkomens in ons land moeten groter worden.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7. De verschillen in inkomens in ons land moeten kleiner worden.
- Weet niet

Hoe erg zou u het vinden als uw mening over verschillen in inkomens niet door de politiek in beleid wordt omgezet?
- Heel erg
- Erg
- Niet zo erg
- Helemaal niet erg

---

19   Nationaal Kiezersonderzoek 2010, vraagteksten en schema's 1e golf pagina 26.

'De bijstand moet verlaagd worden zodat mensen gestimuleerd worden om te gaan werken'. Bent u het hiermee eens of mee oneens?
- Volledig mee eens
- Mee eens
- Mee oneens
- Volledig mee oneens
- Weet niet

Hoe erg zou u het vinden als uw mening over verlaging van de bijstand niet door de politiek in beleid wordt omgezet?
- Heel erg
- Erg
- Niet zo erg
- Helemaal niet erg

Ethisch / Religieus

Sommige mensen vinden dat euthanasie altijd verboden moet zijn. Anderen vinden dat euthanasie mogelijk moet zijn als de patiënt daarom vraagt. Natuurlijk zijn er ook mensen met een mening die daar tussenin ligt[20].

Waar zou u uzelf plaatsen op een lijn van 1 tot en met 7, waarbij de 1 betekent dat euthanasie moet worden verboden en de 7 dat euthanasie mogelijk moet zijn?
- 1. Euthanasie moet verboden zijn.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7. De arts mag altijd euthanasie toepassen als de patiënt daarom vraagt.
- Weet niet

Hoe erg zou u het vinden als uw mening over euthanasie niet door de politiek in beleid wordt omgezet?
- Heel erg
- Erg

---

20   Idem.

- Niet zo erg
- Helemaal niet erg

'Adoptie door homoseksuele paren moet mogelijk zijn'[21]. Bent u het hiermee eens of mee oneens?
- Volledig mee eens
- Mee eens
- Mee oneens
- Volledig mee oneens
- Weet niet

Hoe erg zou u het vinden als uw mening over adoptie door homoseksuele paren niet door de politiek in beleid wordt omgezet?
- Heel erg
- Erg
- Niet zo erg
- Helemaal niet erg

Multicultureel / Homogeen

In Nederland vinden sommigen dat allochtonen hier moeten kunnen leven met behoud van de eigen cultuur. Anderen vinden dat zij zich geheel moeten aanpassen aan de Nederlandse cultuur[22].

Waar zou u uzelf plaatsen op een lijn van 1 tot en met 7, waarbij de 1 behoud van eigen cultuur voor allochtonen betekent en de 7 dat zij zich geheel moeten aanpassen?
- 1. Etnische minderheden moeten in Nederland kunnen leven met behoud van alle gewoonten van de eigen cultuur.
- 2.
- 3.
- 4.
- 5.
- 6.

---

21   Nationaal Kiezersonderzoek 2010, vraagteksten en schema's 1e golf pagina 32.
22   Nationaal Kiezersonderzoek 2010, vraagteksten en schema's 1e golf pagina 30.

- 7. Etnische minderheden moeten zich volledig aanpassen aan de Nederlandse cultuur.
- Weet niet

Hoe erg zou u het vinden als uw mening over de integratie van allochtonen niet door de politiek in beleid wordt omgezet?
- Heel erg
- Erg
- Niet zo erg
- Helemaal niet erg

'Er wonen teveel mensen van een andere nationaliteit in Nederland'[23]. Bent u het hiermee eens of mee oneens?
- Volledig mee eens
- Mee eens
- Mee oneens
- Volledig mee oneens
- Weet niet

Hoe erg zou u het vinden als uw mening over mensen met een niet-Nederlandse nationaliteit niet door de politiek in beleid wordt omgezet?
- Heel erg
- Erg
- Niet zo erg
- Helemaal niet erg

Buitenlandse zaken

De Europese eenwording is in volle gang. De landen van de Europese Unie hebben besloten om steeds nauwer met elkaar te gaan samenwerken. Maar niet iedereen denkt hier hetzelfde over. Sommige mensen en partijen vinden dat de Europese eenwording nog verder zou moeten gaan. Anderen vinden dat de Europese eenwording al te ver is gegaan[24].

---

23   Jaarrapport Integratie 2009: 259.
24   Nationaal Kiezersonderzoek 1998, vraag 23g

Stel dat de mensen en partijen die vinden dat de Europese eenwording nog verder zou moeten gaan aan het begin van de lijn staan (bij cijfer 1) en dat de mensen en partijen die vinden dat de eenwording al te ver is gegaan aan het einde van de lijn staan (bij cijfer 7). Waar zou u zichzelf plaatsen op deze lijn?

- 1. De Europese eenwording zou nog verder moeten gaan.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7. De Europese eenwording is al veel te ver gegaan.
- Weet niet

Hoe erg zou u het vinden als uw mening over de Europese eenwording niet door de politiek in beleid wordt omgezet?

- Heel erg
- Erg
- Niet zo erg
- Helemaal niet erg

'Nederland moet meer geld uitgeven aan ontwikkelingshulp'[25]. Bent u het hiermee eens of mee oneens?

- Volledig mee eens
- Mee eens
- Mee oneens
- Volledig mee oneens
- Weet niet

Hoe erg zou u het vinden als uw mening over de Nederlandse uitgaven aan ontwikkelingshulp niet door de politiek in beleid wordt omgezet?

- Heel erg
- Erg
- Niet zo erg
- Helemaal niet erg

---

25   Nationaal Kiezersonderzoek 2010, vraagteksten en schema's 1e golf pagina 32.

Actueel

'Het gebruik van softdrugs moet volledig verboden worden'[26]. Bent u het hiermee eens of mee oneens?

- Eens
- Oneens
- Weet niet

Hoe erg zou u het vinden als uw mening over softdrugs niet door de politiek in beleid wordt omgezet?

- Heel erg
- Erg
- Niet zo erg
- Helemaal niet erg

'De instelling van de Wietpas vind ik een goed idee'. Bent u het hiermee eens of mee oneens?

- Eens
- Oneens
- Weet niet

Hoe erg zou u het vinden als uw mening over de Wietpas niet door de politiek in beleid wordt omgezet?

- Heel erg
- Erg
- Niet zo erg
- Helemaal niet erg

Welke van onderstaande stellingen komt het dichtst bij uw mening[27]?

- De Europese Unie moet meer bevoegdheden krijgen.
- De EU moet niet meer, ook niet minder bevoegdheden krijgen.
- Nationale staten moeten meer bevoegdheden krijgen.
- Weet niet.

---

26   Peil.nl, 16 mei 2012.
27   TNS Nipo, 1 juni 2012.

Hoe erg zou u het vinden als uw mening over de bevoegdheden van de EU niet door de politiek in beleid wordt omgezet?
- Heel erg
- Erg
- Niet zo erg
- Helemaal niet erg

In hoeverre bent u het eens met de volgende stelling: 'Nederland moet de hypotheekrenteaftrek in zijn geheel afschaffen'[28]?
- Helemaal mee eens
- Mee eens
- Mee eens noch mee oneens
- Mee oneens
- Helemaal mee oneens
- Weet niet

Hoe erg zou u het vinden als uw mening over de hypotheekrenteaftrek niet door de politiek in beleid wordt omgezet?
- Heel erg
- Erg
- Niet zo erg
- Helemaal niet erg

'Ik wil wel dat het boerkaverbod doorgaat'[29]. Bent u het hiermee eens of mee oneens?
- Eens
- Oneens
- Weet niet

Hoe erg zou u het vinden als uw mening over het boerkaverbod niet door de politiek in beleid wordt omgezet?
- Heel erg
- Erg
- Niet zo erg
- Helemaal niet erg

---

28   Ipsos Synovate, 4 juni 2012.
29   Peil.nl, 3 mei 2012.

'De NAVO moet in Syrië ingrijpen[30]'. Bent u het hiermee eens of mee oneens?
- Eens
- Oneens
- Weet niet

Hoe erg zou u het vinden als uw mening over NAVO-ingrijpen in Syrië niet door de politiek in beleid wordt omgezet?
- Heel erg
- Erg
- Niet zo erg
- Helemaal niet erg

---

30 Peil.nl, 2 juni 2012.

More Analyses
Response Time
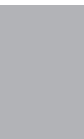
Is response time related to question content? To answer this question with the data from chapter 5, Table B.1 details the response times of the 14 individual items. See Table B.2 for the significance of between-variant-differences in average response time, analysed with a t-test.

Table B.1: Response Time for Individual Questions

	Average Response Time	Standard Deviation	Minimum[31] Response Time	Number of Missings[32]	N
Self-placement Income	23.3	18.4	1.8	119	5609
Old-Age Pension	11.0	9.5	2.0	14	5714
Lowering of Welfare Benefits	12.7	11.7	0.3	19	5709
Self-placement Euthanasia	15.4	12.3	1.6	19	5709
Same-sex Adoption	10.9	9.9	1.6	9	5719
Ova Freezing	11.7	10.7	0.8	14	5714
Self-placement Foreigners	18.3	14.7	1.5	29	5699
Number of non-Dutch	11.0	9.0	1.5	13	5715
Illegal in Netherlands	12.0	10.1	1.2	6	5722
Self-placement EU	21.3	18.3	1.4	39	5689
Developmental Aid	10.0	8.0	1.6	10	5718
UN Power	10.4	10.0	0.7	16	5712
Role Queen	12.6	11.2	1.3	14	5714
Surrogate Motherhood	8.5	7.9	0.3	10	5718

---

31 No maximum is shown here, since 180 was the cut-off point. Out of 5728 participants, the response time of 282 respondents (4.92 percent) was excluded from the analysis for one or more individual questions.

32 Usually the missings are caused by the 180 seconds cut-off point and not because of registered negative values. For the first question, however, both types of missings were relatively high with 26 negative values and 93 other outliers (over 180 seconds).

Table B.2: Significance Between-Variants-Differences in Average Response Time

	1/2	1/3	1/4	2/3	2/4	3/4
Self-placement Income	.084	.001	.000	.097	.055	.793
Old-Age Pension	.078	.000	.005	.027	.301	.254
Lowering of Welfare Benefits	.486	.018	.003	.099	.031	.706
Self-placement Euthanasia	.095	.051	.166	.735	.783	.546
Same-sex adoption	.024	.037	.043	.971	.958	.936
Ova Freezing	.149	.009	.002	.212	.079	.625
Self-placement Foreigners	.017	.057	.035	.710	.901	.820
Number of non-Dutch	.002	.002	.000	.729	.094	.244
Illegal in Netherlands	.093	.001	.000	.118	.020	.410
Self-placement EU	.657	.900	.311	.742	.135	.237
Developmental Aid	.096	.000	.000	.000	.000	.926
UN Power	.004	.525	.848	.013	.004	.642
Role Queen	.006	.321	.053	.119	.499	.389
Surrogate Motherhood	.000	.000	.000	.243	.204	.954

The response time for an individual question on average ranges from 8.5 seconds (for Surrogate Motherhood) to 23.3 seconds for Self-Placement Income. This coincides with the last and the first item in the questionnaire and seems to be in line with Callegaro *et al* (2009) who found that the item response time decreased during the survey, flattening out towards the end (a graphic representation can be found in Figure B.1). The main point, however, is that there are substantial differences between the response times for individual items.

Both the number of missings and the response time can be considered indications of a question being more difficult to answer. The response time is a direct measure of a respondent needing more time to read the question and to decide on an answer. So the higher the number of missings and the amount of response time, the more difficult it apparently was to answer that particular question. In Table B.1, the issues with the most missings and the highest response times are the same: self-placement questions. Self-Placement Income, Self-Placement Foreigners and Self-Placement EU were the items with the highest numbers of missings, between 29 and 119, and the highest average response time ranging from 18.3 to 22.3 seconds per question. These are all self-placement items with a relatively long introduction and many response categories in comparison to the other items; it makes sense that more cognitive effort is required to read and process the question and decide on an answer. Since the fourth highest response time of 15.4 seconds is registered for the only other self-placement question, Self-Placement Euthanasia, the tentative

conclusion is that the length of introduction and the form of the question is more decisive than question content. An analysis of average response time and the number of words in the question (excluding response options)[33] indeed reveals a strong correlation of .89 significant at the .001 level. Longer questions require more response time.

Table B.1 includes a depiction of whether the average item response time differed significantly between the four variants of the questionnaire; the differences between individual items are not shown. Figure B.1 shows the item response times of the four variants during the survey progress and shows a gradual decrease towards the end of the survey. These trends confirm that the item response rate decreases (see also Callegaro et al., 2009; Galesic & Bosnjak, 2009). Note also that the explicit DK variants result in slightly higher response times for all items as compared to the single implicit and forced choice questionnaires. Furthermore, in 13 out of 14 items, the exception being question 10 Self-Placement EU, the double explicit DK variant resulted in a higher response time than the single explicit variant - see the pair-wise comparisons of significance. Finally, no response time trend can be discerned when comparing the single implicit DK and forced choice variant.
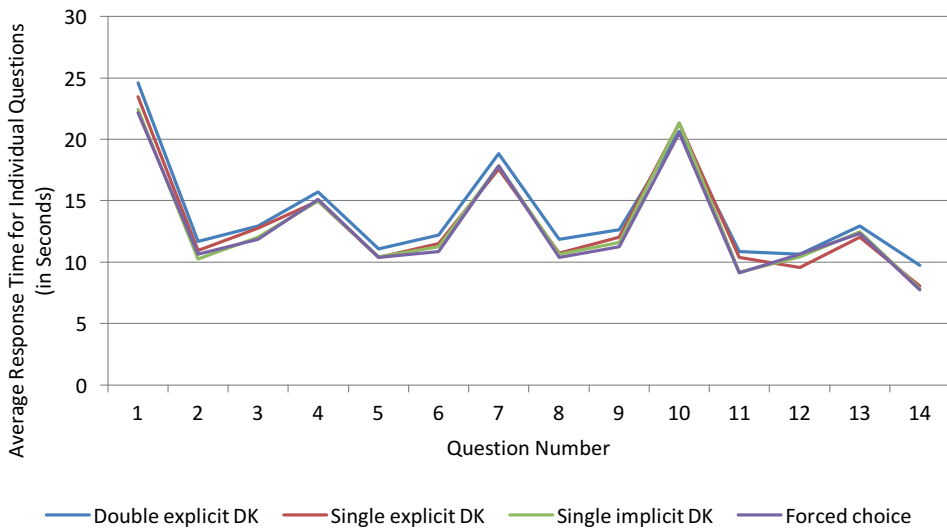
Figure B.1: Item Response Times for Four Questionnaire Variants

33   The number of words is counted in the explicit DK variant; the questions in implicit DK variant 3 and forced choice variant 4 are identical. In the double explicit DK variant, each question included a mention of the explicit DK option. The response categories were not taken into account.

APPENDIX C   Additional Analyses,
Tables and Figures
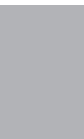
I   Additions to Chapter 5: Don't Know

Distributions of Opinions

In the presentation below, a distinction is made between self-placement items and other items, since the significance of differences between distributions of questionnaire variants are analyzed differently (by comparing means) for the self-placement items. Figure C.1 to Figure C.4 present the distributions of opinions of the self-placement items followed by the significance tests (Table C.1); Figure C.5 to Figure C.14 show the distributions of the other items, also followed by significance tests (Table C.3). Item nonresponse is treated as missing data and excluded from the findings.
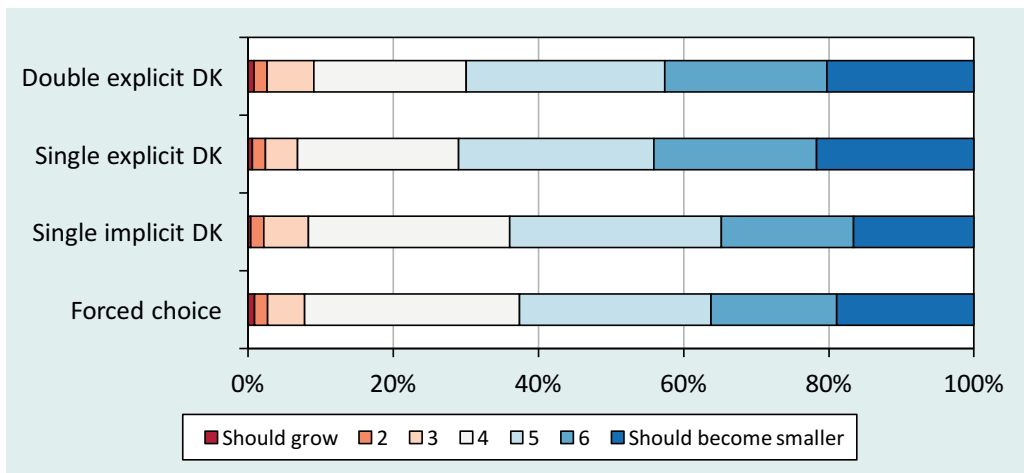
Figure C.1: Distribution of Opinions Self-Placement Income Differences
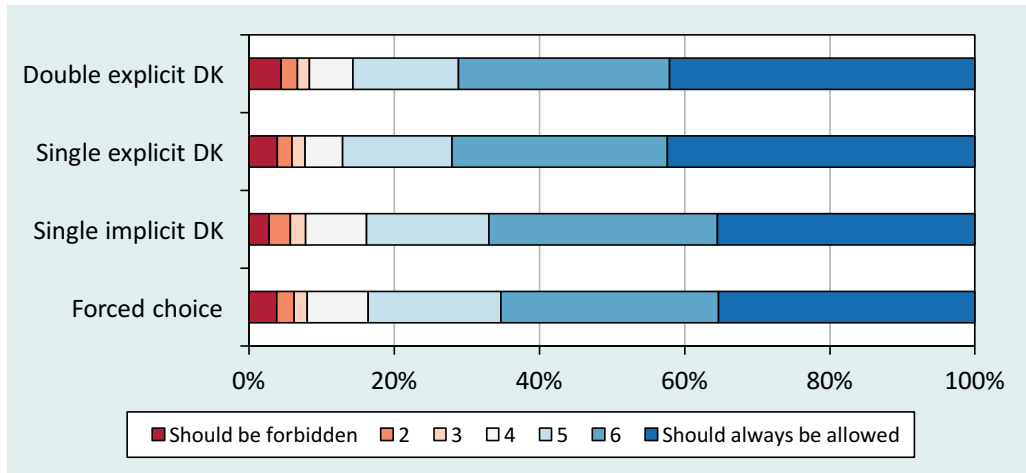
Figure C.2: Distribution of Opinions Self-Placement Euthanasia
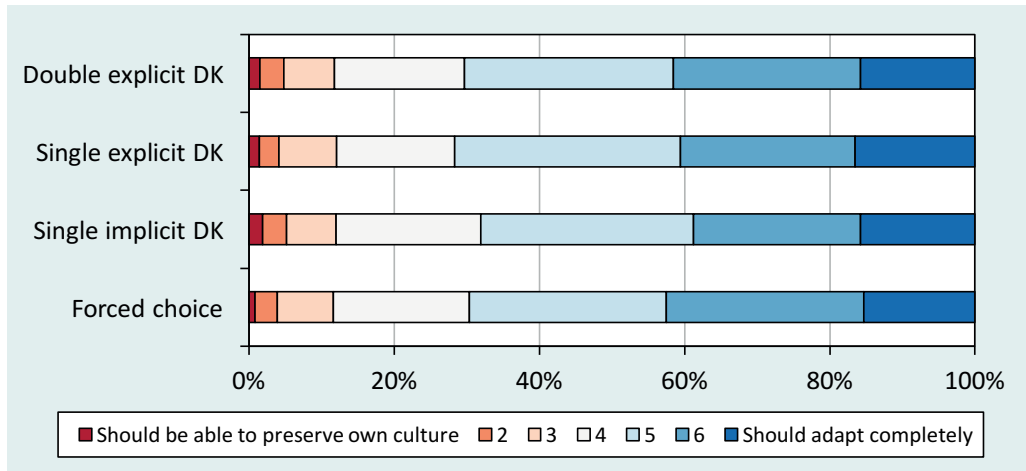
Figure C.3: Distribution of Opinions Self-Placement Foreigners
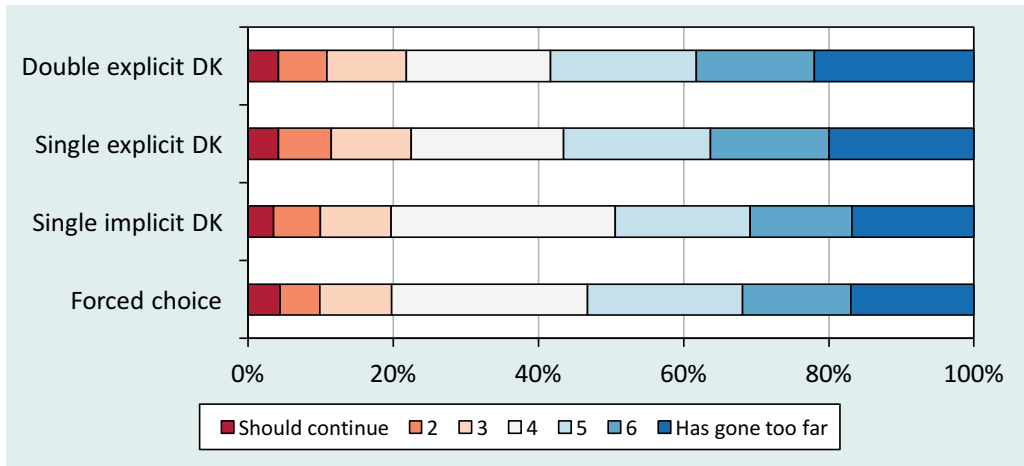
Figure C.4: Distribution of Opinions Self-Placement European Unification

Table C.1: Significance of Differences in Means Self-Placement Items[34]

Question	1/2	1/3	1/4	2/3	2/4	3/4
Self-placement Income Differences	.202	.001	.005	.000	.000	.721
Self-placement Euthanasia	.415	.103	.024	.013	.002	.520
Self-placement Foreigners	.731	.234	.801	.125	.926	.147
Self-placement European Unification	.279	.005	.031	.091	.300	.492
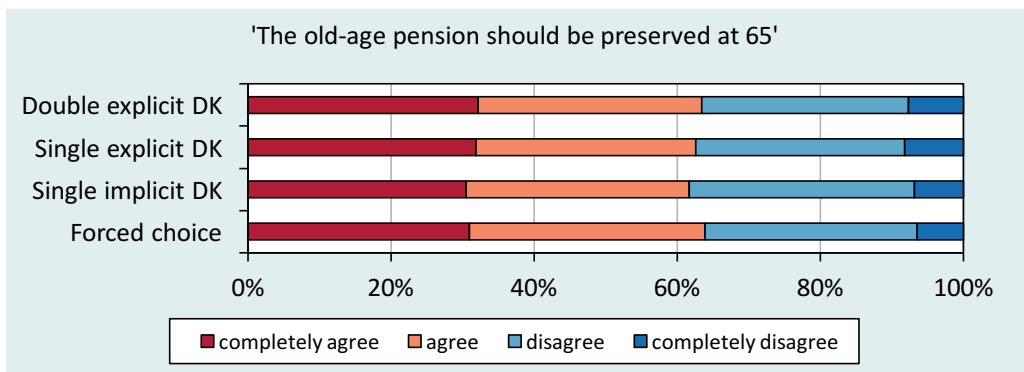
Figure C.5: Distribution of Opinions Old-Age Pension

34   The significance of differences between the means is calculated for all possible pairs of questionnaire variants. Item nonresponse is excluded from the analysis.
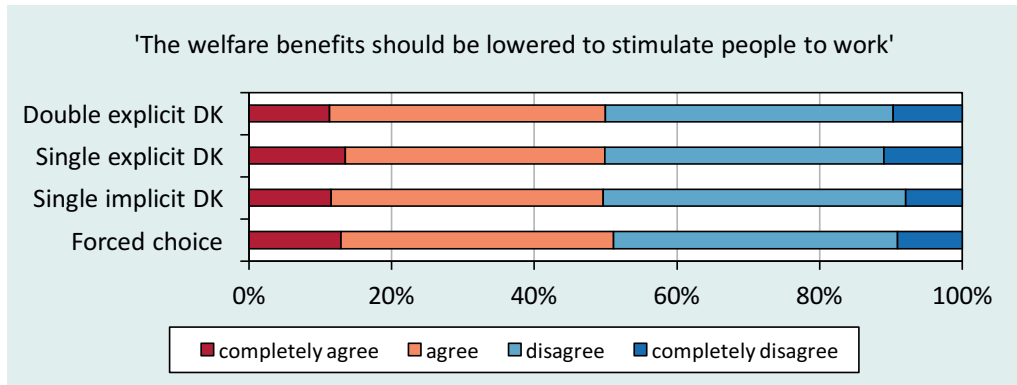
Figure C.6: *Distribution of Opinions Lowering Welfare Benefits*
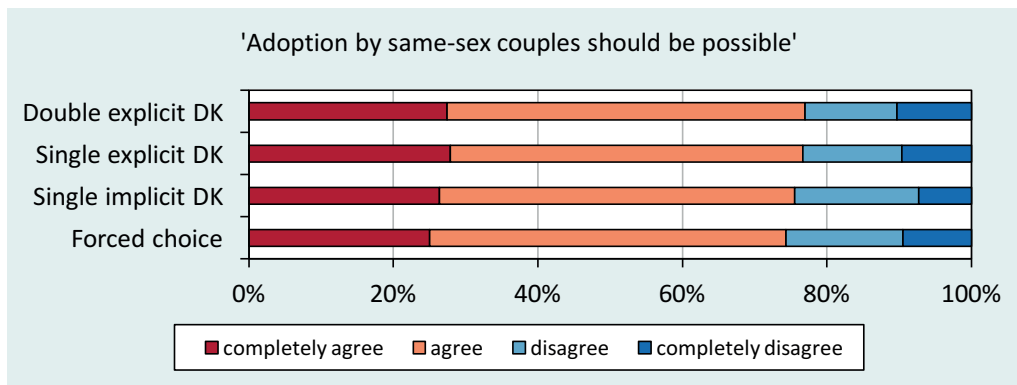
Figure C.7: *Distribution of Opinions Same-Sex Adoption*
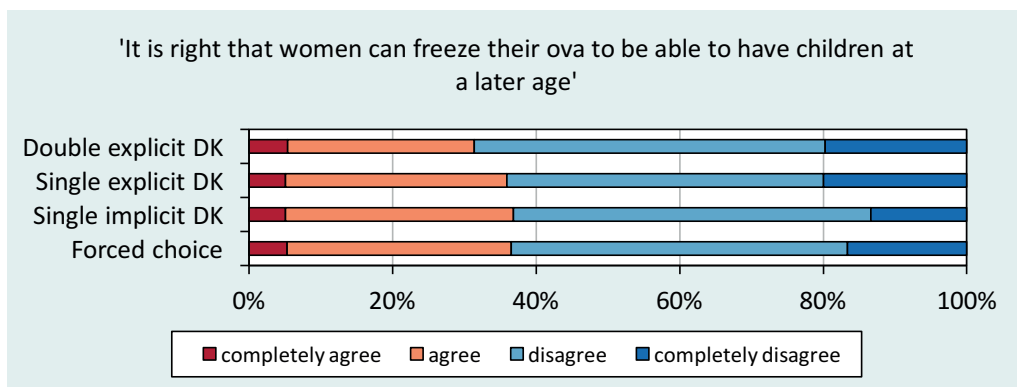
Figure C.8: *Distribution of Opinions Ova Freezing*
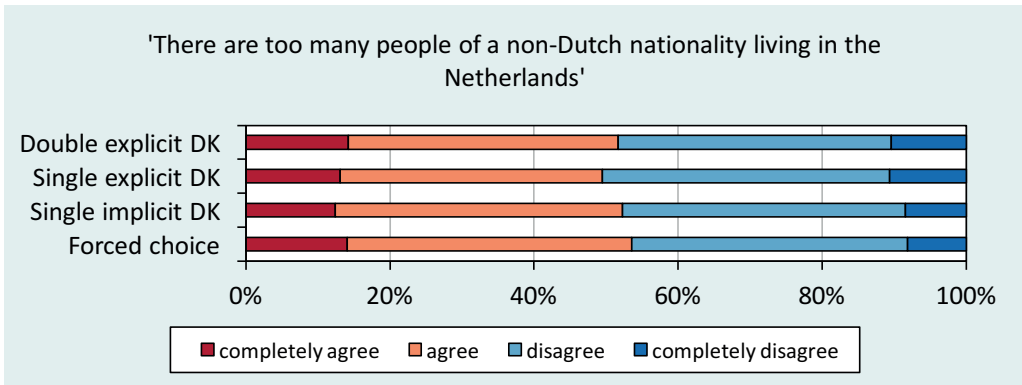
Figure C.9: Distribution of Opinions Number of non-Dutch
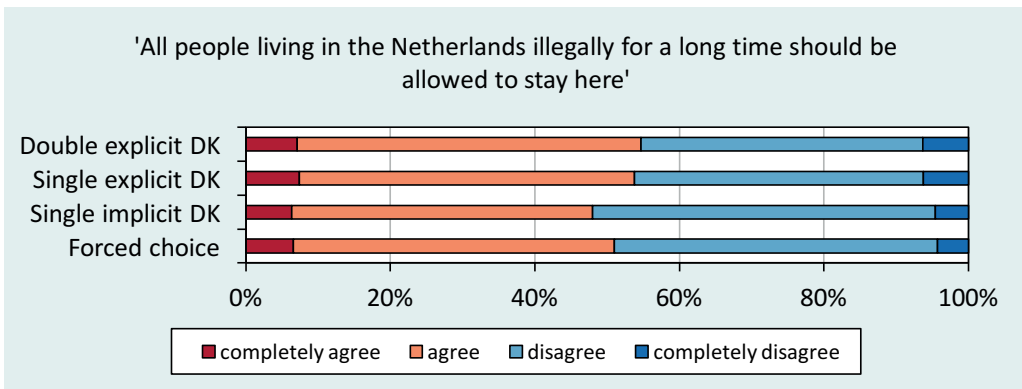
Figure C.10: Distribution of Opinions Illegal in Netherlands
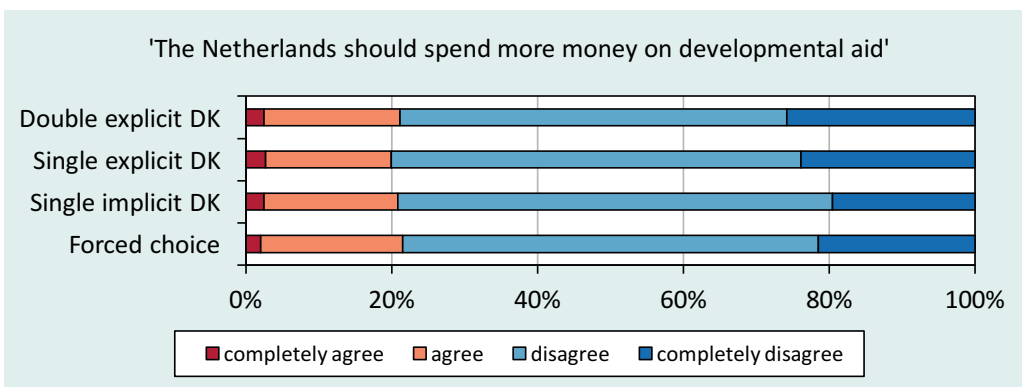
Figure C.11: Distribution of Opinions Developmental Aid
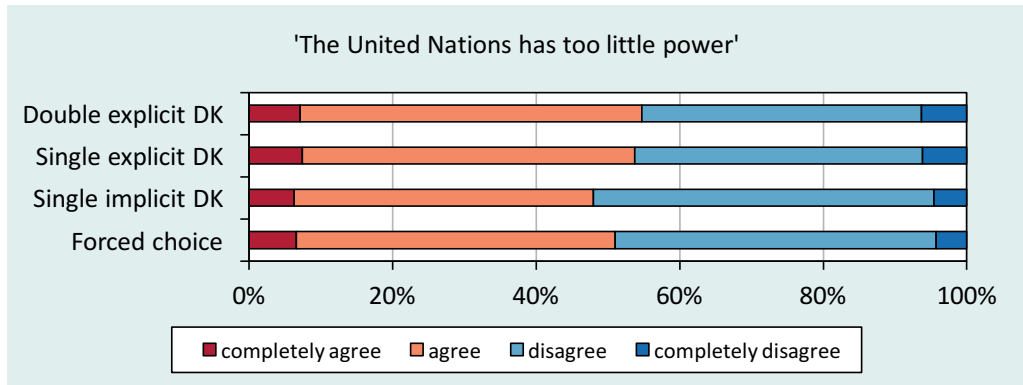
Figure C.12: Distribution of Opinions UN Power
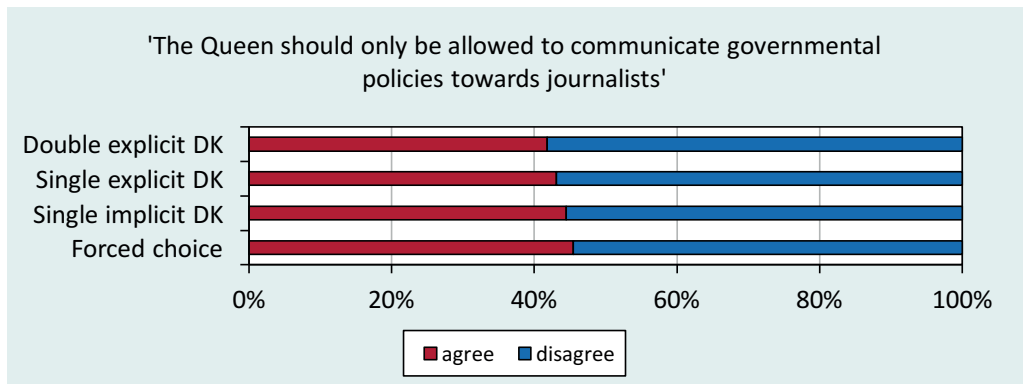
Figure C.13: Distribution of Opinions Role Queen
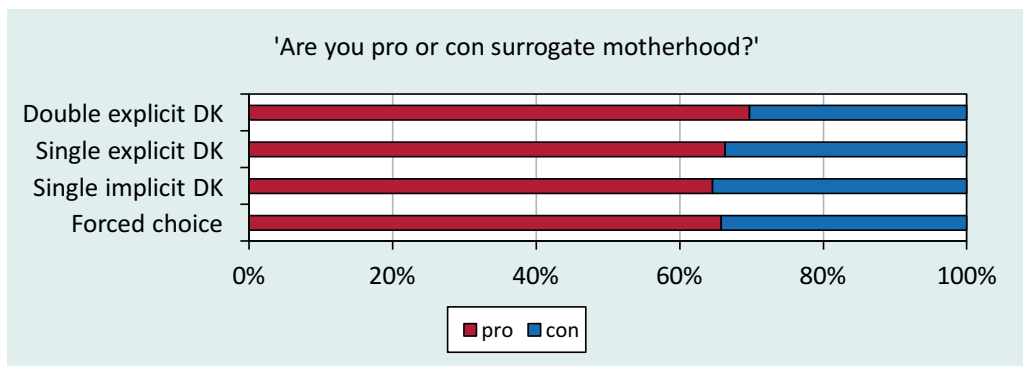
Figure C.14: Distribution of Opinions Surrogate Motherhood

Table C.2: Significance of Differences in Distribution of Opinions Other Items

Question	1/2	1/3	1/4	2/3	2/4	3/4
Old-Age Pension	.967	.423	.461	.367	.266	.686
Lowering of Welfare	.201	.357	.625	.012	.325	.361
Same-sex Adoption	.829	.001	.057	.016	.173	.160
Ova Freezing	.040	.000	.017	.000	.149	.105
Number of non-Dutch	.703	.137	.197	.137	.065	.626
Illegal in Netherlands	.775	.126	.243	.519	.562	.705
Developmental Aid	.398	.001	.048	.059	.184	.338
UN Power	.956	.001	.018	.004	.043	.470
Role Queen	.555	.200	.077	.505	.250	.619
Surrogate Motherhood	.104	.011	.046	.409	.829	.523

The significance levels in this table are indicators for differences in the distributions of opinions of two questionnaire variants. Item nonresponse is excluded from the analysis.

Figure C.15 - Figure C.24 display the two pictures of public opinion: one excluding DK answers (in variant 1 and 2) and skipped questions (in variant 3) as missing data and one including item nonresponse as substantively relevant part of public opinion. For these analyses, the answers were dichotomized. The self-placement items were excluded.
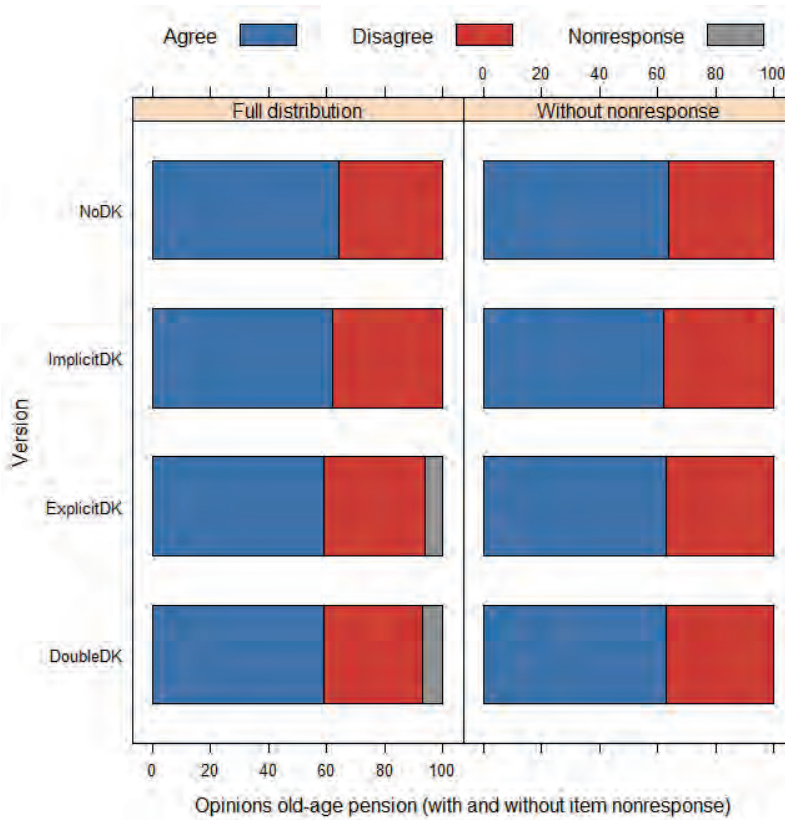
*Figure C.15: Distribution (%) of Opinions **Old-Age Pension** With and Without Item Nonresponse*
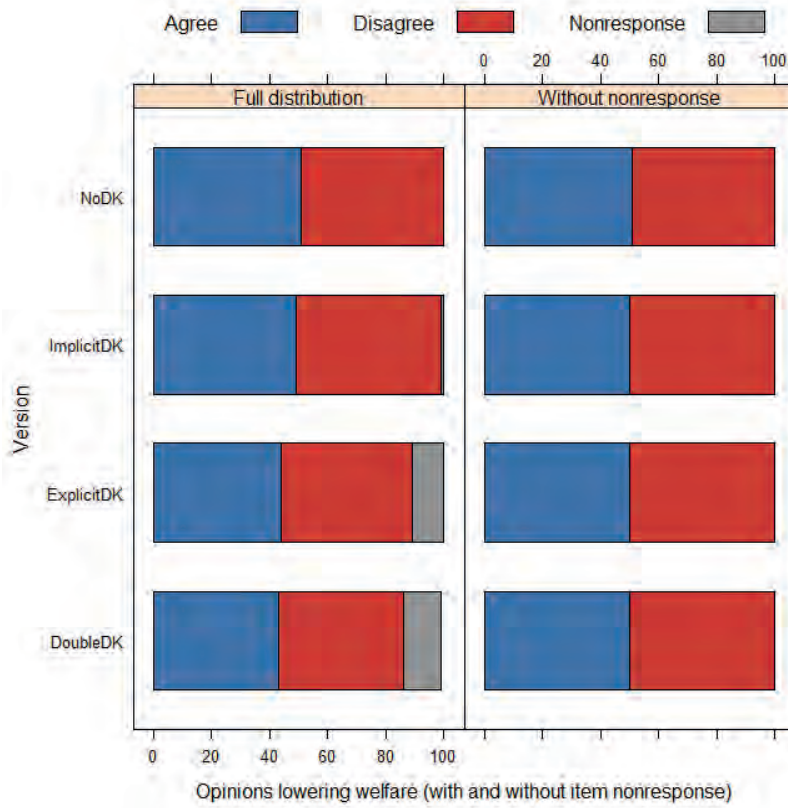
*Figure C.16: Distribution (%) of Opinions **Lowering Welfare** With and Without Item Nonresponse*
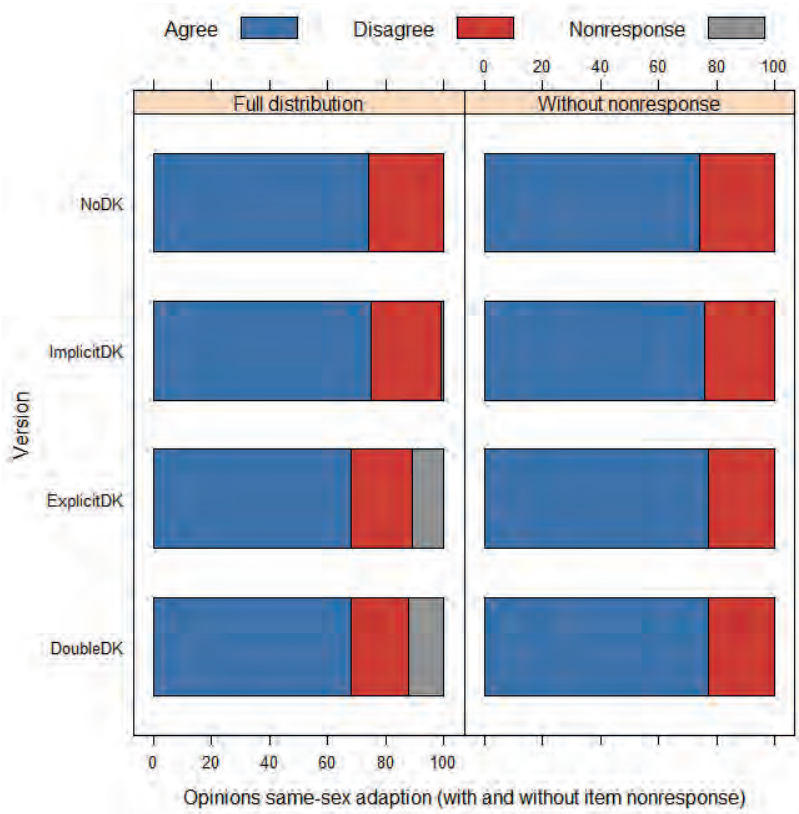
Figure C.17: Distribution (%) of Opinions **Same-Sex Adoption** With and Without Item Nonresponse
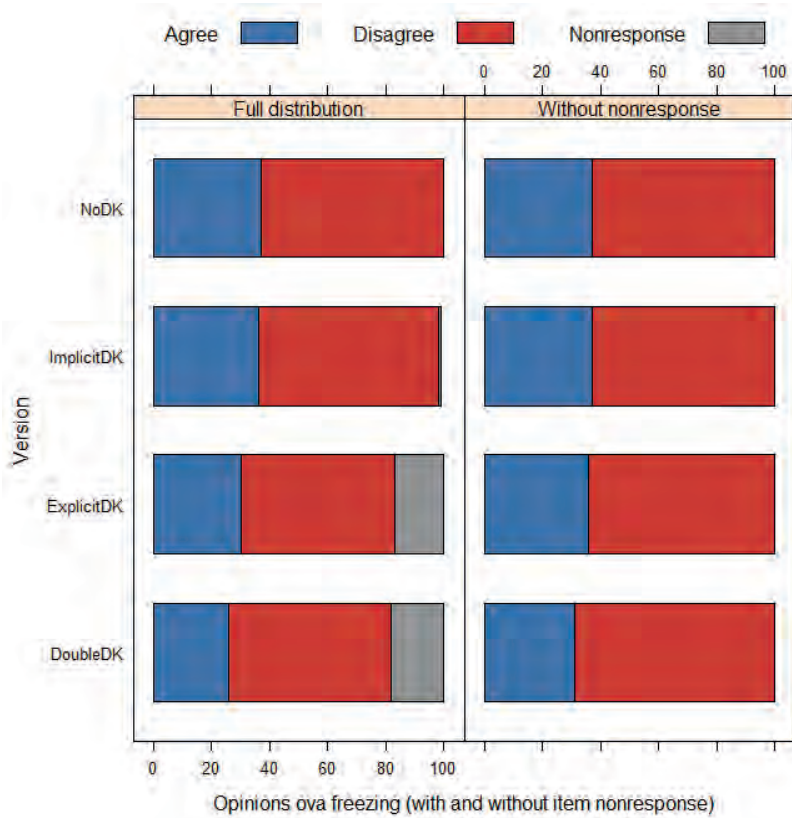
Figure C.18: Distribution (%) of Opinions **Ova Freezing** *With and Without Item Nonresponse*
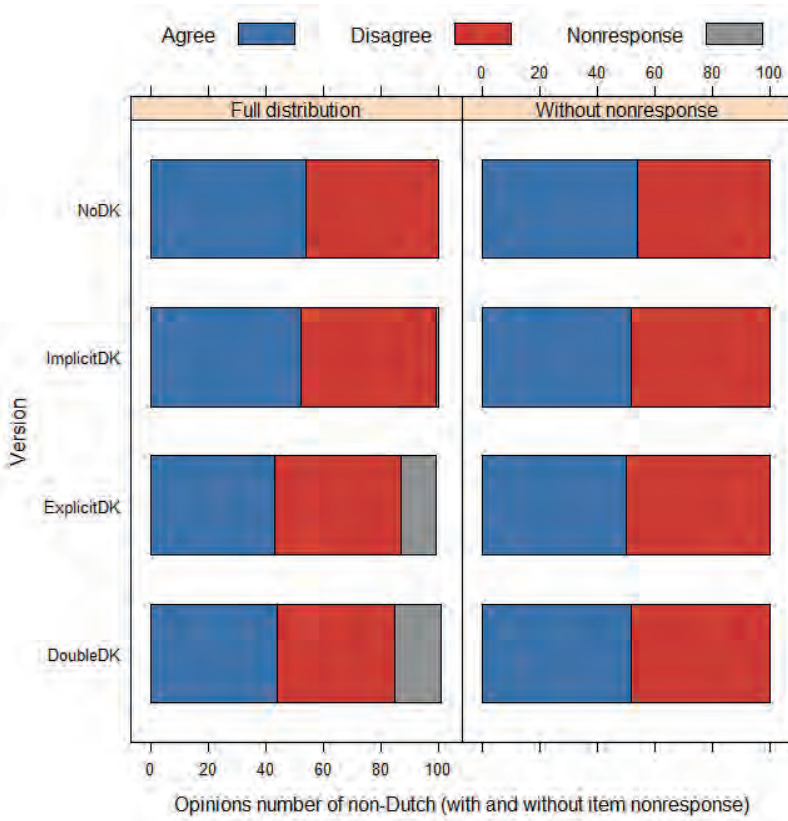
Figure C.19: Distribution (%) of Opinions **Number of non-Dutch** With and Without Item Nonresponse
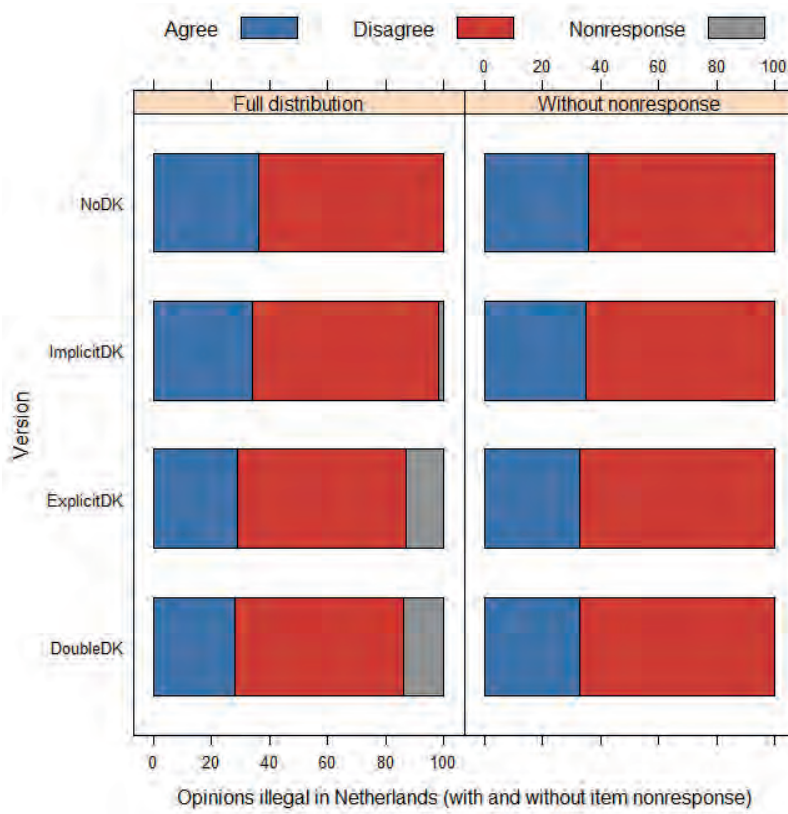
*Figure C.20: Distribution (%) of Opinions **Illegal in Netherlands** With and Without Item Nonresponse*
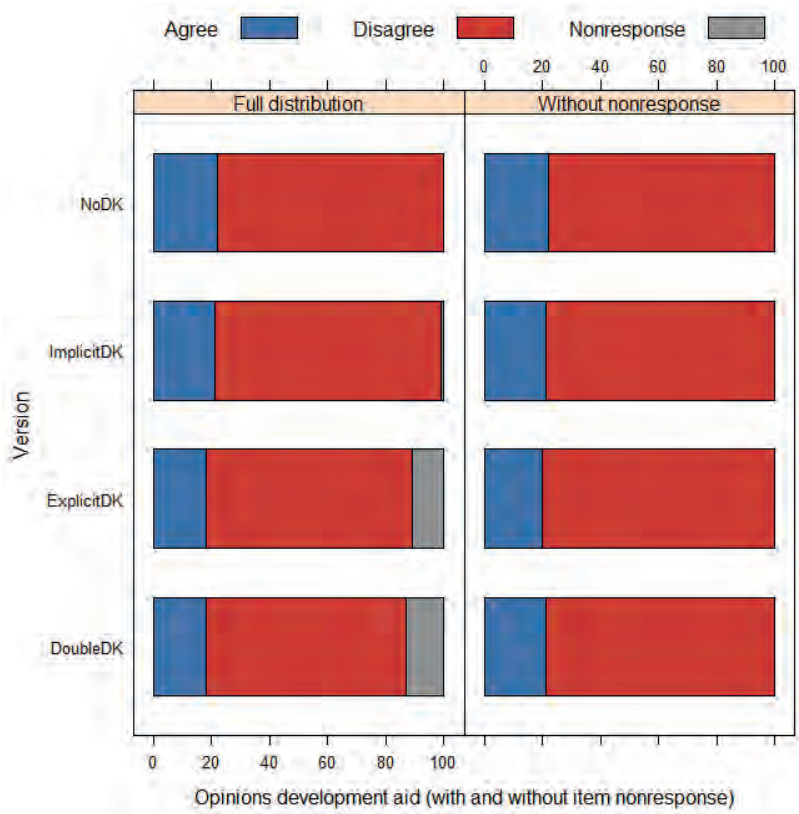
*Figure C.21: Distribution (%) of Opinions **Developmental Aid** With and Without Item Nonresponse*
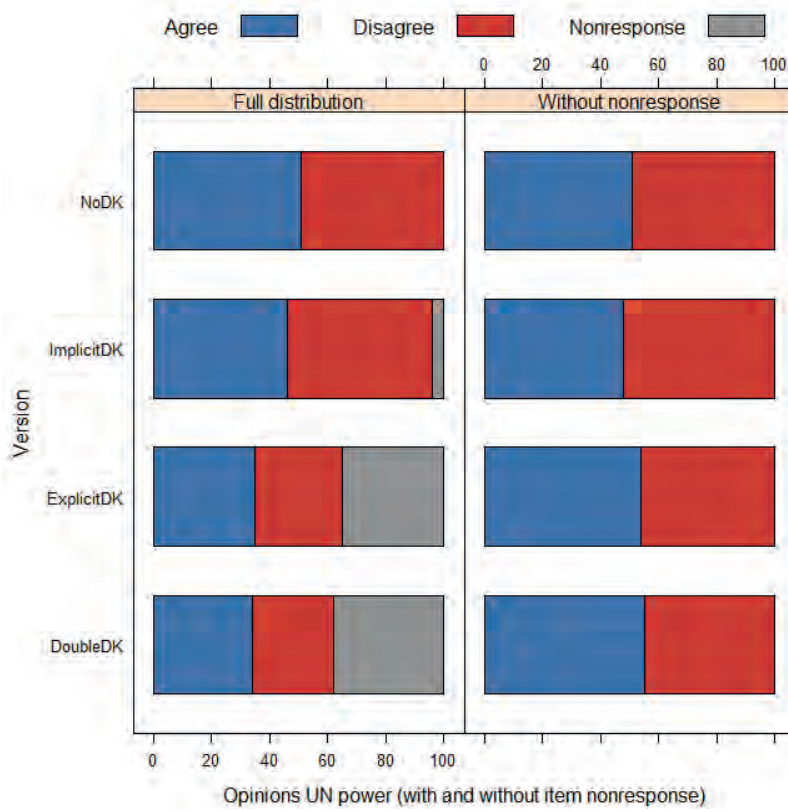
*Figure C.22: Distribution (%) of Opinions **UN Power** With and Without Item Nonresponse*
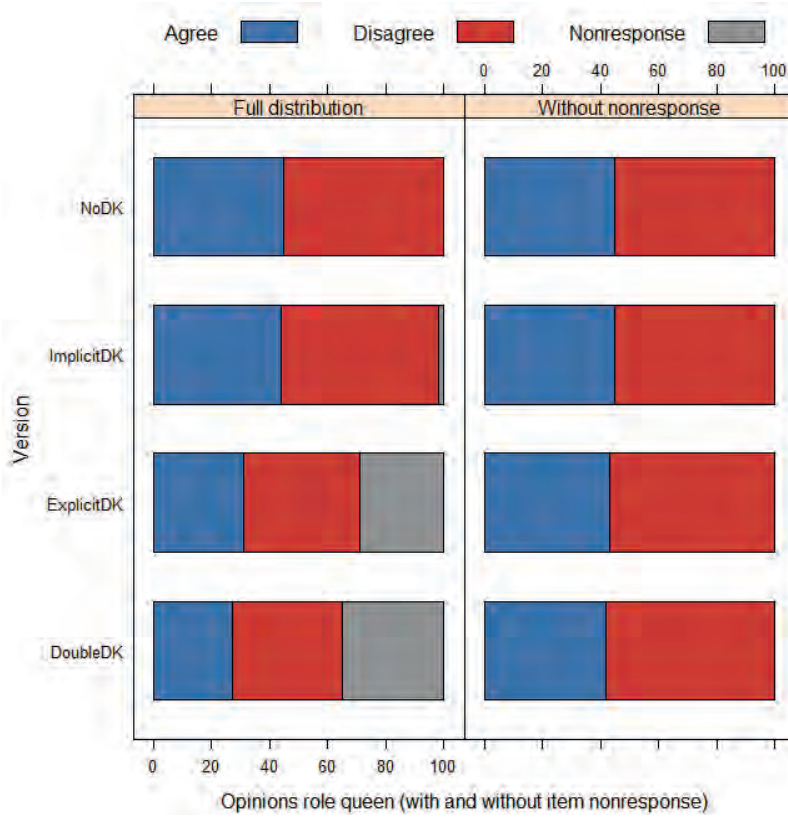
*Figure C.23: Distribution (%) of Opinions **Role Queen** With and Without Item Nonresponse*
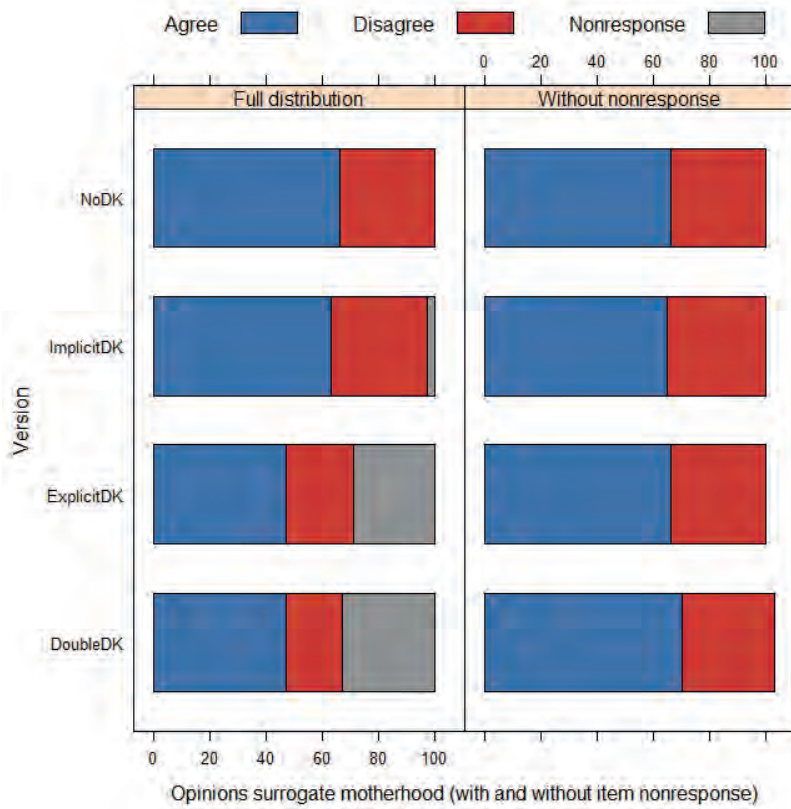
*Figure C.24: Distribution (%) of Opinions **Surrogate Motherhood** With and Without Item Nonresponse*

Table C.3: Mean, Mode and Standard Deviation of Self-Placement Items

Question		1. Double Explicit DK	2. Single Explicit DK	3. Single Implicit DK	4. Forced choice
Self-placement Income Differences***	Mean	5.3	5.3	5.0	5.1
	SD	1.3	1.3	1.3	1.3
	Mode	5	5	5	4
Self-placement Euthanasia**	Mean	5.8	5.8	5.7	5.7
	SD	1.5	1.5	1.5	1.5
	Mode	7	7	7	7
Self-placement Foreigners	Mean	5.1	5.1	5.0	5.1
	SD	1.4	1.4	1.4	1.3
	Mode	5	5	5	6
Self-placement European Unification*	Mean	4.8	4.7	4.6	4.7
	SD	1.7	1.7	1.6	1.6
	Mode	7	4	4	4

*Significant at the .05 level, **Significant at the .01 level, ***Significant at the .001 level
The significance levels indicate the difference in means between all four variants, compared at the same time

Table C.4: Significance Between-Variant-Differences in Item Nonresponse (Themes)

	1 / 2	1 / 3	2 / 3
Socio-Economic	.015	.000	.000
Ethical	.335	.000	.000
Multicultural	.045	.000	.000
Foreign Affairs	.055	.000	.000
Foreign Affairs Without UN	.170	.000	.000
Current Affairs	.000	.000	.000

Table C.5: Significance Between-Variant-Differences in Use of Neutral Response Category

	1/2	1/3	1/4	2/3	2/4	3/4
Self-placement Income	.181	.000	.000	.000	.000	.277
Self-placement Euthanasia	.189	.010	.007	.000	.000	.946
Self-placement Foreigners	.134	.102	.411	.006	.049	.443
Self-placement EU	.198	.000	.000	.000	.000	.037

II   Additions to Chapter 6: Filter Question

Table C.6 presents the item nonresponse which is caused by the use of the Don't Know option – either by choosing the explicit option or by skipping the question in the implicit variants. The explicit DK variants are grouped together on the left of the table; the implicit DK variants on the right.

Table C.6: DK Item Nonresponse (%) of Individual Items

Question	1A Strong Filter, Explicit DK	1B Strong Filter, Implicit DK	2A Weak Filter, Explicit DK	2B Weak Filter, Implicit DK	3A No Filter, Explicit DK	3B No Filter, Implicit DK
Self-placement Income Differences	1	1	1	1	2	1
Welfare should be lowered in order to stimulate people to work	2	2	4	1	1	1
Self-placement Euthanasia	0	0	1	1	1	0
Adoption by same-sex couples should be possible	2	1	5	1	1	1
Self-placement Foreigners	0	0	0	1	1	0
There are too many people of a non-Dutch nationality living in the Netherlands	3	2	4	1	1	1
Self-placement European Unification	0	0	2	1	1	0
The Netherlands should spend more money on development aid	3	3	5	1	1	1
What do you think should happen to the mortgage interest deduction?	0	0	3	0	1	1
What do you think is the best solution for the impending deficits of pension funds?	25	20	30	4	3	6
The Netherlands should in the next year quit the euro and go back to the gulden	7	5	8	1	1	2
Do you think that the King or Queen should have political influence, or should s/he restrict herself to ceremonial roles?	2	1	3	1	1	1

Table C.6: Continued

Question	1A Strong Filter, Explicit DK	1B Strong Filter, Implicit DK	2A Weak Filter, Explicit DK	2B Weak Filter, Implicit DK	3A No Filter, Explicit DK	3B No Filter, Implicit DK
Do you think that Maxima's father can or cannot be present at the coronation?	4	2	8	1	1	2
I think that Libya will, in time, become a normal democratic country	21	18	23	2	1	2
The Netherlands should be actively involved to help Libya establish a democratic regime	8	7	9	2	2	2
How long do you think this cabinet will remain in office?	6	4	6	1	1	1
Suppose that next year another 5 billion euro in budget cuts have to be made. Do you think the PVV will stop supporting the government?	18	17	18	2	2	2
N	4412	4327	3931	4329	4283	3591

Distributions of Opinions

Figure C.25 to Figure C.40 present the distributions of opinions of all survey items, with the self-placement items grouped together (Figure C.25 to Figure C.28). Item nonresponse is excluded as missing data.
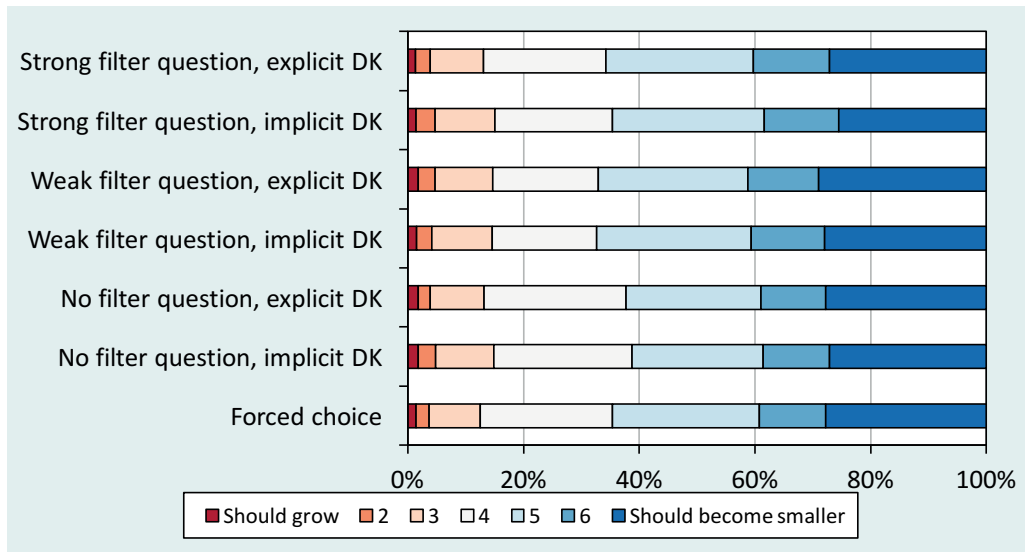
Figure C.25: Distribution of Opinions Self-Placement Income Differences
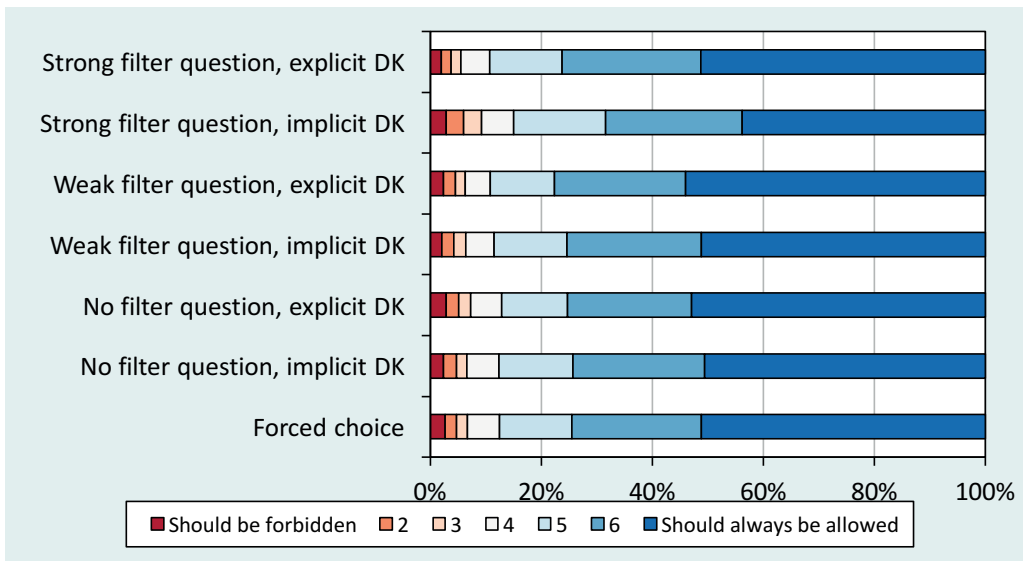
Figure C.26: Distribution of Opinions Self-Placement Euthanasia
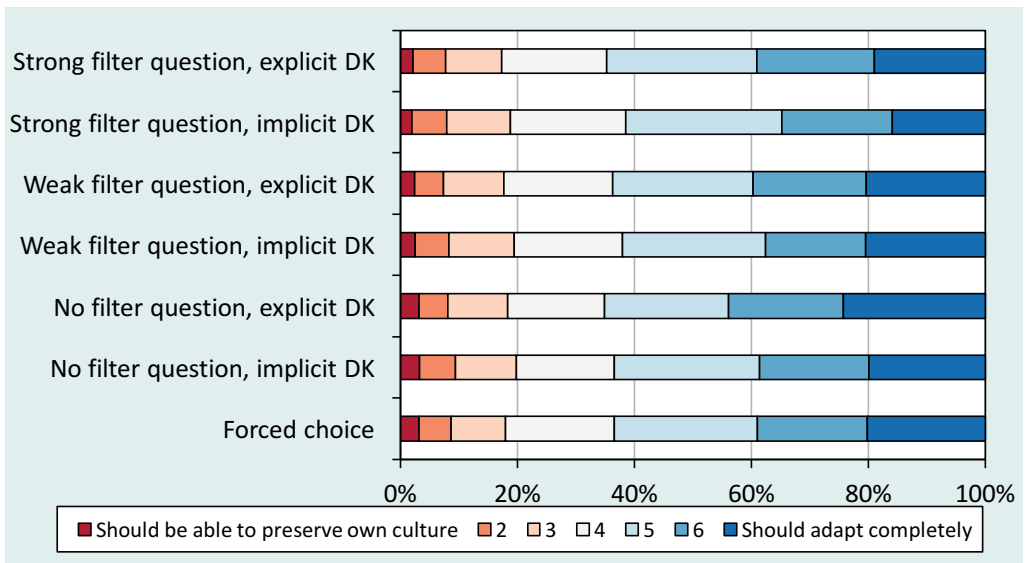
Figure C.27: Distribution of Opinions Self-Placement Foreigners
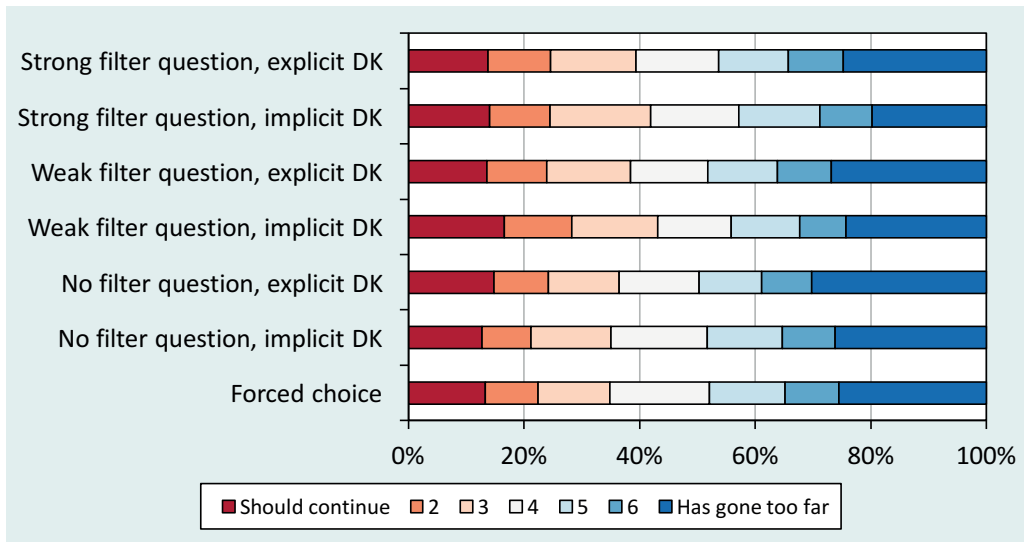
Figure C.28: Distribution of Opinions Self-Placement European Unification
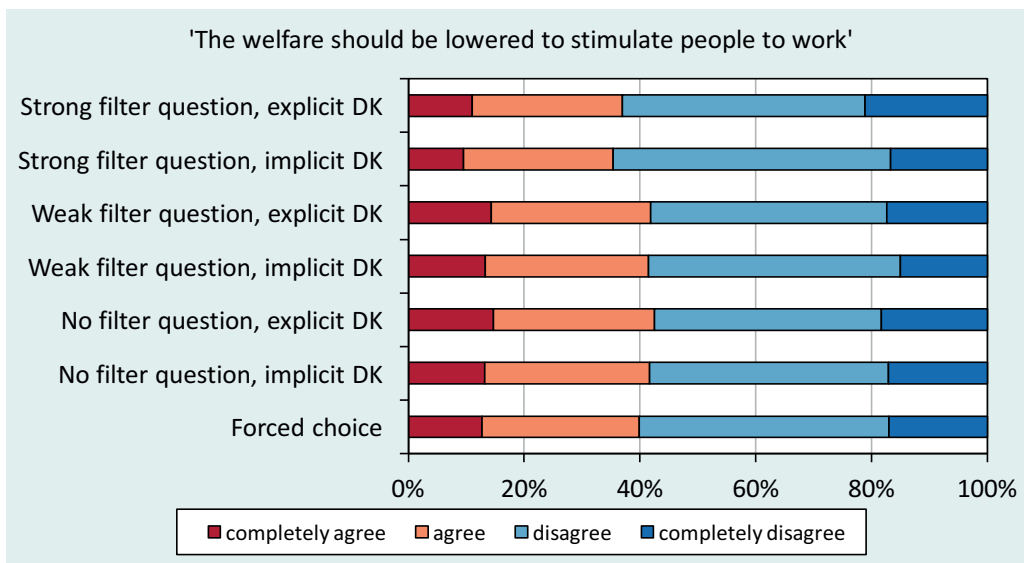
Figure C.29: Distribution of Opinions Lowering Welfare
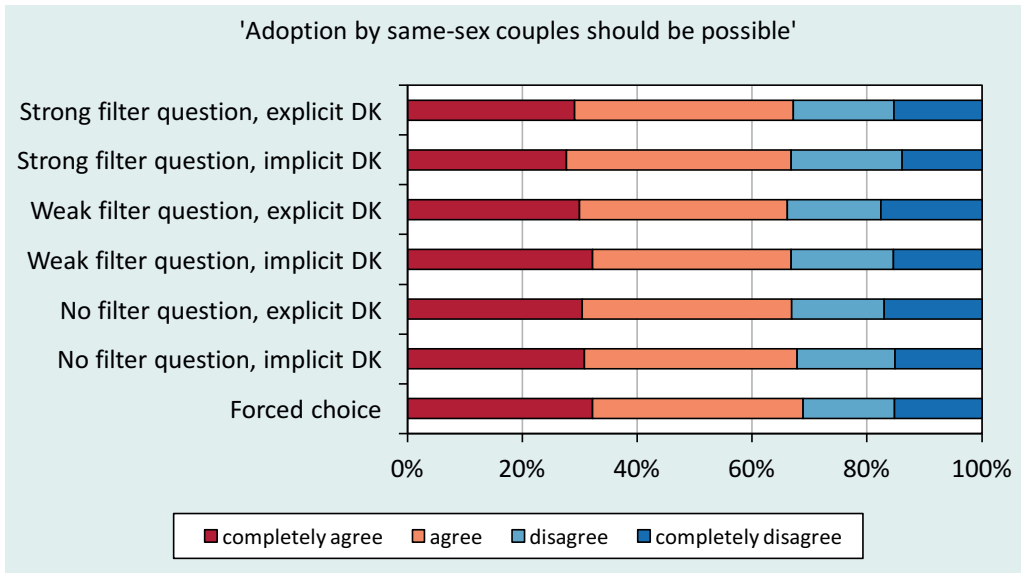
Figure C.30: Distribution of Opinions Same-Sex Adoption
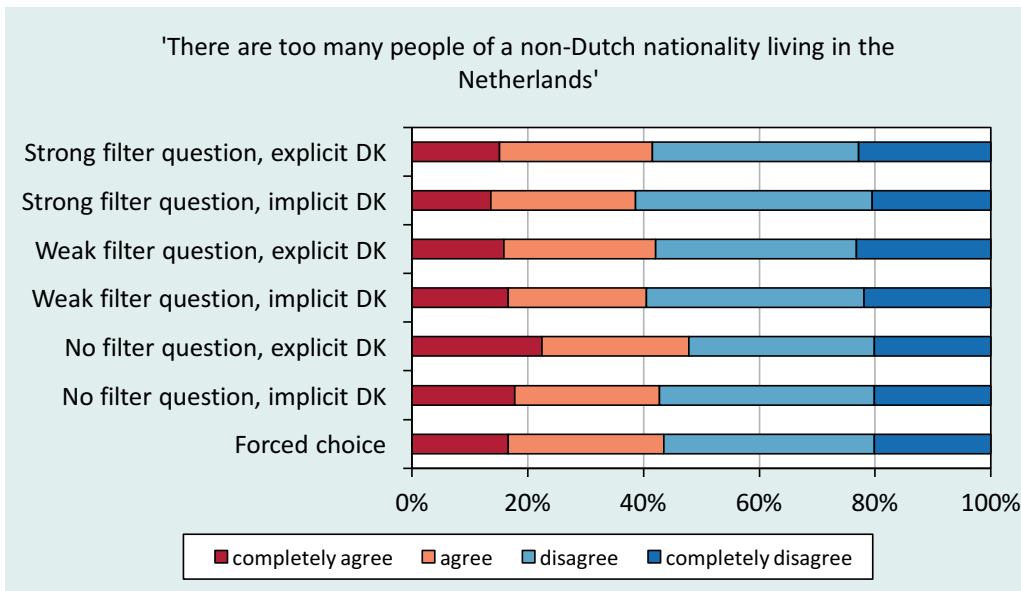
Figure C.31: Distribution of Opinions Non-Dutch Nationality

'What do you think should happen to the mortgage interest deduction?'
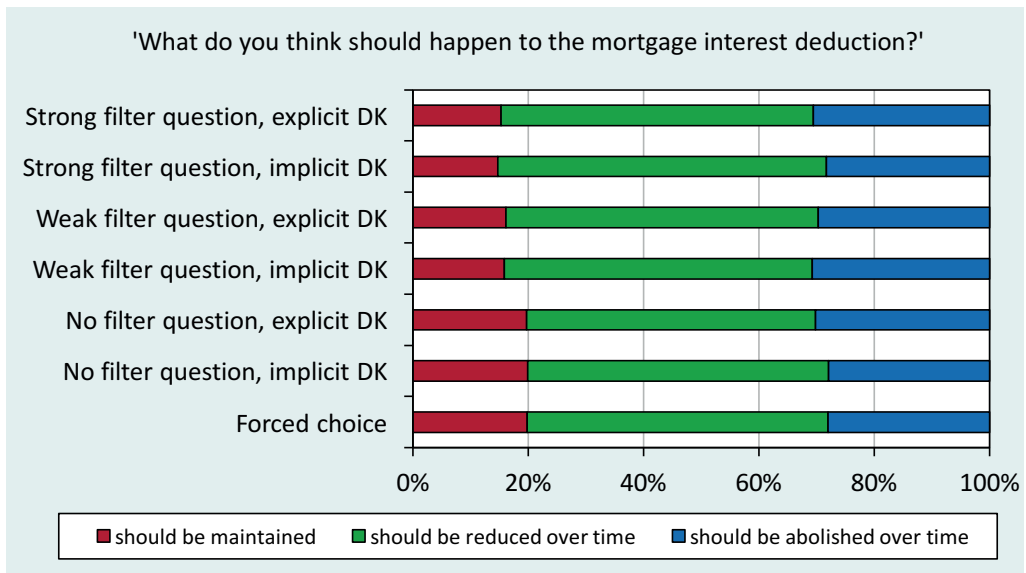
Figure C.32: Distribution of Opinions Mortgage Interest Deduction

'What do you think is the best solution for the pending deficits of pension funds?'
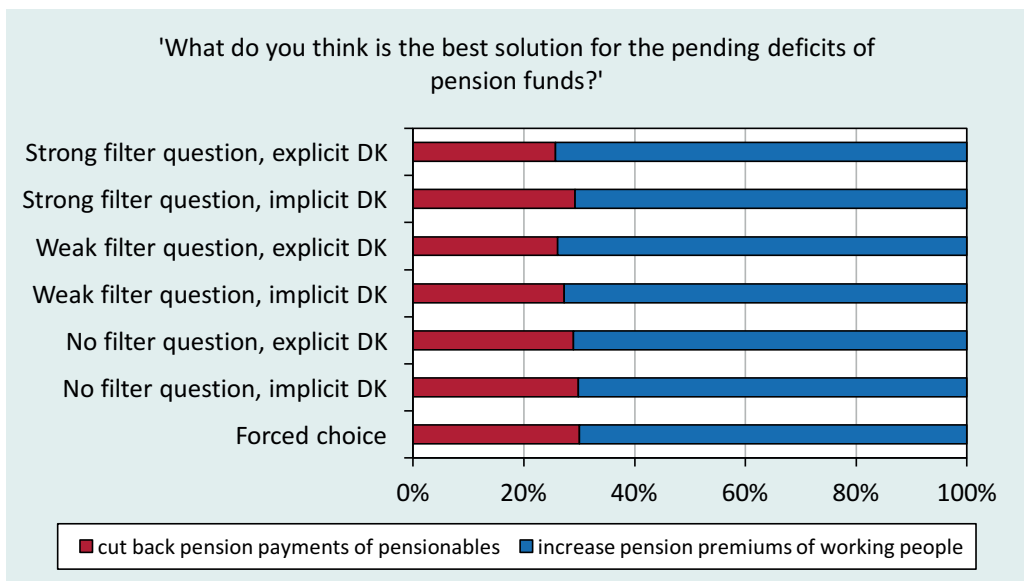
Figure C.33: Distribution of Opinions Pension Funds
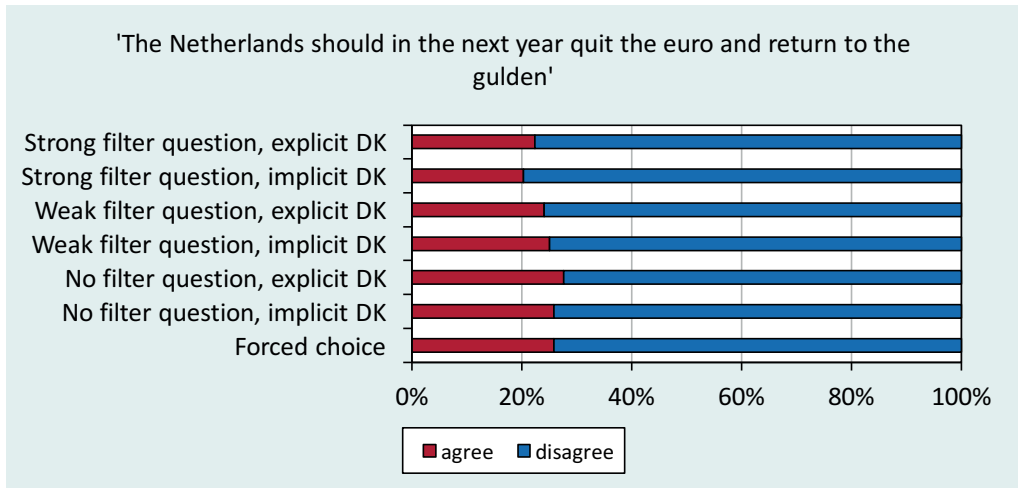
Figure C.34: Distribution of Opinions Gulden
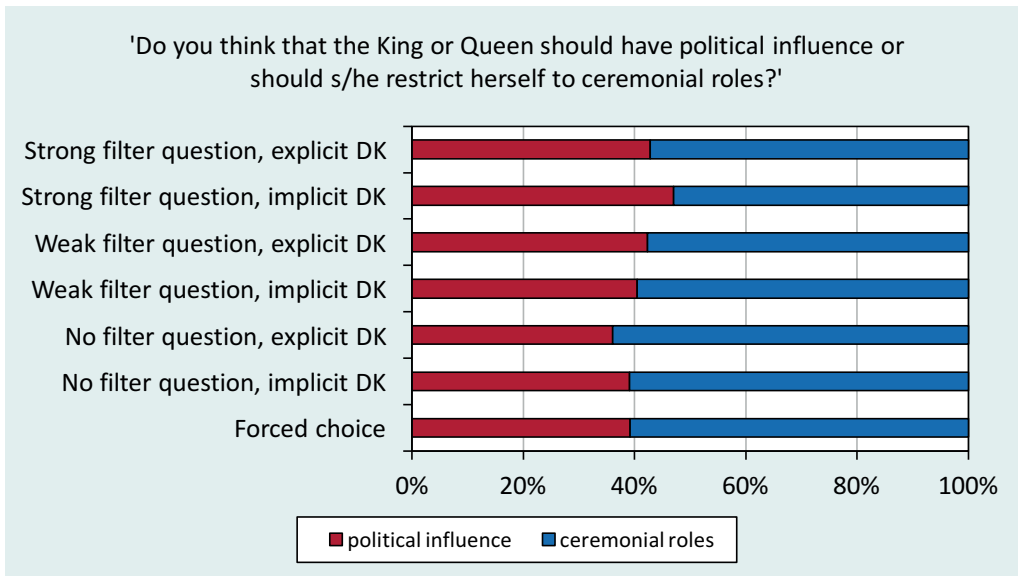
Figure C.35 : Distribution of Opinions King
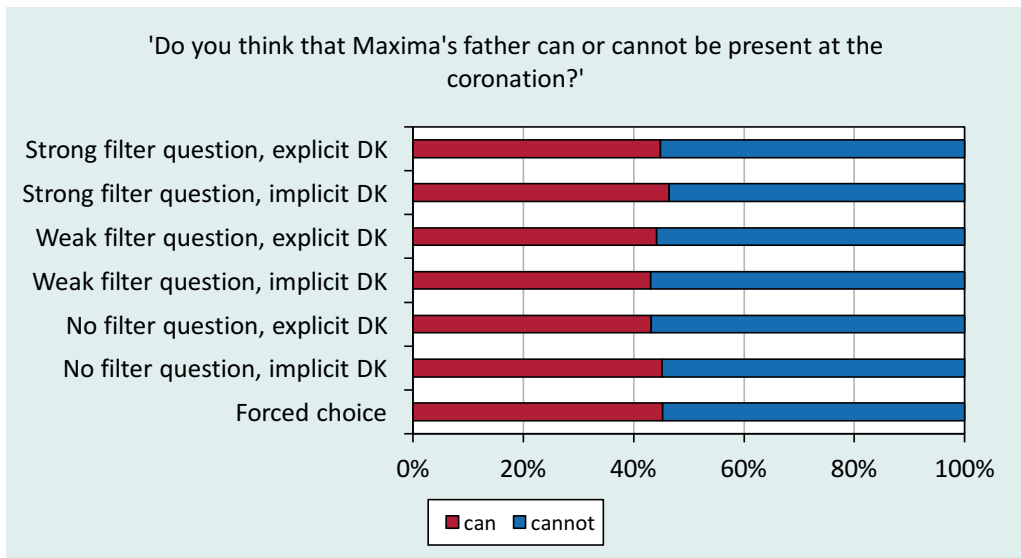
Figure C.36: Distribution of Opinions Maxima's Father
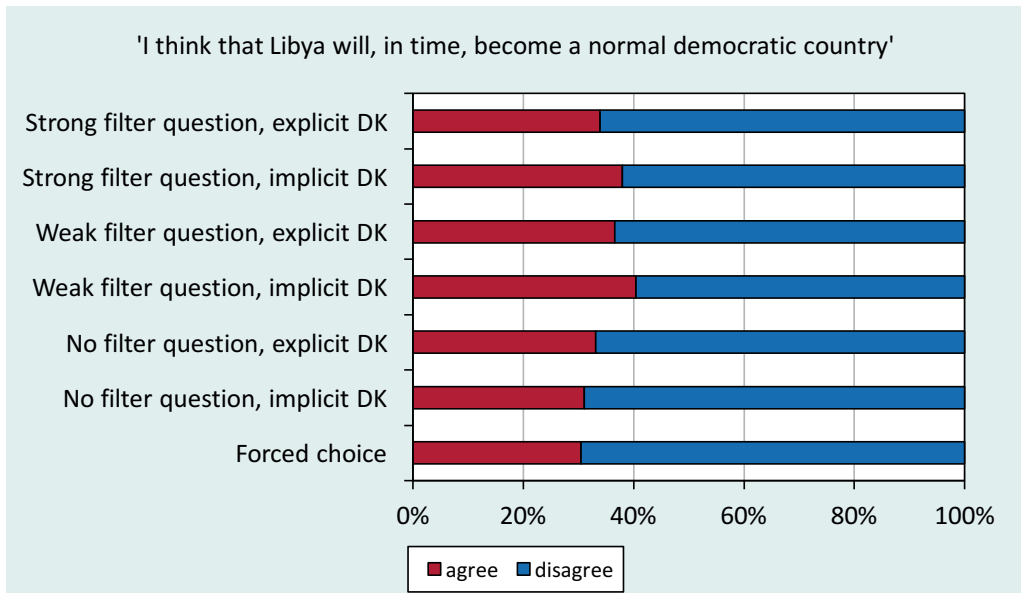
Figure C.37: Distribution of Opinions Libya 1
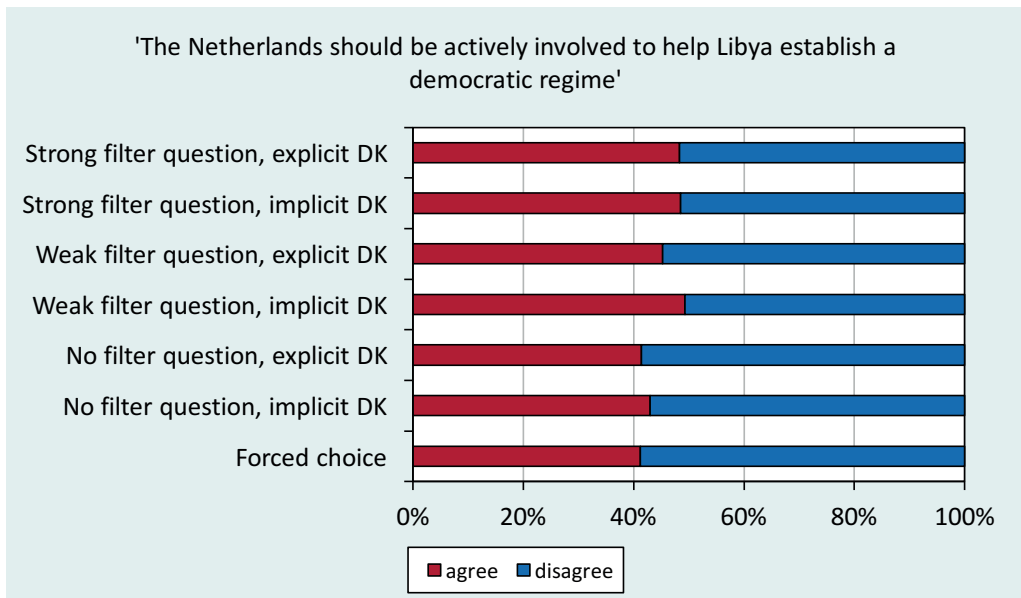
Figure C.38: Distribution of Opinions Libya 2
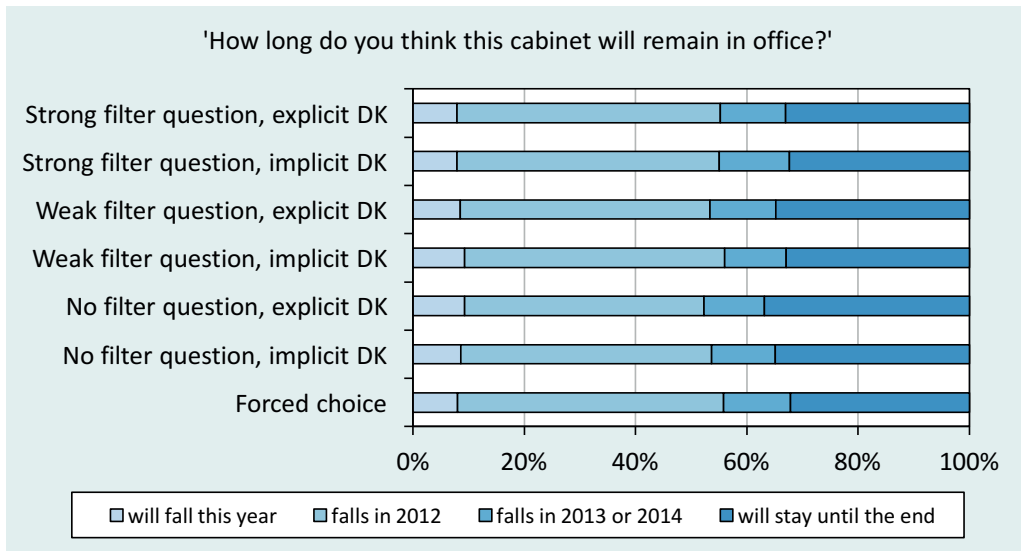
Figure C.39: Distribution of Opinions Cabinet
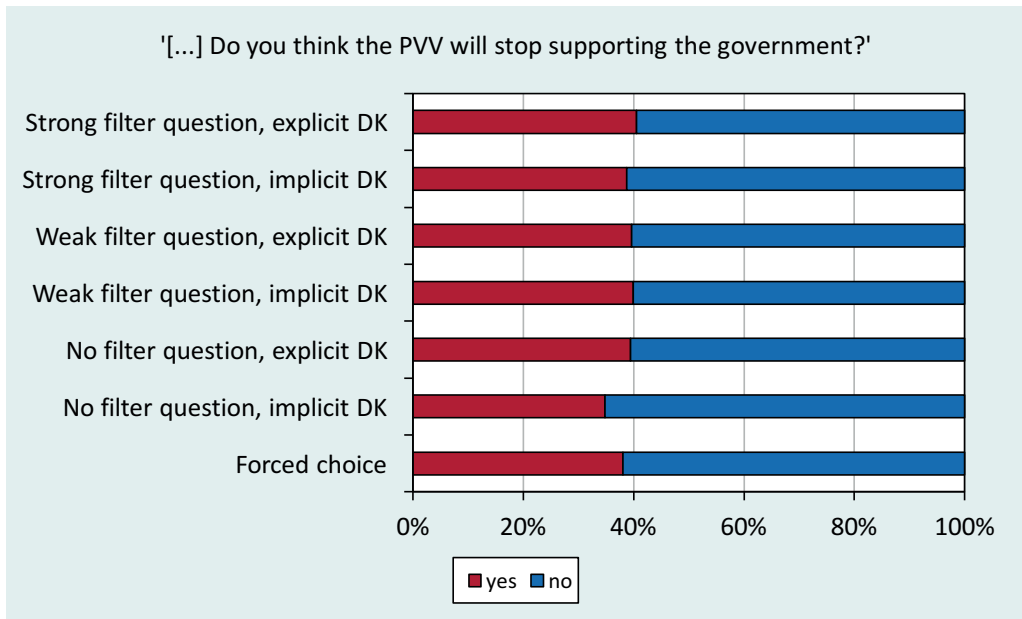
Figure C.40: Distribution of Opinions PVV

Table C.7 is an additional analysis of the distributions of opinions of the four self-placement items.

Table C.7: Mean, Mode and Standard Deviation of Self-placement Items

Question		1A Strong Filter, Explicit DK	1B Strong Filter, Implicit DK	2A Weak Filter, Explicit DK	2B Weak Filter, Implicit DK	3A No Filter, Explicit DK	3B No Filter, Implicit DK	4 Forced Choice
Income Differences*	Mean	5.2	5.1	5.2	5.2	5.1	5.1	5.1
	SD	1.5	1.5	1.6	1.5	1.5	1.6	1.5
	Mode	7	5	7	7	7	7	7
Euthanasia**	Mean	6.1	5.8	6.1	6.0	6.0	6.0	6.0
	SD	1.3	1.5	1.4	1.4	1.5	1.4	1.4
	Mode	7	7	7	7	7	7	7
Foreigners **	Mean	5.0	4.8	5.0	4.9	5.0	4.9	4.9
	SD	1.5	1.5	1.6	1.6	1.6	1.6	1.6
	Mode	5	5	5	5	7	5	5
European Unification**	Mean	4.3	4.1	4.4	4.1	4.4	4.4	4.4
	SD	2.1	2.0	2.1	2.2	2.2	2.1	2.1
	Mode	7	7	7	7	7	7	7

*Significant at the .010 level, **Significant at the .001 level
The significance levels indicate the difference in means between all seven variants, compared at the same time.

Figure C.41 to Figure C.53 display the two pictures of public opinion for all seven variants: one excluding item nonresponse as missing data and one including item nonresponse as part of the outcome. Self-placement items are not included in the analysis below, because they could not be dichotomized.
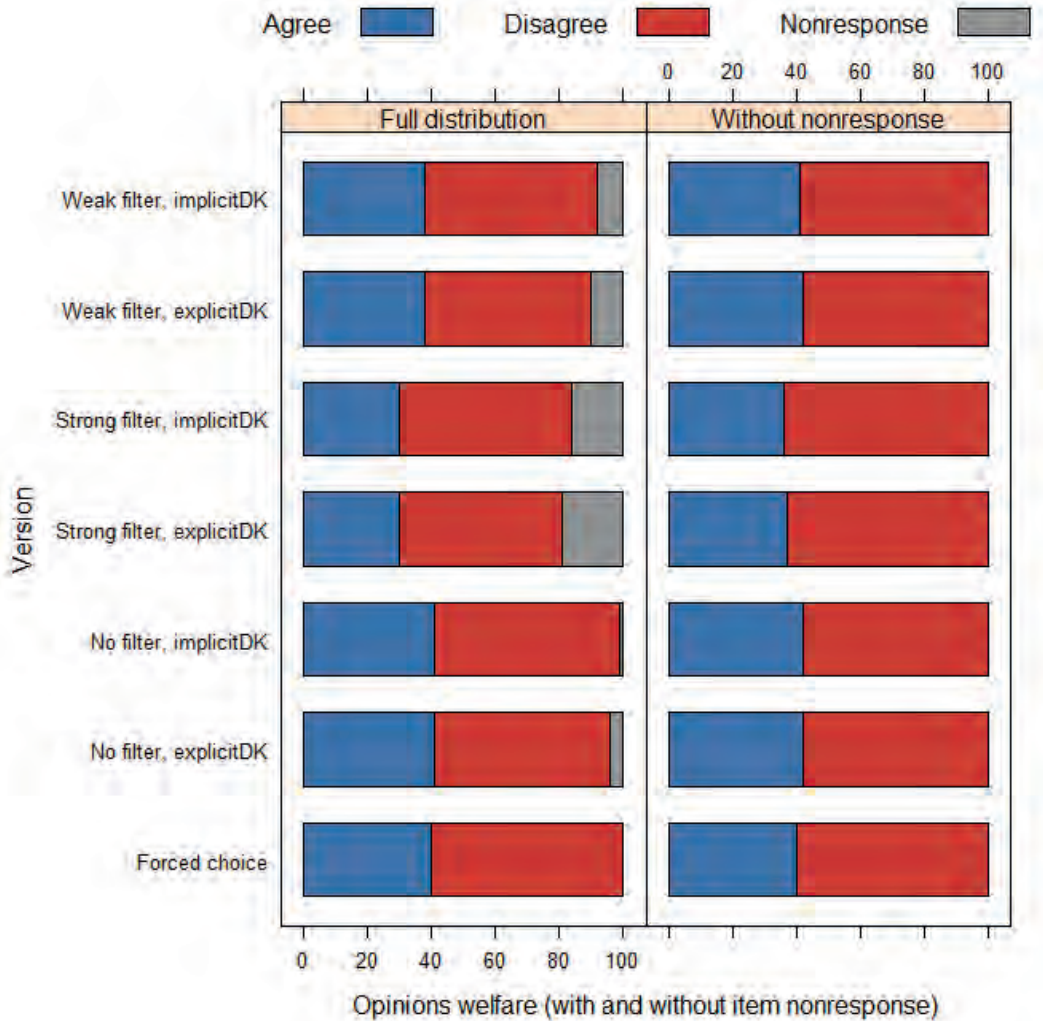
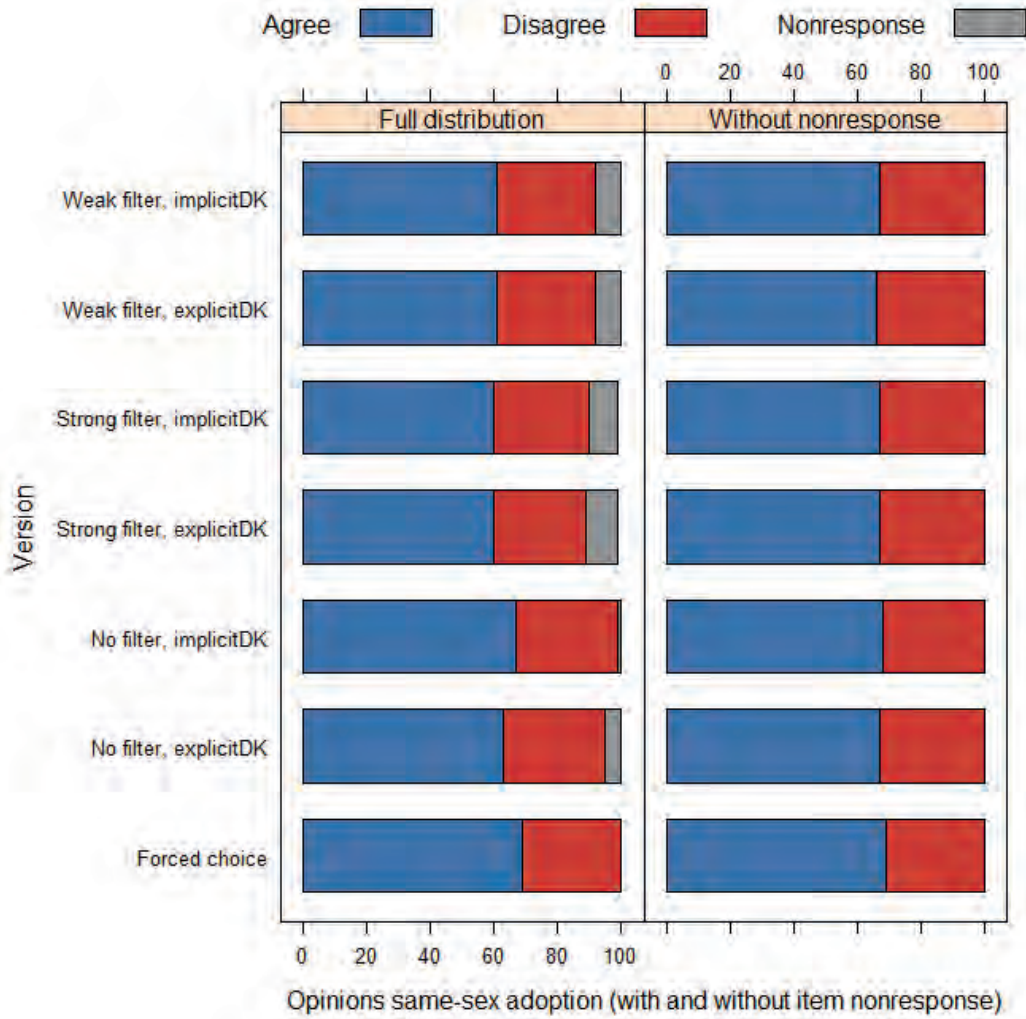*Figure C.41: Distribution (%) of Opinions WELFARE With and Without Item Nonresponse*

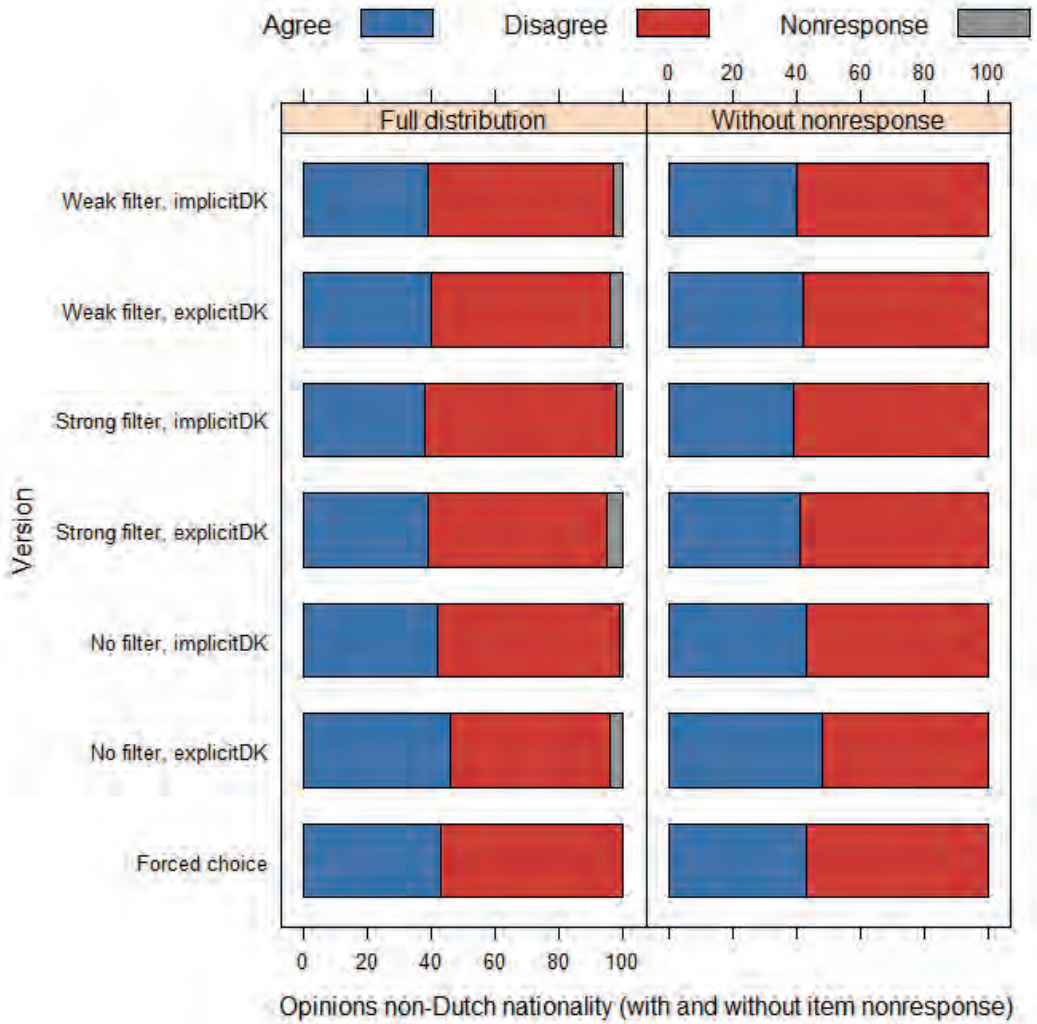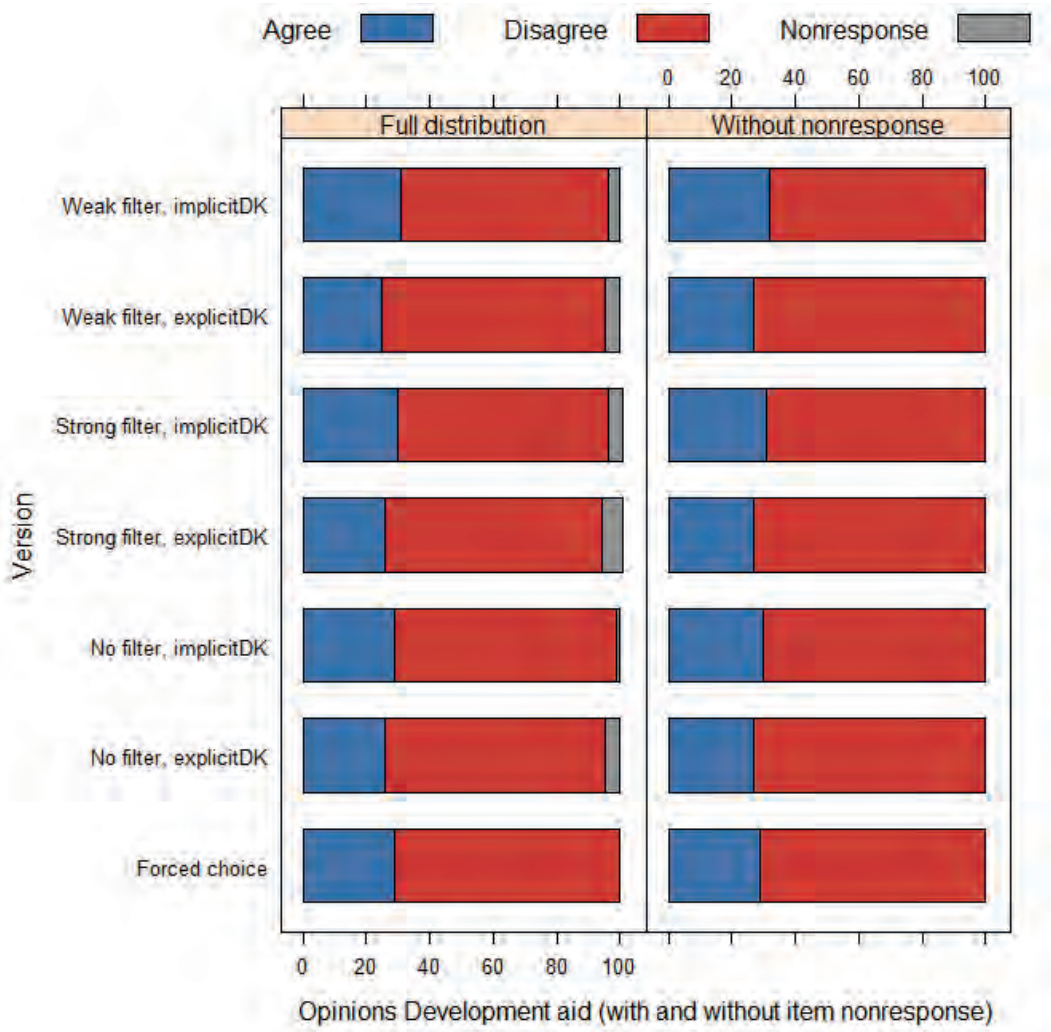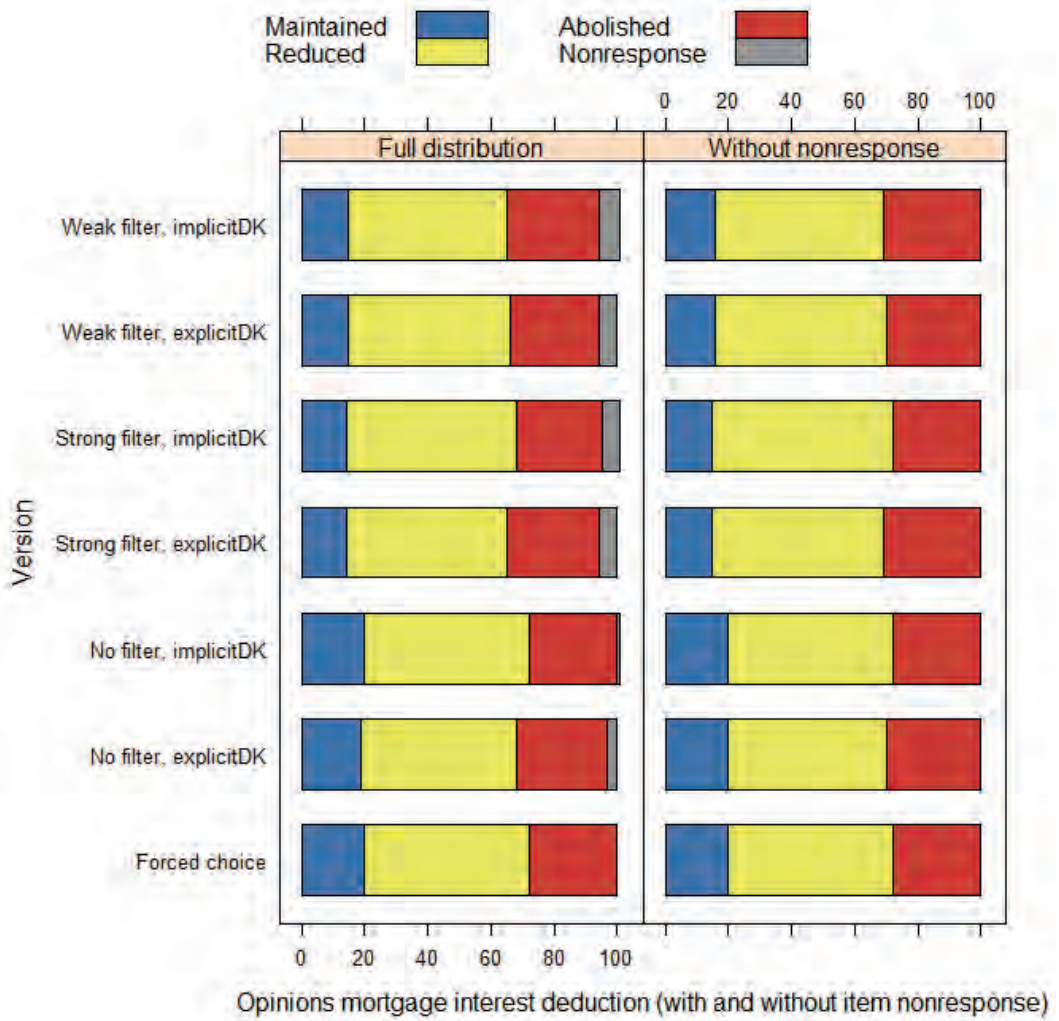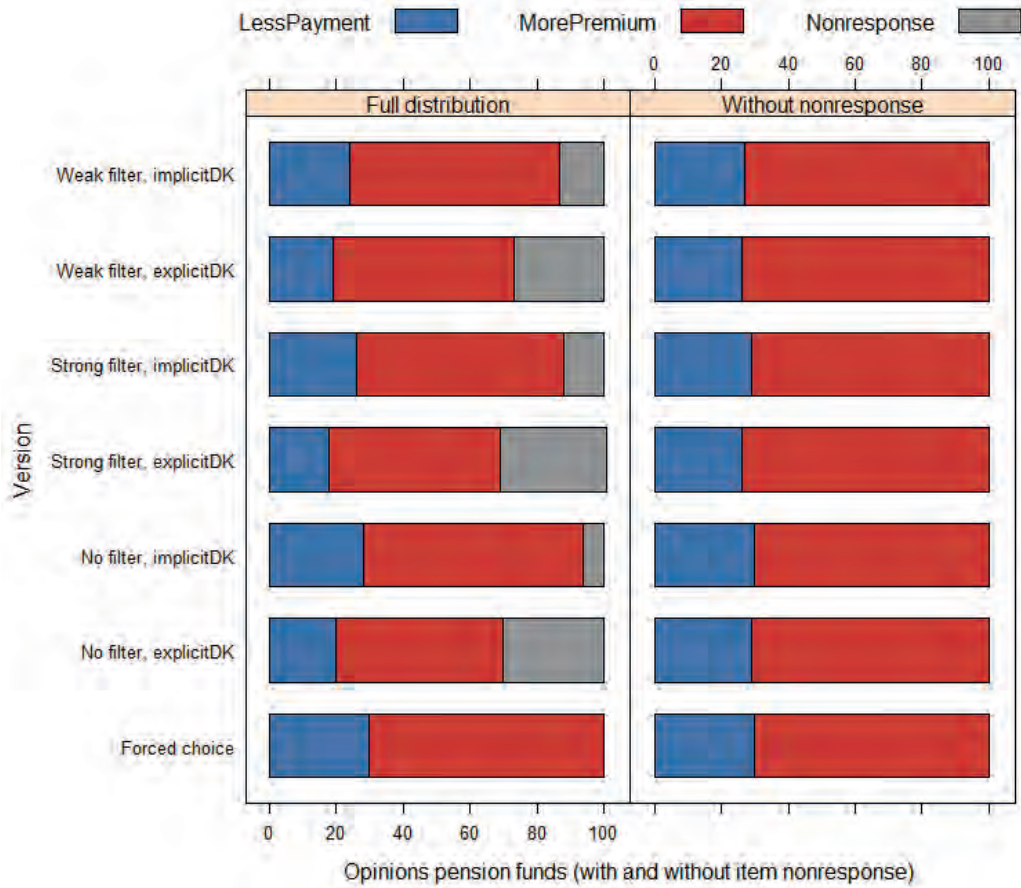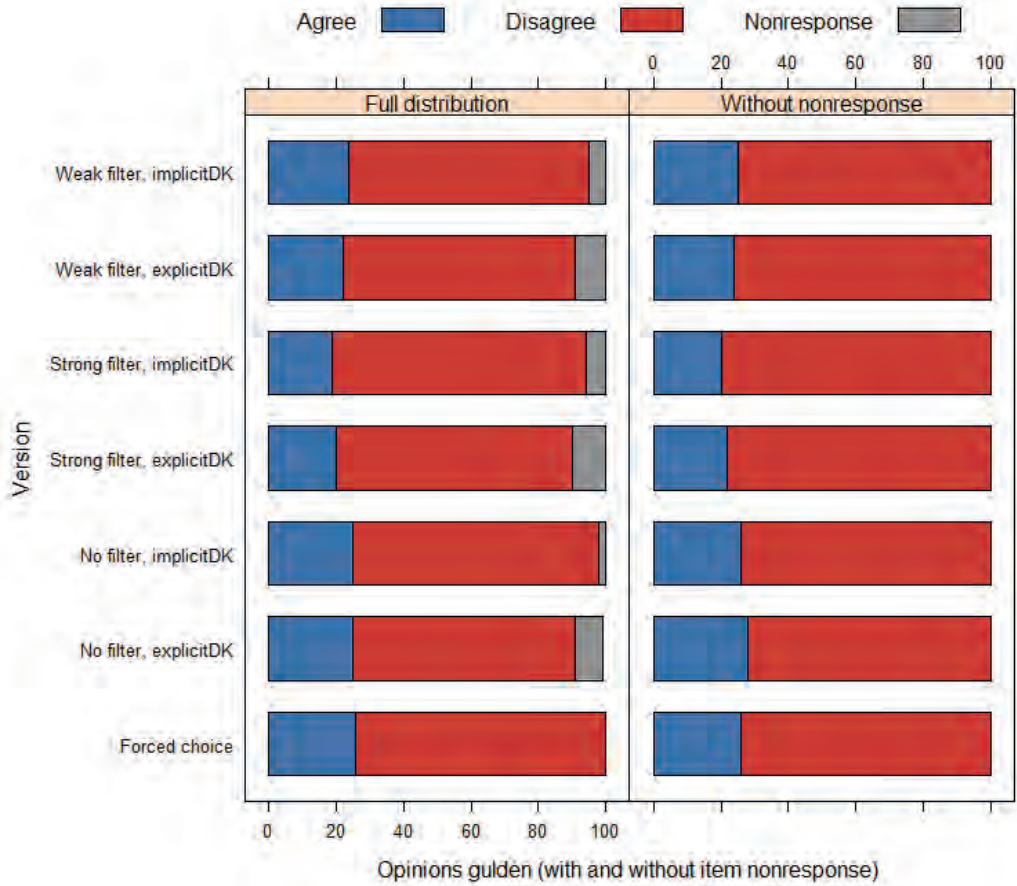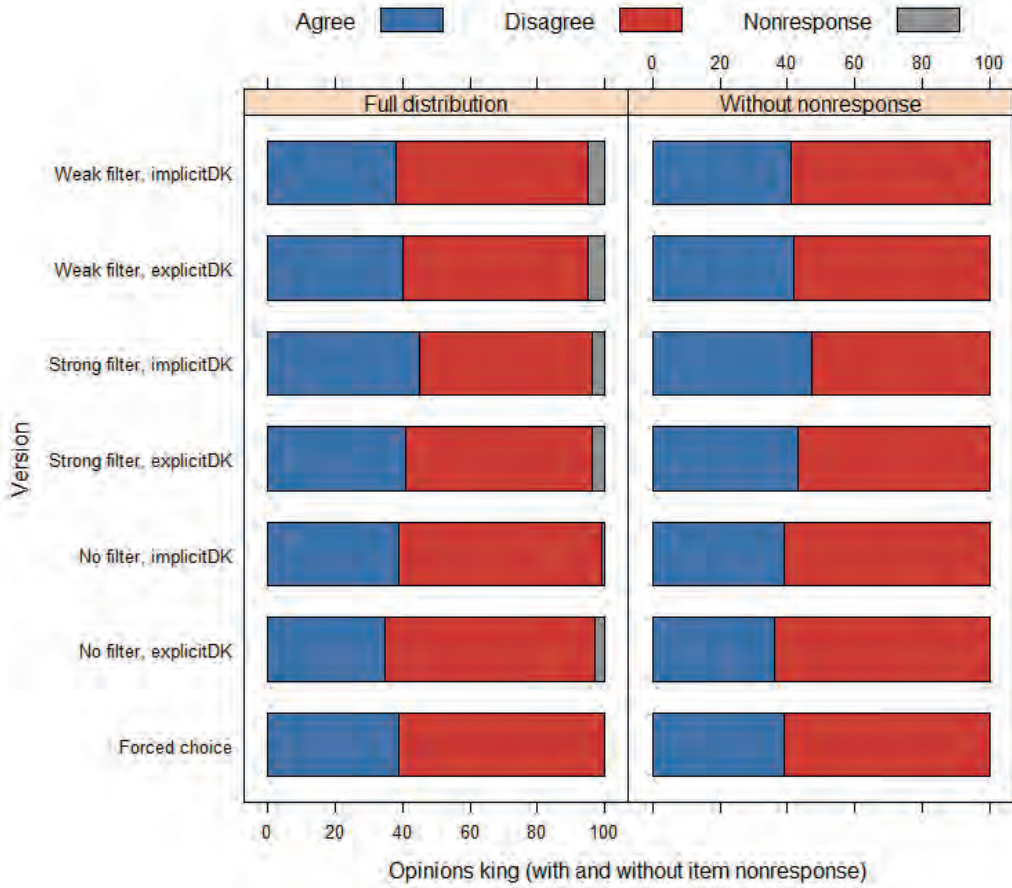*Figure C.42: Distribution (%) of Opinions **SAME-SEX ADOPTION** With and Without Item Nonresponse*

Figure C.43: Distribution (%) of Opinions **NON-DUTCH NATIONALITY** *With and Without Item Nonresponse*

*Figure C.44: Distribution (%) of Opinions **DEVELOPMENTAL AID** With and Without Item Nonresponse*

Figure C.45: Distribution (%) of Opinions **MORTGAGE INTEREST DEDUCTION** With and Without Item Nonresponse

*Figure C.46: Distribution (%) of Opinions **PENSION FUNDS** With and Without Item Nonresponse*

*Figure C.47: Distribution (%) of Opinions **GULDEN** With and Without Item Nonresponse*

Figure C.48: Distribution (%) of Opinions **KING** With and Without Item Nonresponse
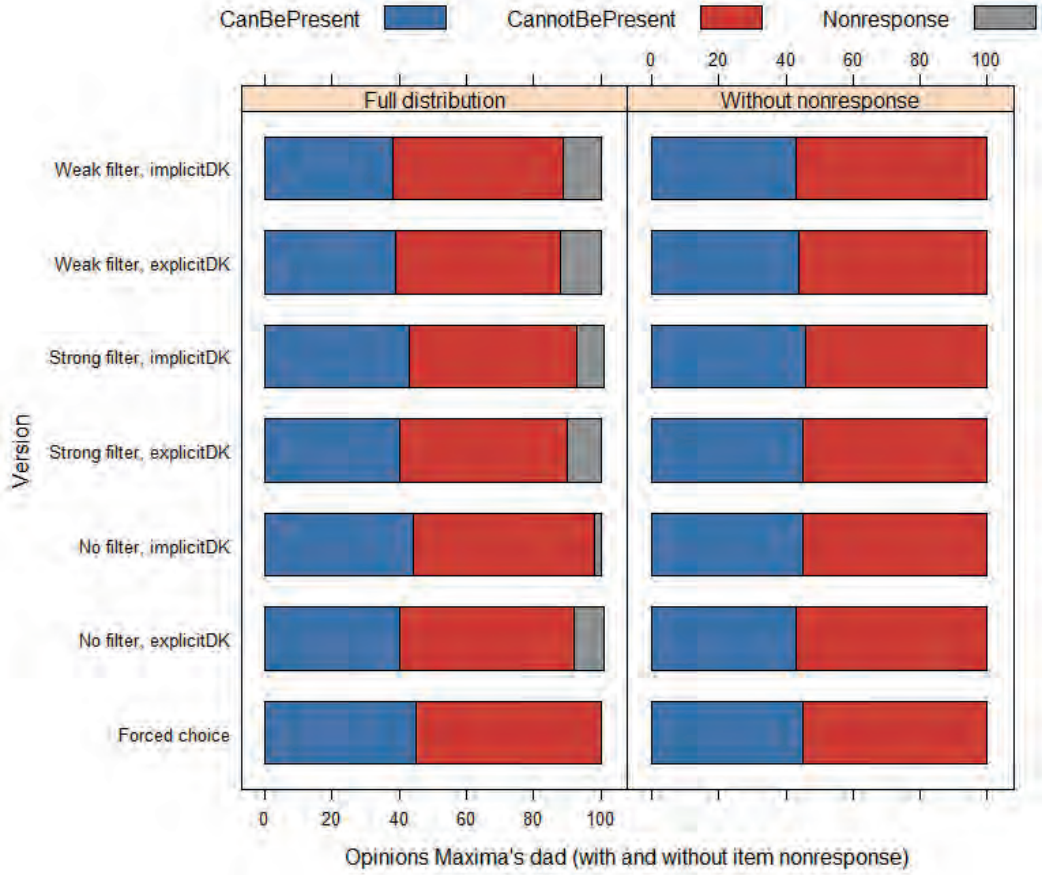
*Figure C.49: Distribution (%) of Opinions **MAXIMA'S FATHER** With and Without Item Nonresponse*
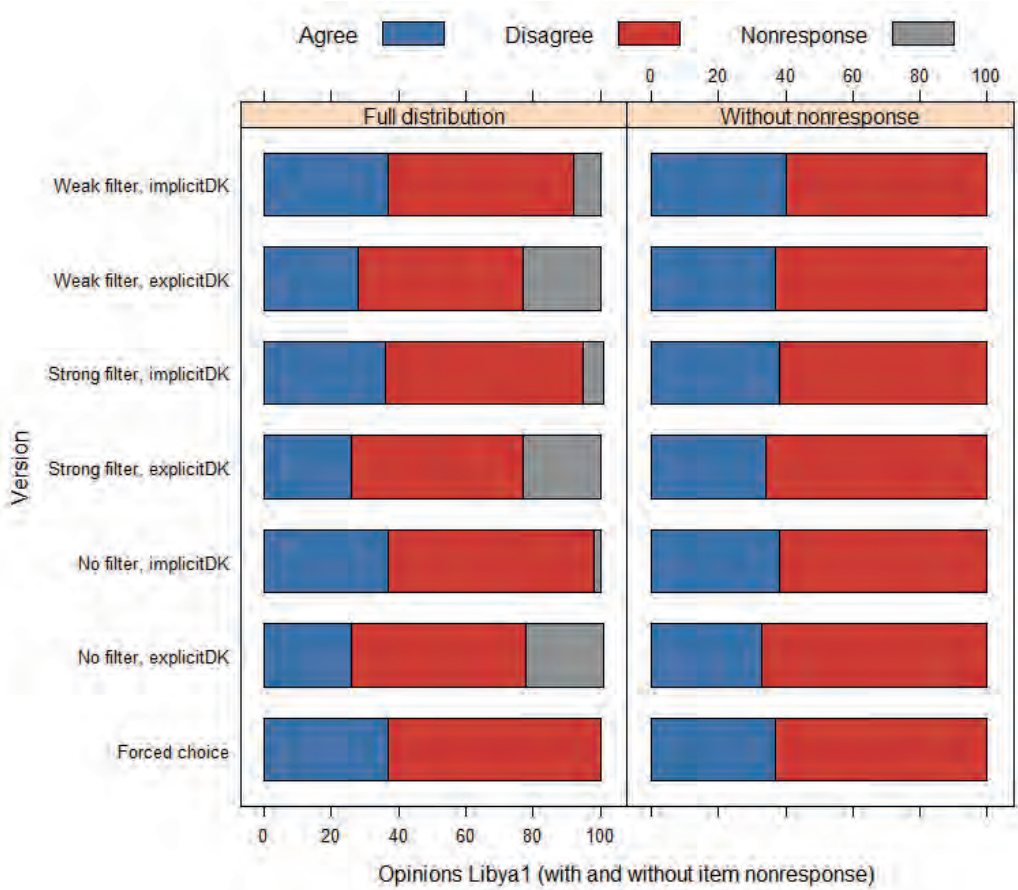
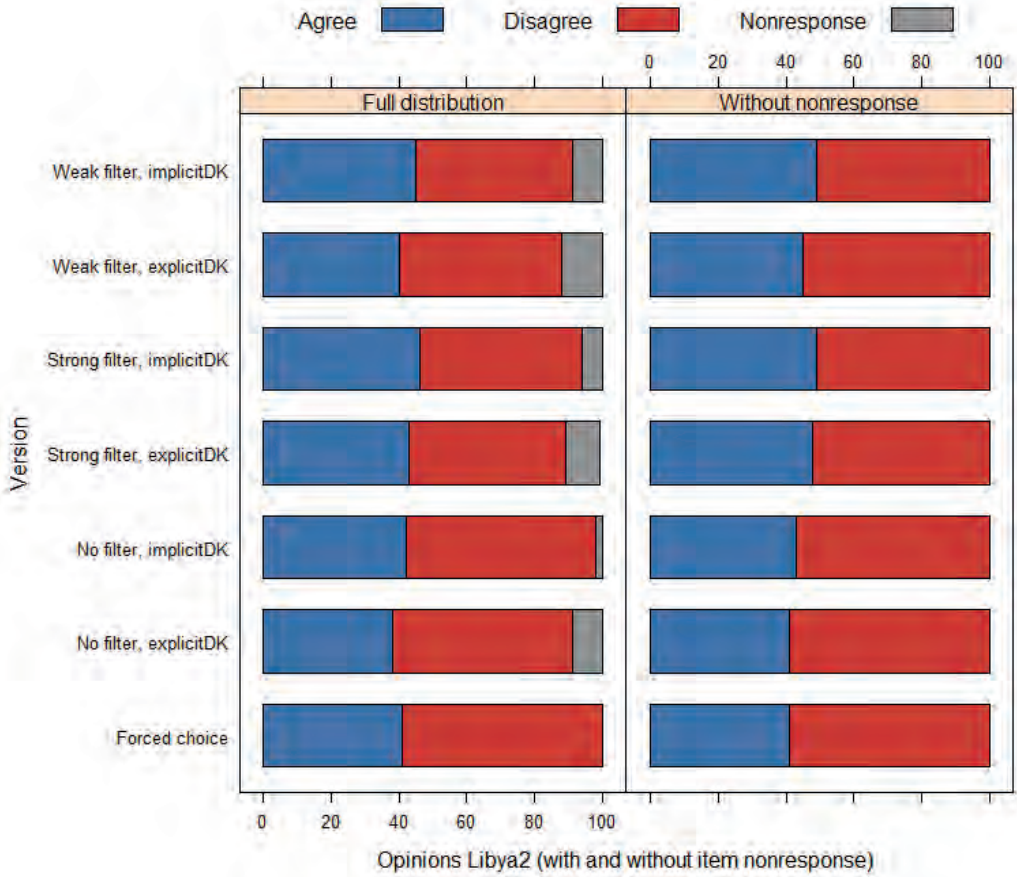*Figure C.50: Distribution (%) of Opinions **LIBYA1** With and Without Item Nonresponse*

*Figure C.51: Distribution (%) of Opinions **LIBYA2** With and Without Item Nonresponse*
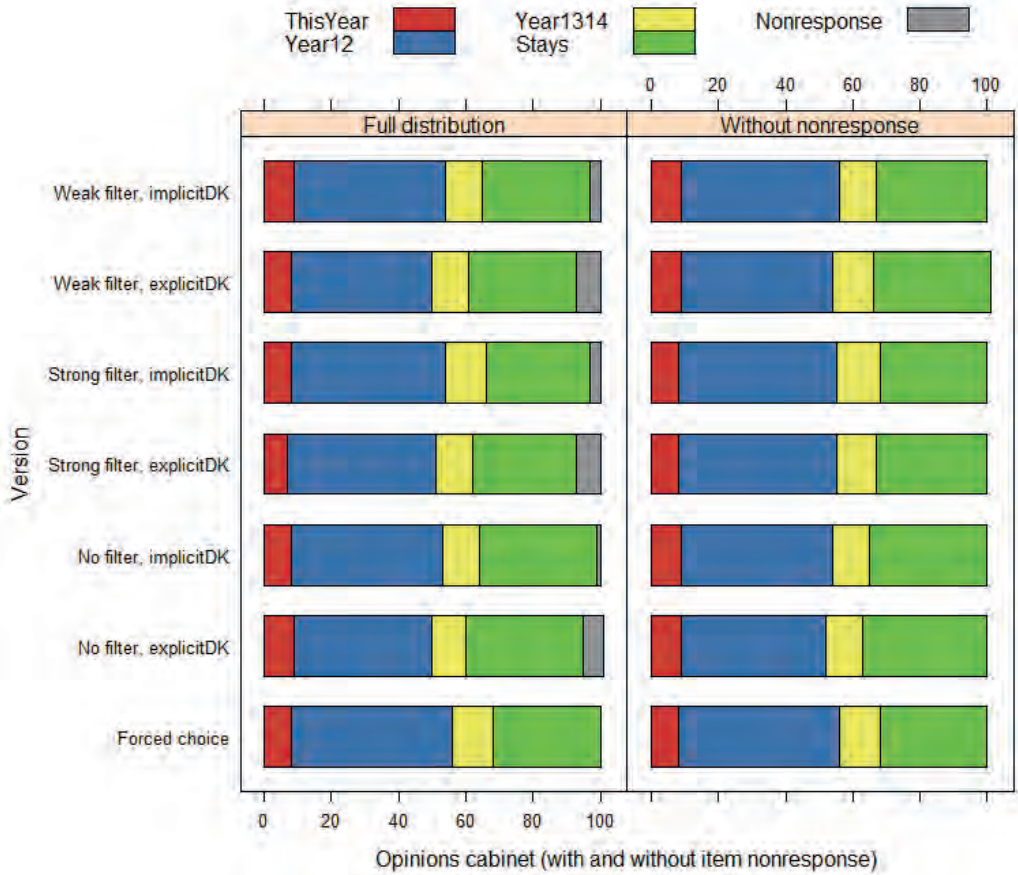
Figure C.52: Distribution (%) of Opinions **CABINET** With and Without Item Nonresponse
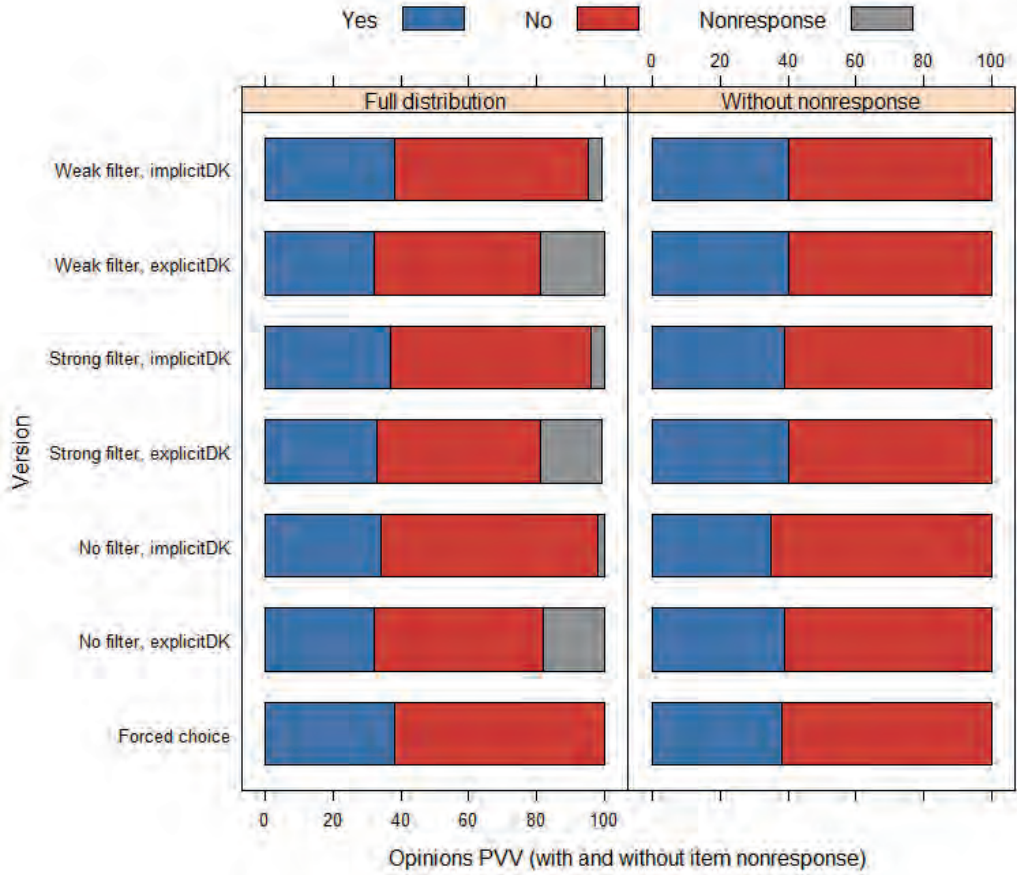
*Figure C.53: Distribution (%) of Opinions **PVV** With and Without Item Nonresponse*

III   Additions to Chapter 7: Follow-Up Question

Table C.8 is an additional analysis of the percentage of directive opinions for all individual survey questions. Item nonresponse is excluded; only substantive answers to the opinion question (which were followed by a follow-up question) are included.

Table C.8: Directive Opinions (%) Individual Survey Questions – Excluding Item Nonresponse

Issue		1A: single explicit DK	1B: single implicit DK	1C: forced choice	2A: filter single explicit	2B: filter single implicit	2C: filter forced choice
Self-Placement Income Differences	Very Upset	17	13	15	17	19	19
	Upset	57	59	52	62	57	50
The welfare should be lowered in order to stimulate people to work	Very Upset	21	21	16	22	23	18
	Upset	55	52	57	55	54	56
Self-Placement Euthanasia	Very Upset	29	28	31	33	36	34
	Upset	58	57	52	54	51	50
Adoption by same-sex couples should be possible	Very Upset	17	17	17	20	23	21
	Upset	50	49	48	49	45	45
Self-Placement Foreigners	Very Upset	19	21	14	20	18	17
	Upset	57	55	60	60	61	60
There are too many people of a non-Dutch nationality living in the Netherlands	Very Upset	13	15	14	18	17	18
	Upset	55	51	50	53	51	54
Self-Placement EU	Very Upset	20	16	14	20	21	19
	Upset	53	54	50	57	58	61
The Netherlands should spend more money on developmental aid	Very Upset	19	22	18	25	24	20
	Upset	50	48	52	51	48	57
The use of softdrugs should be completely prohibited	Very Upset	21	19	14	20	17	19
	Upset	42	48	47	50	56	50
'Establishing a 'Weed Permit' is a good idea'	Very Upset	18	15	10	15	16	18
	Upset	43	44	46	48	45	50
Powers EU	Very Upset	17	13	12	13	17	16
	Upset	53	53	57	59	56	63
'The Nederlands should abolish the mortgage interest deduction completely'	Very Upset	33	32	29	36	40	30
	Upset	46	41	47	49	49	54

Table C.8: Continued

Issue		1A: single explicit DK	1B: single implicit DK	1C: forced choice	2A: filter single explicit	2B: filter single implicit	2C: filter forced choice
'I want the Burqa Ban to proceed'	Very Upset	29	26	27	32	31	32
	Upset	48	41	38	48	42	47
'NATO should intervene in Syria'	Very Upset	19	13	14	22	15	22
	Upset	56	50	51	56	55	57
N		251	249	250	251	250	252

Figure C.54 to Figure C.61 display the distributions of opinions (including item nonresponse and the distinction between directive and permissive opinions) for all six variants. The self-placement items, powers of the EU and mortgage interest deduction are excluded. The substantive answer categories (completely agree, agree, disagree, completely disagree) of the welfare, same-sex adoption, non-Dutch nationality and developmental aid items were dichotomized.
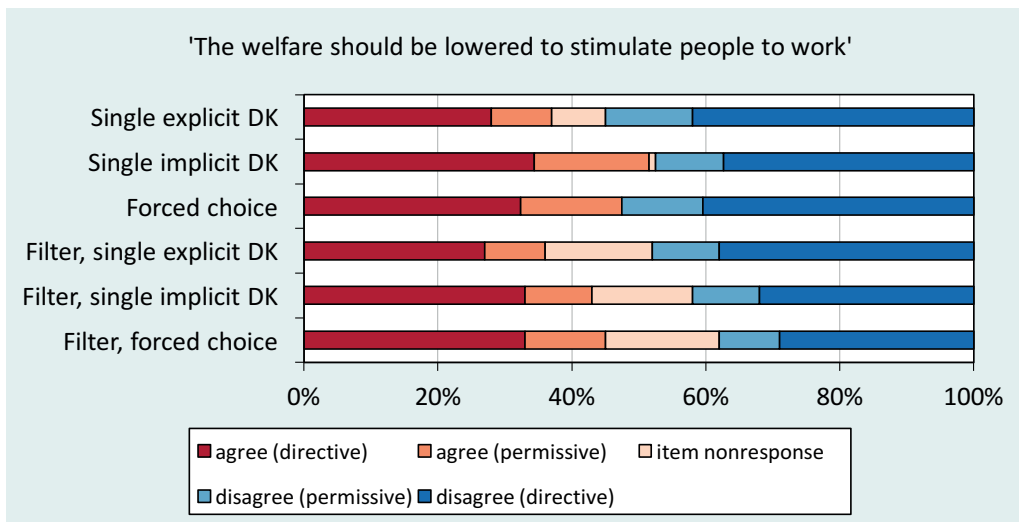

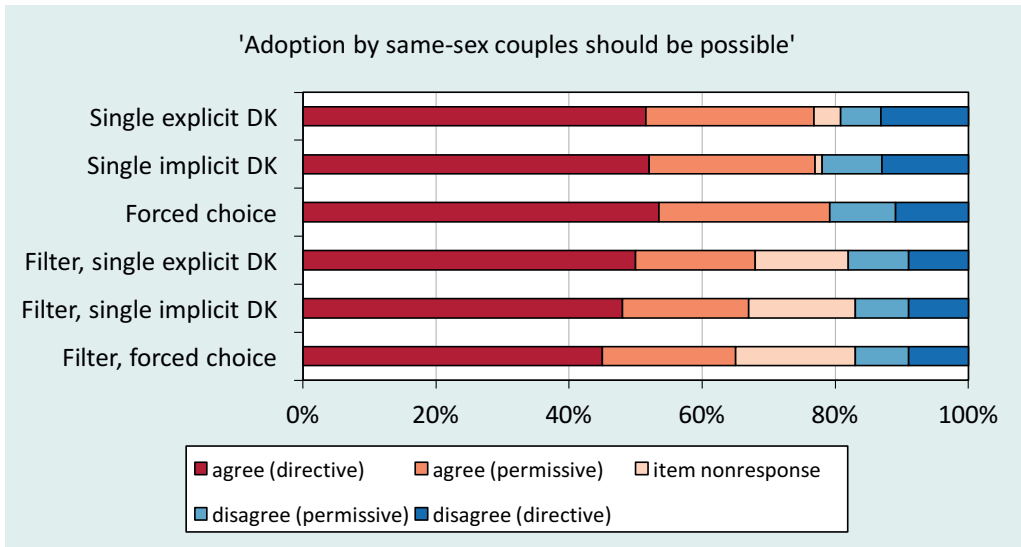
Figure C.54: Distribution of Opinions Lowering Welfare

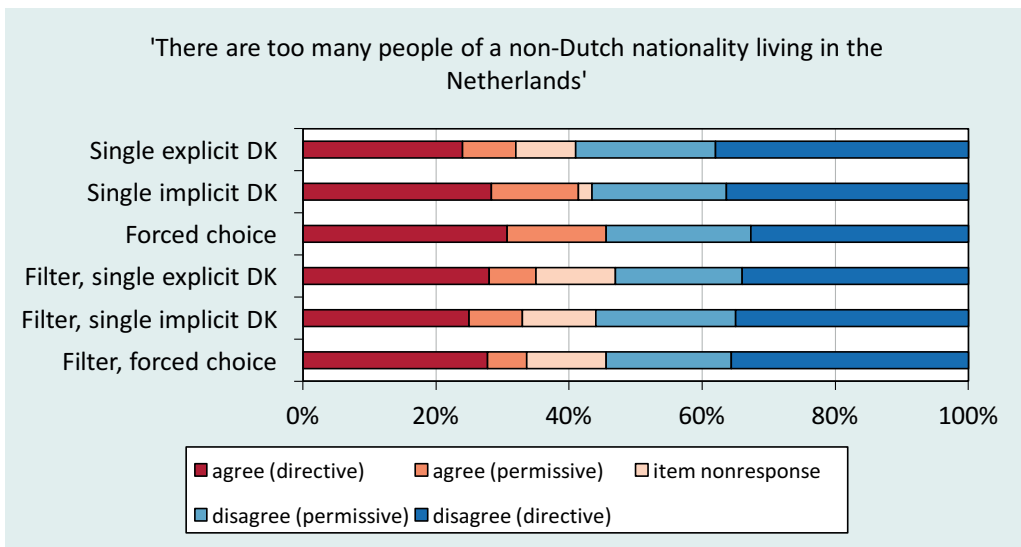*Figure C.55: Distribution of Opinions Same-Sex Adoption*

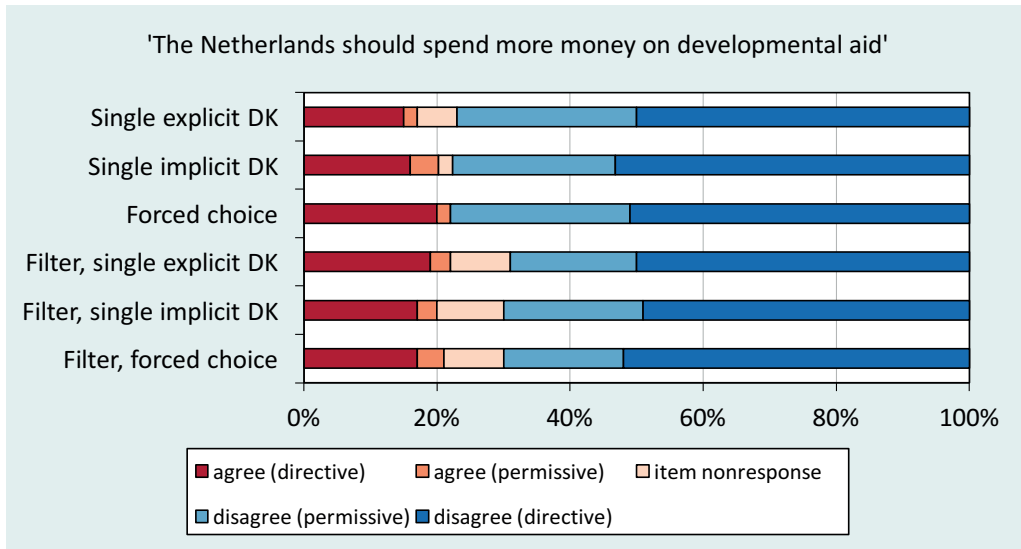*Figure C.56: Distribution of Opinions Non-Dutch Nationality*

'The Netherlands should spend more money on developmental aid'

Single explicit DK

Single implicit DK

Forced choice

Filter, single explicit DK

Filter, single implicit DK

Filter, forced choice

0%   20%   40%   60%   80%   100%

■ agree (directive)   ■ agree (permissive)   □ item nonresponse
□ disagree (permissive) ■ disagree (directive)

Figure C.57: Distribution of Opinions Developmental Aid

'The use of softdrugs should be completely prohibited'

Single explicit DK

Single implicit DK

Forced choice

Filter, single explicit DK

Filter, single implicit DK

Filter, forced choice

0%   20%   40%   60%   80%   100%

■ agree (directive)   ■ agree (permissive)   □ item nonresponse
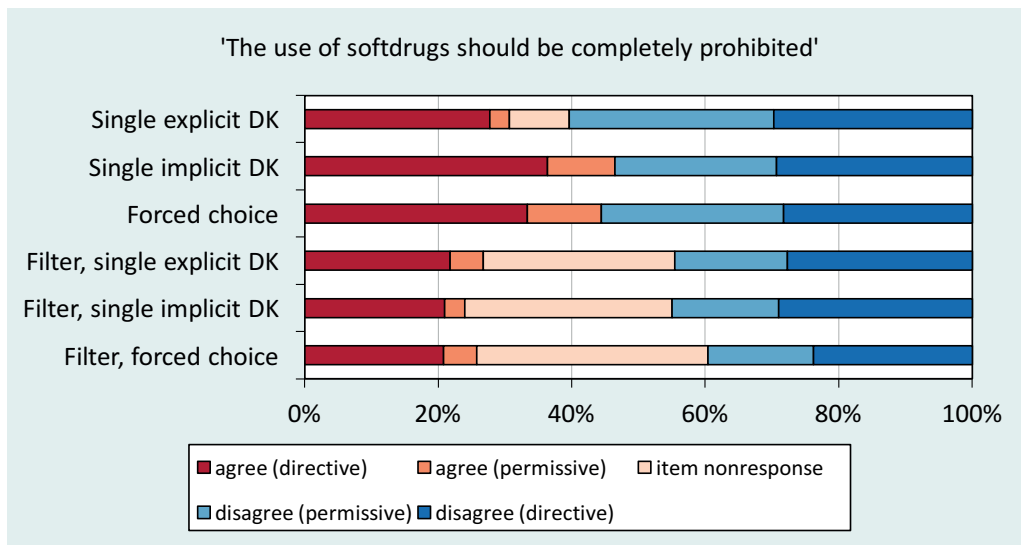□ disagree (permissive) ■ disagree (directive)

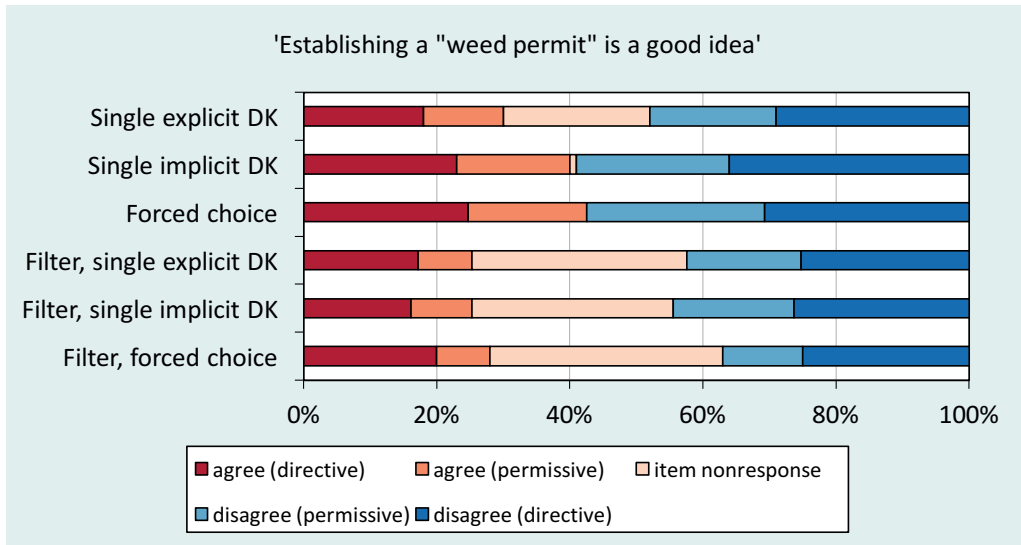*Figure C.58: Distribution of Opinions Softdrugs*

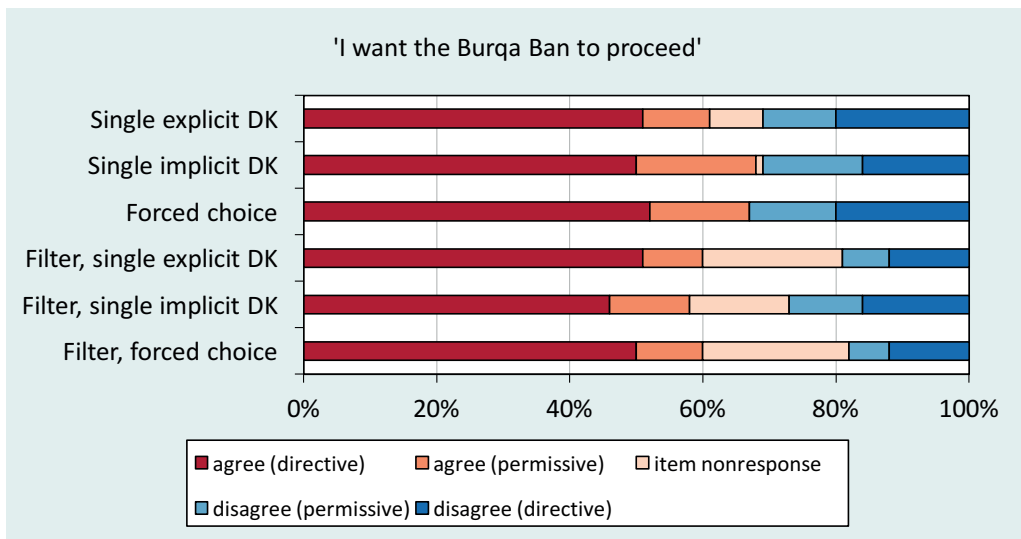*Figure C.59: Distribution of Opinions Weed Permit*

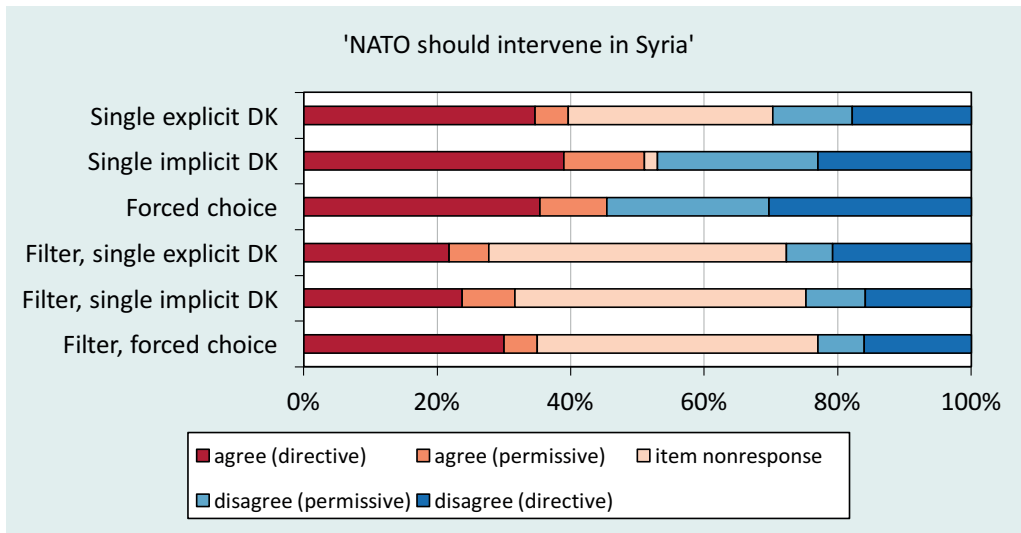*Figure C.60: Distribution of Opinions Burqa Ban*

Figure C.61: Distribution of Opinions NATO in Syria

IV   Additions to Chapter 8: Comparing Survey Experiments

Figure C.62 contains an additional analysis of the average total item nonresponse resulting from specific questionnaire variants. The analysis in Chapter 8 displayed the effect of a single non-substantive response option, i.e. a DK option or filter question. Figure C.62 displays the total item nonresponse, combining two non-substantive response options for the filter variants.
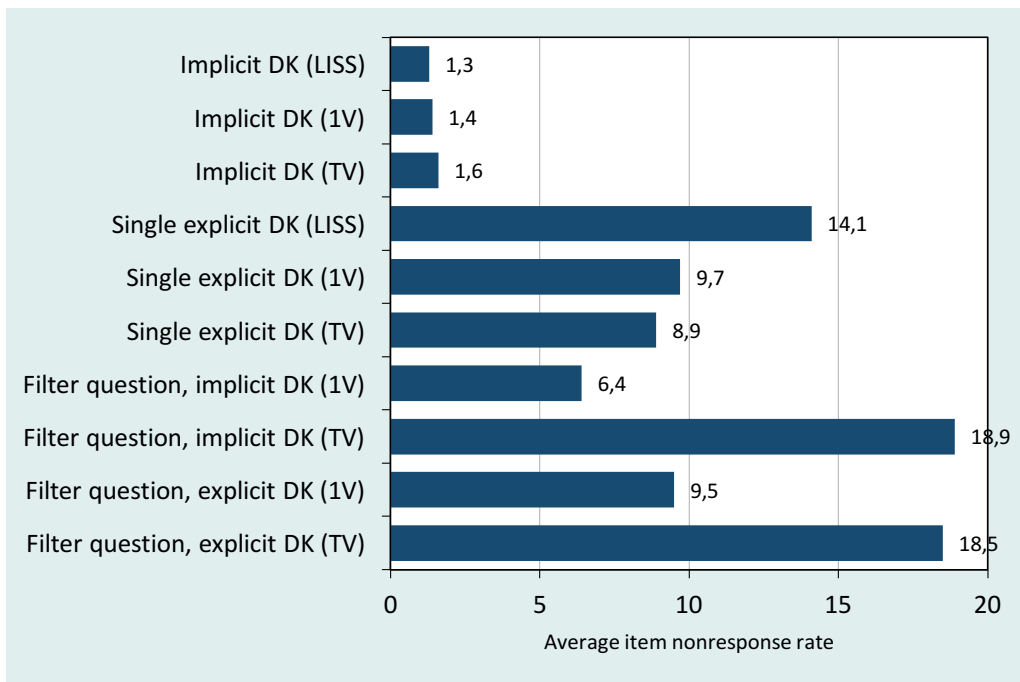

Figure C.62: Average Total Item Nonresponse (%)

The average total item nonresponse was computed by adding up the level of item nonresponse of each issue in one variant of the questionnaire, both as a DK answer and as a 'no' to the preceding filter question, and dividing this sum by the total number of questions in the questionnaire. Since no DK option was offered in the Forced Choice variants, these are excluded from this analysis.

Dutch Summary
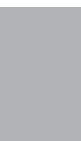
Publieke opinie zonder opinies? Item nonrespons en (de afwezigheid van) inhoudelijke meningen in surveys

Dit onderzoek gaat over zogeheten item nonrespons en niet-inhoudelijke antwoorden bij survey- of vragenlijstonderzoek. Item nonrespons houdt, kort gezegd, in dat een deelnemer aan of respondent bij een survey specifieke vragen (of 'items') niet heeft beantwoord. Als gevolg hiervan ontbreken inhoudelijke antwoorden op afzonderlijke vragen, waarbij het veelal om gesloten vragen gaat waarbij de respondent een beperkt aantal antwoordmogelijkheden krijgt aangeboden. Die ontbrekende antwoorden als gevolg van item nonrespons zijn van belang en onderzoek naar dergelijke ontbrekende inhoudelijke antwoorden is belangrijk, en wel om twee redenen. Allereerst vormt een hoge mate van item nonrespons een bedreiging van de geldigheid of validiteit van de resultaten en voor de mate waarin de bevindingen van steekproef naar populatie gegeneraliseerd kunnen worden, aangezien a) minder data beschikbaar zijn, waardoor (statistische) analyses beperkingen ondervinden; en b) item nonrespons allicht niet willekeurig 'verdeeld' is, in welk geval resultaten mogelijk niet valide zijn of een vertekening bevatten. Ten tweede verschaffen niet-inhoudelijke antwoorden waardevolle informatie over welk deel van de populatie of het publiek bepaalde vragen in een survey niet kan of wil beantwoorden.

De meeste studies met betrekking tot dit onderwerp leggen de nadruk op het effect van niet-inhoudelijke antwoordcategorieën op item nonrespons en bezien hoeveel niet-inhoudelijke antwoorden worden geregistreerd (bijv. Bishop, 2005; Bradburn, Sudman & Wansink, 2004; Schuman & Presser, 1996). Daarnaast is er onderzoek naar en literatuur over de vraag of de ontbrekende data, oftewel item nonrespons of niet-inhoudelijke antwoorden, al dan niet willekeurig verdeeld zijn (Tourangeau et al., 2013; De Leeuw et al., 2003). Wat grotendeels ontbreekt, is nadrukkelijke, gerichte aandacht voor het beeld van de publieke opinie dat uiteindelijk uit de survey ontstaat of wordt ontwikkeld. Hoe ziet dat beeld eruit als de niet-inhoudelijke antwoorden op uiteenlopende manieren worden geregistreerd? Hebben verschillende manieren om item nonrespons te faciliteren en registreren uiteenlopende gevolgen voor het resulterende beeld van de publieke opinie?

De centrale onderzoeksvraag van deze studie is: *hoe beïnvloedt het vraagontwerp van niet-inhoudelijke antwoordcategorieën de uitkomsten van een survey?* De nadruk ligt in het bijzonder op het gebruik van de 'weet niet' antwoordcategorie, de zogeheten filtervraag en de vervolgvraag. Zowel de niet-inhoudelijke antwoorden als de verdeling van de verkregen meningen, oftewel de inhoudelijke resultaten, worden bestudeerd als betekenisvolle uitkomsten. Het doel van het voorliggende onderzoek is te kijken of de

verschillende manieren waarop niet-inhoudelijke antwoorden geregistreerd kunnen worden, invloed hebben op de resultaten voor specifieke antwoordcategorieën.

Om de invloed van niet-inhoudelijke antwoordcategorieën te onderzoeken, zijn drie originele survey-experimenten uitgevoerd op basis van drie Nederlandse internetopiniepanels: de 'weet niet' optie bij het LISS Panel, de filtervraag bij het EenVandaag Opiniepanel, en de vervolgvraag bij het internetpanel van Team Vier. Door een *between-subjects-design* toe te passen en de respondenten gerandomiseerd of willekeurig aan de verschillende varianten van de vragenlijst toe te wijzen, werd de interne validiteit van de bevindingen gewaarborgd. De varianten verschilden per experiment slechts in de manier(en) waarop de niet-inhoudelijke antwoordcategorie werd(en) aangeboden. Hierbij moet overigens worden aangetekend dat de zogeheten externe validiteit van de twee *convenience samples* of gemakssteekproeven – het EenVandaag Opiniepanel en het internetpanel van Team Vier – enigszins beperkt is: omdat de steekproef niet aselect is getrokken, is generalisatie naar de populatie problematisch. Voor deze studie is dit probleem echter van ondergeschikt belang. Ten eerste is het uitvoeren van een survey met een internetpanel dagelijkse praktijk en zijn is de werkwijze en zijn de uitkomsten van de experimenten dan ook in die zin 'typisch' voor de momenteel dominante aanpak van opinieonderzoek. Ten tweede is het primaire doel van deze studie het bereiken van een hoge mate van interne validiteit en hiertoe zijn (survey-)experimenten bij uitstek geschikt.

Welk(e) effect(en) hebben niet-inhoudelijke antwoordcategorieën op item nonrespons? Vier mogelijkheden van deze niet-inhoudelijke antwoordcategorieën zijn verkend: geen niet-inhoudelijke antwoordcategorie (ook wel *forced choice* genoemd), de impliciete (weet niet) optie waarbij vragen kunnen worden overgeslagen, de expliciete weet niet optie (zowel enkelvoudig als antwoordcategorie als 'dubbel expliciet' met eveneens een verwijzing naar de weet niet optie in de vraag) en de filtervraag (in een zwakke en sterke formulering). Deze mogelijkheden werden onafhankelijk van elkaar bezien, of in een combinatie van een filtervraag gevolgd door een inhoudelijke meningsvraag plus weet niet optie. De belangrijkste verwachting was dat hoe explicieter een niet-inhoudelijke antwoordcategorie werd gepresenteerd, des te meer item nonrespons zich zou manifesteren. Deze verwachting wordt krachtig ondersteund door de data: de expliciete weet niet optie resulteert in meer item nonrespons dan de impliciete weet niet optie en de filtervraag resulteert in meer item nonrespons dan de expliciete weet niet optie in althans één van de toepassingen van de filtervraag. Het aanbieden van een niet-inhoudelijke antwoordcategorie moedigt respondenten kennelijk aan om daadwerkelijk een niet-inhoudelijk antwoord te geven, en het voorafgaan van een filtervraag aan een

(inhoudelijke) meningsvraag heeft een sterker effect dan een weet niet optie. De omvang van dit effect hangt overigens enigszins samen met het specifieke panel waarmee de data verzameld zijn en met het betreffende experiment; de filtervraag had niet het sterkste effect op item nonrespons in alle experimenten. De filtervraag resulteerde in ongeveer 5 à 6 procent (EenVandaag Opiniepanel) of ongeveer 18 procent item nonrespons (Team Vier internetpanel), wat aanzienlijk en opmerkelijk lager was dan eerdere bevindingen deden vermoeden (Bishop, 2005; Bishop *et al.*, 1983; Eckman *et al.*, 2014; Schuman & Presser, 1979). Niettemin hadden zowel de mate van explicietheid als het type niet-inhoudelijke antwoordcategorieën duidelijk invloed op het relatieve aantal niet-inhoudelijke antwoorden.

Een tweede belangrijke vraag was of en zo ja hoe de inhoudelijke verdeling van meningen eventueel verandert door het vraagontwerp, in het bijzonder door de beschikbare niet-inhoudelijke antwoordcategorie. Een hogere mate van item nonrespons biedt weliswaar in beginsel meer mogelijkheden voor het optreden van vertekening, maar of daadwerkelijk een dergelijke vertekening plaatsvindt, is mede afhankelijk van de vraag of de antwoorden of data willekeurig (*at random*) ontbreken of niet. Het resulterende beeld van de publieke opinie, opgevat als aggregatie van individuele opinies, bleek echter nauwelijks beïnvloed te worden door dit aspect van het vraagontwerp. De uitkomsten waren robuust en impliceren een *at random* ontbreken van inhoudelijke antwoorden en geven, op enkele uitzonderingen na, geen aanleiding tot zorgen over of en hoe respondenten kunnen aangeven dat zij geen mening hebben of willen geven, in ieder geval niet voor een indruk van wat 'het publiek' wil en vindt. Het aanbieden van niet-inhoudelijke antwoordcategorieën is allicht te prefereren om vast te stellen welk deel van het publiek een mening heeft en geeft, maar voor een valide indruk van de publieke opinie of van welk beleid(svoorstel) de voorkeur krijgt in de samenleving maakt het - verrassend - geen substantieel verschil of een niet-inhoudelijke antwoordcategorie wordt aangeboden of niet. Kort samengevat: het effect van het aanbieden van een niet-inhoudelijke antwoordcategorie is dat (de mate van) item nonrespons *wel* verandert, maar *niet* het daaruit voorkomende beeld van de publieke opinie in termen van absolute of relatieve meerderheden onder de relevante populatie.

De verwachting was dat het effect van een niet-inhoudelijke antwoordcategorie zou variëren voor inhoudelijk verschillende onderwerpen, waarbij meer specifiek verwacht werd dat onderwerpen die gerelateerd zijn aan een belangrijke onderliggende politieke scheidslijn in minder item nonrespons zouden uitmonden en onderwerpen die te maken hebben met het voor een groot publiek 'lastige' buitenlands beleid in meer item nonrespons. De resultaten waren echter niet

eenduidig: beide verwachtingen werden in twee van de drie experimenten ondersteund. Meer onderzoek is nodig om nader te kunnen differentiëren naar inhoud.

Het derde belangwekkende element dat in deze studie werd onderzocht is de vervolgvraag. In de hier onderzochte vervolgvraag werd respondenten gevraagd hoe erg zij het zouden vinden als hun mening, dat wil zeggen de opvatting die zij in antwoord op de voorafgaande vraag hebben geuit, niet vervolgens in beleid zou worden omgezet. Met behulp van deze vraag kan een onderscheid tussen meer 'permissieve' en 'directieve' meningen worden aangebracht. Permissieve meningen, oftewel opvattingen van mensen die het niet erg zouden vinden als geen beleidsmatig vervolg aan hun mening wordt gegeven, kunnen, evenals item nonrespons, niet of althans relatief moeilijk worden gezien als richtinggevende input voor bijvoorbeeld beleidsmakers en politici. In die zin kunnen permissieve meningen beschouwd worden als niet-inhoudelijke antwoorden: een deel van het publiek maakt het niet uit welk beleid wordt nagestreefd, zelfs al hebben zij een bepaalde mening (gegeven) ten aanzien van het betreffende 'dossier'. Ongeveer twee derde van de respondenten uitte een directieve mening, wat meer was dan verwacht (met name op basis van Moore, 2008). Verder werd de verwachting dat onderwerpen die gerelateerd konden worden aan een belangrijke politieke dimensie in meer directieve meningen zouden uitmonden en dat onderwerpen gerelateerd aan buitenlands beleid in relatief minder directieve meningen zouden uitmonden, empirisch ondersteund.

Effecten van niet-inhoudelijke antwoordcategorieën op de uitkomsten van een survey vormden de kern van deze studie. Daarnaast werd een aantal andere methodologische vragen geadresseerd. Vragen met meer inhoudelijke antwoordcategorieën resulteerden in minder item nonrespons en bij varianten zonder niet-inhoudelijke antwoordcategorie, of met een minder expliciete optie, gebruikten respondenten de zogenoemde middencategorie van een antwoordschaal relatief vaak. Deze bevindingen zijn echter onder enig voorbehoud, omdat zij niet de kern van het hier gepresenteerde onderzoek betreffen; meer systematische en uitgebreide analyses zijn nodig.

Bogart (1972) had gelijk dat de vraag *wat* mensen vinden van onderwerpen in zekere zin secundair en ondergeschikt is aan de vraag *of* zij er iets van vinden en erover nadenken. Dit wordt weergegeven door de niet-inhoudelijke antwoordcategorieën, die in deze studie op basis van eigen origineel onderzoek dan ook de aandacht kregen die zij verdienen.

Curriculum Vitae

Jannine van de Maat (1985, Rijssen) is a political scientist and survey methodologist. She attended pre-university education (VWO) at Reggesteyn in Rijssen and Nijverdal, the Netherlands, and graduated in 2004. She obtained a Bachelor's degree in Political Science at Leiden University in 2007 and a Research Master's degree at the same university in 2009. At the same time she worked as a teaching assistant at the Institute for Political Science. She conducted her PhD research at the Institute for Political Science at Leiden University, first as a part of a PhD appointment (2009-2013) and later alongside a contract as Lecturer in Political Science (2013-2017). She has presented her work at various international conferences, including the ECPR Joint Sessions, the ISA World Congress of Sociology, the annual WAPOR conference and the ESRA conference. She also worked as a survey methodologist and advisor for the research project Confronting Caribbean Challenges (2014-2016, KITLV). She currently works as a Methodologist and Datamanager at the Netherlands Institute for Social Research (Sociaal en Cultureel Planbureau, SCP).