UNIVERZA V LJUBLJANI

FAKULTETA ZA DRUŽBENE VEDE

Jernej Berzelak

Učinki načina anketiranja v spletnih anketah

Mode effects in web surveys

Doktorska disertacija

Ljubljana, 2014

UNIVERZA V LJUBLJANI

FAKULTETA ZA DRUŽBENE VEDE

Jernej Berzelak

Mentor: prof. dr. Vasja Vehovar
Somentorica: prof. dr. Edith D. de Leeuw

Učinki načina anketiranja v spletnih anketah

Mode effects in web surveys

Doktorska disertacija

Ljubljana, 2014

Acknowledgments

The statement of authorship on next page suggests that this dissertation is the product of my sole research work. This is far from the truth. The dissertation would not have happened without many people who supported my work.

I would like to thank Vasja Vehovar for his supervision and productive discussions on important conceptual issues. My sincere thanks go to Edith de Leeuw for her detailed and cheerful feedback on my work. Kind suggestions for improvements were also provided by my committee members, Katja Lozar Manfreda and Mojca Bavdaž.

I have greatly benefited from the opportunity to collaborate with the GGP project team. I am very grateful to Gregor Petrič and Andrej Kveder, who invited me to get involved in the project activities. Many thanks also to Rok Platinovšek for his generous help with the preparation of the empirical data for analysis.

This work would not have been done without my friends, who prevented me from becoming asocial. And finally, but by no means less important, I am indebted to my family for their patience, outstanding support, and assistance.

The research was part-financed by the Slovenian Research Agency and the EU 7th Framework Programme.

Univerza
v Ljubljani

Fakulteta
za družbene vede

Kardeljeva ploščad 5
1000 Ljubljana, Slovenija
telefon 01 58 05 120

I Z J A V A O A V T O R S T V U
doktorske disertacije

Podpisani Jernej Berzelak, z vpisno številko 74080800, sem avtor doktorske disertacije z naslovom: Učinki načina anketiranja v spletnih anketah (Mode Effects in Web Surveys).

S svojim podpisom zagotavljam, da:

- je predložena doktorska disertacija izključno rezultat mojega lastnega raziskovalnega dela;
- sem poskrbel, da so dela in mnenja drugih avtorjev oz. avtoric, ki jih uporabljam v predloženem delu, navedena oz. citirana v skladu s fakultetnimi navodili;
- sem poskrbel, da so vsa dela in mnenja drugih avtorjev oz. avtoric navedena v seznamu virov, ki je sestavni element predloženega dela in je zapisan v skladu s fakultetnimi navodili;
- sem pridobil vsa dovoljenja za uporabo avtorskih del, ki so v celoti prenesena v predloženo delo in sem to tudi jasno zapisal v predloženem delu;
- se zavedam, da je plagiatorstvo – predstavljanje tujih del, bodisi v obliki citata bodisi v obliki skoraj dobesednega parafraziranja bodisi v grafični obliki, s katerim so tuje misli oz. ideje predstavljene kot moje lastne – kaznivo po zakonu (Zakon o avtorski in sorodnih pravicah (UL RS, št. 16/07-UPB3, 68/08, 85/10 Skl.US: U-I-191/09-7, Up-916/09-16)), prekršek pa podleže tudi ukrepom Fakultete za družbene vede v skladu z njenimi pravili;
- se zavedam posledic, ki jih dokazano plagiatorstvo lahko predstavlja za predloženo delo in za moj status na Fakulteti za družbene vede;
- je elektronska oblika identična s tiskano obliko doktorske disertacije ter soglašam z objavo doktorske disertacije v zbirki »Dela FDV«.

V Ljubljani, dne 24.3.2014 Podpis avtorja: _____

Povzetek

Utemeljitev problema in cilji

Spletne ankete so postale eden vodilnih načinov za zbiranje anketnih podatkov. Ponujajo številne prednosti, predvsem visoko hitrost zbiranja podatkov, napredne funkcionalnosti vprašalnikov ter nižje stroške raziskovanja. Zlasti slednje postaja zaradi naraščajočih pritiskov po nižanju izdatkov v vseh sektorjih raziskovalne dejavnosti, ki od raziskovalcev zahtevajo iskanje optimalnih ravnovesij med kakovostjo podatkov in stroški, vse pomembnejše. Spletne ankete so zato obetajoča alternativa tradicionalnim načinom anketiranja, kot so osebne, telefonske in poštne ankete.

Uporaba kateregakoli načina anketiranja zahteva upoštevanje morebitnih negativnih vplivov na kakovost podatkov. V disertaciji se osredotočamo na problem učinkov načina anketiranja v spletnih anketah kot na potencialno oviro pri zagotavljanju točnosti ocen. Učinki načina anketiranja so v najsplošnejšem smislu sestavni del anketne napake, ki nastane zaradi uporabe določenega načina anketiranja. Temeljna teza naše obravnave problema je, da so učinki načina anketiranja v spletnih anketah rezultat nožice dejavnikov, povezanih ne le z načinom anketiranja, temveč tudi s specifičnimi implementacijami ankete.

Glavni namen disertacije je prispevati celostno sliko učinkov načina anketiranja v spletnih anketah, ki bo omogočila boljše razumevanje te problematike. V ta namen sledimo štirim osrednjim ciljem:

1. vzpostavitvi trdnejšega konceptualnega okvira učinkov načina anketiranja v spletnih anketah,
2. celostni analizi oblik in virov učinkov načina anketiranja,
3. empiričnemu prikazu nestanovitne narave problema ter
4. izpostavitvi implikacij za nadaljnje raziskovanje in anketno prakso.

Razvoj konceptualnega okvira

Vzpostavitev konceptualnega okvira učinkov načina anketiranja pričnemo z opredelitvijo temeljnih konceptov. Pregled literature kaže velike razlike v razumevanju in nezadostno opredeljenost terminov »način« in »učinek načina«, kar je presenetljivo glede na pogosto uporabo v anketni terminologiji. Za zagotovitev jasne uporabe teh pojmov v celotni disertaciji jih uvrstimo v širši okvir anketnega procesa in anketnih napak ter oblikujemo njihove operacionalne opredelitve. Način anketiranja opredeljujemo kot skupek postopkov za zbiranje podatkov, ki določajo osnovna načela komunikacije in prenosa informacij med anketirancem in anketnim vprašalnikom. Ta osnovna načela so utemeljena s šestimi inherentnimi značilnostmi načina anketiranja: z glavnim kanalom za predstavitev vprašanj (vhodnim kanalom), kanalom za podajanje odgovora (izhodnim

kanalom), vključenostjo anketarja, bližino interakcije med anketarjem in anketirancem, uporabo računalniške tehnologije za zbiranje podatkov ter medijem za prenos informacij. Nadalje identificiramo več izvedbeno specifičnih in kontekstualnih značilnosti načina, ki so odvisne od lastnosti določenega načina uporabe ankete.

Skladno s predlagano opredelitvijo razumemo učinke načina anketiranja kot vse neposredne in posredne učinke inherentnih značilnosti načina na točnost dobljenih anketnih ocen. Ker lahko vsi sestavni deli anketne napake medsebojno vstopajo v močne interakcije, učinki načina anketiranja niso omejeni le na neposreden vpliv inherentnih značilnosti, temveč so lahko posredovani tudi prek izvedbeno specifičnih in kontekstualnih značilnosti ter drugih dejavnikov.

V disertaciji obravnavamo vplive spletnega načina anketiranja na merske napake. Ker je ta vrsta napake tesno povezana z anketirančevimi odgovori, v razpravo vključujemo ustrezne vidike teorij odgovarjanja na anketna vprašanja. Posebno pozornost namenjamo modelu obdelave informacij, ki ga je razvil Tourangeau s sodelavci. Ta model nadalje razširjamo z nekaterimi drugimi prispevki na tem področju, vključno z drugimi modeli obdelave informacij, modeli odklonov v procesu odgovarjanja ter modeli interakcije med anketirancem in vprašalnikom.

Elaboracija virov učinkov načina anketiranja v spletnih anketah

Osrednji del disertacije predstavlja elaborat potencialnih virov učinkov načina anketiranja v spletnih anketah in njihovega vpliva na točnost ocen. Delo temelji na izčrpni evalvaciji in integraciji obstoječega empiričnega raziskovanja. Pri tem vključujemo primerjalne študije, ki se ukvarjajo z najrazličnejšimi dejavniki, ki vplivajo na razlike v ocenah med spletnimi anketami in drugimi načini anketiranja. Na tej osnovi oblikujemo razširitev konceptualnega modela učinkov načina, ki ga je predlagal Tourangeau s sodelavci. Z vzpostavitvijo potencialnih posredniških odnosov med inherentnimi, izvedbeno specifičnimi in kontekstualnimi značilnostmi spletnih anket pridobimo celovit vpogled v kritične dejavnike za nastanek učinkov načina anketiranja.

Elaborat izpostavlja samoanketiranje kot najsplošnejši vir učinkov načina anketiranja v spletnih anketah. Odsotnost anketarja lahko vodi do povečanja obremenitve anketiranca in hkrati izrazito zmanjšuje možnosti njegovega zunanjega motiviranja. S tem se pri anketirancu povečuje verjetnost za iskanje bližnjic v procesu odgovarjanja, kar vodi do manj celostne izvedbe procesa in posledično do nižanja točnosti odgovorov. Nekateri drugi potencialni viri učinkov načina anketiranja v spletnih anketah so še: izrazito poudarjene kontekstualne informacije zaradi vizualne predstavitve vprašanj, negativni vplivi računalniške izvedbe anketiranja in računalniško posredovane interakcije ter tehnično in fizično okolje, v katerem anketiranje poteka.

Analiza obstoječih raziskav odkriva tudi pomembne razlike v ugotovitvah in pogosto neskladne zaključke, kar močno nakazuje na nestanovitnost učinkov načina anketiranja. Čeprav predlagani konceptualni model opisuje kritične interakcije med značilnostmi načina anketiranja, ki lahko vodijo do pojava učinkov, njihov dejanski nastanek pogojuje

tudi veliko število drugih dejavnikov, povezanih z lastnostmi določene ankete in anketirancev.

Metodologija empirične študije

Za prikaz in nadaljnje odkrivanje narave učinkov načina anketiranja smo analizirali eksperimentalno izvedbo ankete *Generations and Gender Survey*. Analiza se je osredotočila na primerjavo 89 postavk iz vprašanj v obliki lestvice med spletnim, telefonskim (CATI) in osebnim (CAPI) načinom zbiranja podatkov. Hkratna vključitev velikega števila spremenljivk je koristna zlasti zaradi zagotavljanja celostnejšega opazovanja razlik med načini.

S pomočjo empirične študije smo preverili štiri splošne hipoteze. Dve se nanašata na odsotnost in dve na prisotnost negativnih vplivov spletnega načina anketiranja. Pričakovali smo, da bodo spletni anketiranci manj težili k upravljanju z vtisom (H_1) in redkeje izbirali skrajne odgovore na lestvici (H_2), vendar pa pogosteje izbirali sredinske odgovore (H_3) in se verjetneje posluževali nediferenciacije odgovorov kot posebne oblike zadovoljevanja (H_4).

Vzorec anketirancev smo pridobili s komercialnega spletnega panela, s čimer smo zmanjšali možnost vplivov napake nepokritja na primerljivost eksperimentalnih skupin. Za analizo podatkov smo uporabili dve metodi statističnega modeliranja: regresijsko analizo po metodi najmanjših kvadratov (OLS) in model parcialno sorazmernih obetov (GO-logit). S tem smo omogočili odkrivanje več vrst razlik med načini anketiranja, ki so ključni kazalci prisotnosti učinkov načina. Rezultate smo interpretirali z upoštevanjem statističnih značilnosti, prilagojenih večkratnim preizkusom statističnih predpostavk, ter ocen velikosti učinkov.

Rezultati empirične študije

Izsledki empirične študije delno potrjujejo zastavljene hipoteze. Večina analiziranih postavk kaže razmeroma majhne velikosti učinkov. Primerjava med spletno in telefonsko anketo je pokazala statistično značilne razlike v ocenah povprečij za 22 % postavk, primerjava med spletno in osebno anketo pa za 27 % postavk ($p < 0.01$). Veliko večje število postavk z značilnimi razlikami smo odkrili z analizo porazdelitev posameznih odgovorov. To kaže, da se nekateri učinki načina anketiranja ne odražajo na merah srednje vrednosti, vendar še vedno značilno vplivajo na verjetnosti izbora posameznih kategorij odgovorov.

Rezultati trdno podpirajo hipotezo o nižji stopnji upravljanja z vtisom med spletnimi anketiranci. Podobno velja tudi za drugo hipotezo: spletni anketiranci so manj verjetno izbirali skrajne odgovore na zgornjem in spodnjem delu lestvice kot anketiranci v telefonski ali osebni anketi. Posebej izrazita razlika se je pokazala pri vprašanjih, občutljivih na upravljanje z vtisom. Podatki pa kažejo le omejeno podporo za hipotezi, ki se nanašata na negativne učinke spletnega načina. Čeprav so spletni anketiranci pogosteje izbirali srednje vrednosti kot anketiranci v drugih dveh načinih anketiranja,

ostaja nejasno, koliko je to mogoče pripisati nižji točnosti odgovorov v spletnem načinu. Tudi preverjanje hipoteze o večji stopnji nediferenciacije odgovorov kaže mešane rezultate: spletni anketiranci so izkazovali nižjo stopnjo razlikovanja pri mnenjskih in vrednotnih postavkah, anketiranci v telefonski in osebni anketi pa na postavkah, občutljivih na upravljanje z vtisom.

Kljub ne povsem jasni sliki o konsistentnosti in velikosti učinkov načina anketiranja so rezultati pretežno skladni s predhodnimi raziskavami in obetavni za spletne ankete. Večina potencialno negativnih učinkov spletnega anketiranja je bila majhna. Ugotovitve tako nadalje potrjujejo spletne ankete kot ustrezen način anketiranja, ki omogoča zbiranje visokokakovostnih podatkov.

Izvirni prispevek in implikacije

Ključna dodana vrednost disertacije izhaja iz sistematične evalvacije dejavnikov učinkov načina anketiranja v spletnih anketah. Podana splošna elaboracija ponuja celosten pogled na problem, ki v obstoječi literaturi večinoma ni na voljo. Disertacija izpostavlja pomembnost obravnave učinkov načina anketiranja kot rezultata kompleksnega skupka različnih dejavnikov. Predlagani konceptualni model prikazuje številne vire učinkov, ki jih je treba upoštevati pri pripravi spletne ankete, da bi se izognili nepredvidenim vplivom na kakovost podatkov. Zavedanje o potencialnih pasteh je še posebej pomembno zaradi visoke fleksibilnosti spletnih anket, ki zagotavlja skoraj neomejene možnosti za izdelavo vprašalnika. Uporabniki anket morajo zato pazljivo pretehtati smiselnost izkoriščanja specifičnih značilnosti spletnega načina, zlasti ob uporabi metodološko nezadostno proučenih funkcionalnosti.

Razumevanje virov učinkov načina anketiranja postaja nadalje kritično z naraščajočim vključevanjem spletnih anket v kombinirane načine anketiranja, kjer je primerljivost rezultatov med posameznimi načini običajno bistvenega pomena. Izvedeno proučevanje učinkov uvaja konceptualno orodje za identifikacijo možnih virov razlik zaradi kombiniranja spletnega anketiranja z drugimi načini. To je prispevek k znanju o implementiranju primerljivejših vprašalnikov prek različnih načinov anketiranja.

Disertacija ponuja tudi smernice za nadaljnjo metodološko obravnavo problema. Celosten pregled literature kaže premalo raziskane teme, ki bi jim bilo treba nameniti dodatno raziskovalno pozornost. Prikazana prikrita narava učinkov načina anketiranja poziva raziskovalce k večsmernemu pristopu, ki vključuje pazljivo načrtovanje eksperimentalnih načrtov in uporabo različnih analitičnih tehnik za odkrivanje vpliva na anketne ocene. Za bistven napredek pri obvladovanju učinkov načina anketiranja pa je zlasti pomembno povečati uporabo metaanalitičnih študij ter spodbujati skupne, inovativne in strateško usmerjene raziskovalne projekte.

Ključne besede: učinek načina anketiranja, spletna anketa, način anketiranja, napaka merjenja, proces odgovarjanja na anketna vprašanja

Abstract

Rationale and objectives

Web surveys have already become one of the leading approaches to survey data collection. They offer significant advantages, including lower research costs, fast data collection, and broader utilization possibilities due to advanced questionnaire features. Their role is becoming even more important as cost-reduction pressures in all sectors urge researchers to seek the optimal balance between data quality and costs. Web surveys are thus often considered as a promising alternative to traditional survey modes, like face-to-face, telephone, and mail surveys.

The utilization of any survey mode needs to take into account its possible negative influences on data quality. The dissertation addresses the problem of mode effects in web surveys as one of potential threats to the accuracy of obtained estimates. In the most general sense, mode effects are a component of survey error that arise because a specific survey mode is used to collect data. We based our elaboration on the central thesis that mode effects in web surveys are the result of a broad set of factors related not only to the mode itself but also to specific survey implementations.

The main purpose of the dissertation is to contribute to the big picture to enable a more thorough understanding of mode effects in web surveys. To accomplish this, we pursued four central objectives: 1) to establish a more solid conceptual framework of mode effects in web surveys, 2) to offer a comprehensive analysis of the forms and sources of mode effects, 3) to empirically demonstrate the volatile nature of the problem, and to 4) draw implications for further research and survey practice.

Development of the conceptual framework

To establish a conceptual framework of mode effects, we began by defining basic concepts. The literature review shows large variations in understanding and inadequate definition of terms “mode” and “mode effects”, which is surprising due to their common use in survey terminology. To be able operate with these concepts clearly throughout the dissertation, we placed them into a broader framework of survey process and survey errors and determined their explicit operational definitions. We define mode as a set of data collection procedures that determine the basic principles of communication and information transmission between the respondent and the survey questionnaire. These basic principles are based on six inherent mode characteristics: the main question presentation (input) channel, the response (output) channel, interviewer involvement, closeness of interaction between interviewer and respondent, use of computer technology for data collection, and medium of information transmission. In addition, we identify

several implementation-specific and contextual characteristics which depend on the properties of a specific survey application.

Following the proposed definition of survey mode, we understand mode effects as all direct and indirect effects of inherent mode characteristics on the accuracy of obtained survey estimates. Because all survey error components can have strong interactions with each other, mode effects are not limited to a direct impact of inherent mode characteristics but are mediated through implementation-specific and contextual characteristics as well as other factors.

We narrowed the focus to the influences of web mode on measurement errors. Because this type of error is closely related to the respondent's answers, we further considered the relevant aspects of survey response theories. We devoted particular attention to the information processing model by Tourangeau and colleagues, and expanded it with applicable conceptualizations by other authors. These include some other information processing models, models of response process deviations, and models of interaction between the respondent and the questionnaire.

Elaboration of sources of mode effects in web surveys

The central part of the dissertation elaborates the potential causes of mode effects in web surveys and their impact on the accuracy of estimates. We based this on an extensive evaluation and the integration of existing empirical research. We took into account various comparative studies dealing with a broad range of factors of differences in estimates between web surveys and other modes. On this basis, we formed an extension to an earlier conceptual model of mode effects proposed by Tourangeau and colleagues. By establishing potential mediating relations between inherent, implementation-specific and contextual characteristics of web mode, we obtained an inclusive insight into the critical factors of mode effects.

Overall, our elaboration exposed self-administration as the most general source of mode effects in web surveys. Lack of interviewers can increase the burden of respondents and at the same time severely limit the possibilities of providing additional extrinsic motivation. This increases the likelihood for respondents to resort to shortcutting strategies, leading to a less thorough performance of the response process and ultimately resulting in lower answer accuracy. In addition to self-administration, some other potential sources of mode effects in web surveys include specific contextual information due to visual presentation of questions, negative influences of computer-administration and computerized interaction, and the technical and physical environment in which the surveying takes place.

However, our investigation showed substantial variations in findings and largely inconsistent conclusions across different studies. This strongly indicates the volatile nature of mode effects. While the proposed conceptual model describes the critical interactions between mode characteristics that may lead to the emergence of mode effects, their actual occurrence also depends on a large number of other survey-related and respondent-related factors.

Methodology of the empirical study

To demonstrate and further explore the nature of mode effects in web surveys, we analysed data from an experimental application of the *Generations and Gender Survey*. The analysis focused on a comparison of estimates between web, telephone (CATI), and face-to-face (CAPI) data collection on 89 scale items. The simultaneous consideration of a large number of variables is advantageous as it allows us to observe differences between modes more thoroughly.

We used the empirical study to verify four general hypotheses. Two of them state the absence of negative influences of web mode and two the presence of such influences. We expected web respondents to express lower impression management tendencies (H_1) and to be less likely to select extreme scale values (H_2), but also to more frequently select middle scale values (H_3) and be more likely to resort to non-differentiation as a form of satisficing (H_4).

In order to minimize confounding effects of non-coverage, we obtained the sample of respondents from a commercial online access panel. We analysed data using two main modelling techniques: ordinary least square (OLS) regressions and partial proportional odds modelling (GO-logit). This allowed us to detect various forms of between-mode differences as key indicators of the presence of mode effects. The results were interpreted by considering significance levels adjusted for multiple testing and estimates of effect sizes.

Results of the empirical study

The results of the empirical study offer a mixed support for our general hypotheses. A majority of analysed items exhibited comparably small effect sizes. Mean differences were significant with 22% of items in web–CATI comparisons, and 27% of items in web–CAPI comparisons. However, the analysis of response distributions revealed a substantially higher number of significant effects. This indicates that some mode effects do not reflect in the measures of central tendency, but may still significantly change the selection probabilities for some answer categories.

The results firmly confirm the hypothesis about lower impression management tendencies among web respondents. The same holds true for the second hypothesis: web respondents were less likely to select lower or upper extreme answers than CATI and CAPI respondents. The difference was especially pronounced for questions susceptible to impression management. On the other hand, we found only limited support for both hypotheses about negative effects of web mode. Although web respondents tended to select mid-point answers more frequently than respondents in both interviewer-administered modes, it remains unclear whether this can be attributed to a lower accuracy of answers in web mode. Finally, observed between-mode differences in non-differentiation were largely mixed. Web respondents differentiated somewhat less on opinion and value-related questions, while respondents in both interviewer-administered modes exhibited lower differentiation on items susceptible to impression management.

Despite inconclusive evidence about the consistency and size of mode effects, the results are largely in line with previous research findings and encouraging for web surveys. Even where we found potentially negative effects of web mode, effect sizes were generally small. This further strengthens the position of web surveys as a viable survey mode capable of providing high data quality.

Novel contribution and implications

The key added value of the dissertation lies in a systematic evaluation of factors of mode effects in web surveys. The provided general elaboration benefits from an integrative treatment of the problem, which is largely absent in the existing literature. The dissertation underlines the importance of treating mode effects as a result of a complex conglomerate of various factors. The proposed conceptual model exposes numerous sources of effects that should be considered when designing a web survey in order to prevent unforeseen impacts on data quality. Because web surveys are highly flexible and allow virtually endless possibilities of questionnaire construction, awareness of potential pitfalls is even more important. Survey practitioners should therefore carefully weigh benefits of exploiting specific characteristics of web mode, especially when using features with yet unclear methodological implications.

The understanding of different sources of mode effects is becoming even more critical with the increasing inclusion of web surveys into mixed-mode survey designs, where comparability between utilized modes is usually of a paramount importance. The conducted investigation of mode effects introduces a conceptual tool for identifying possible sources of differences due to combining web surveys with other modes. This contributes a tool for achieving more comparable implementation of questionnaires across different modes.

The dissertation also offers guidance for further methodological treatment of the problem. A thorough literature review clearly shows under-explored topics to which further research attention should be devoted. The demonstrated evasive nature of mode effects also urge researchers to approach the problem from various directions by relying on carefully planned experimental designs and considering various analytical techniques to uncover influences on survey estimates. However, for a breakthrough advancement in the treatment of mode effects it will be especially important to increase the utilization of meta-analyses and stimulate concerted, innovative, and strategically oriented research efforts.

Keywords: mode effect, web survey, survey mode, measurement error, survey response process

Contents

Introduction

Web surveys and trends in survey methodology

Surveys present the central data collection tool for many academic, governmental, and commercial research organizations. Application of a standardized questionnaire to a sample from the target population allows a large-scale estimation of population characteristics. Web surveys are the most highlighting recent development in survey methodology. Highly flexible questionnaire design, access to the target population beyond time and space boundaries, fast data collection, and low implementation costs made them one of the most widely used survey modes. Their adoption was particularly fast in the commercial sector (Comley 2002), while governmental and academic organizations remain more reserved due to two major disadvantages of the web mode: the lacking coverage of the general population with Internet access (Farrell and Petersen 2010; Mohorko et al. 2013b) and lower response rates compared to traditional modes (Lozar Manfreda et al. 2008).

However, cost reduction pressures urge researchers from all sectors to search for the most optimal balance between data quality and costs. Principles of survey quality always operate within the boundaries of available resources. Furthermore, the concept of survey quality was extended early on, although mostly implicitly, from the *accuracy of estimates* to a range of additional quality-related criteria. In his early text on survey errors, W. Edwards Deming exposed the compromised usefulness of a survey if it lacks timeliness, that is by "changes that take place in the universe before tabulations are available" (Deming 1944, 360). The principles of quality assurance like *Total Quality Management* (also importantly contributed by Deming) and *Continuous Quality Improvement* (Biemer and Caspar 1994) started to become an increasingly important guide to what is now commonly known as the *Total Survey Quality* approach. Following this principle, Biemer and Lyberg (2003) list several approaches to the evaluation of overall

survey quality. Eurostat's quality dimensions, for example, require accuracy to be balanced with six other quality dimensions: relevance of concepts, timeliness and punctuality of dissemination, accessibility and clarity of information, comparability, coherence, and completeness (Biemer and Lyberg 2003).

In the 1990s, survey research entered "turbulent times", marked by aggravating problems of low response rates and required cost optimization on one side, and rapid developments of new survey modes on the other (Dillman et al. 2008, 6). Both aspects are strongly related. The pursue of balancing survey quality and costs has traditionally been a driving force of continuous innovation in the field. Developments in information communication technologies opened up a range of new opportunities, and web surveys are currently regarded as one of the most promising of them. However, each survey approach has its own specifics, requiring careful examination by survey methodologists in order to prevent an unforeseen impact on various aspects of data quality and costs.

Mode-related differences in survey data

The problem of differences in obtained answers between various modes was acknowledged already by Deming (1944). The issue, however, did not receive thorough and general attention over subsequent decades. Researchers were predominantly focused on sampling and nonresponse problems of survey research (Platek and Särndal 2001). They were devoting relatively little attention to the treatment of measurement errors in general (Alwin 2007) and even less to those arising from the use of a specific survey mode. A prominent exception is research on differences in reporting sensitive behaviours between self-administered and interviewer-administered modes: mail surveys were consistently reported to elicit higher reporting of such behaviours than telephone or face-to-face surveys (Hochstim 1967; Bradburn et al. 1978).

Results from research on sensitive questions between different survey modes increased awareness that characteristics of a particular mode can contribute to measurement errors. Methodological research started paying more attention to systematic between-

mode comparisons of response effects (Bishop et al. 1988), contextual influences (Schwarz et al. 1991), various data quality indicators (de Leeuw 1992), differences in response process (Dillman 1991), and many other aspects. Explanations for the observed differences in estimates were attributed to various mode characteristics, like the presence or absence of interviewers, properties of the communication and information transmission, and the use of computer technology for data collection (de Leeuw 1992; Tourangeau et al. 2000).

These influences of survey mode on obtained estimates were labelled with the somewhat vaguely defined term *mode effects*. The term seems to have become commonly used in survey methodology in the early 1990s (for example Aquilino and Lo Sciuto 1990; Dillman and Tarnai 1991), generally referring to observed differences in estimates obtained using various modes. Mode effects are generally considered as a special type of measurement error, arising because a specific survey mode is used to collect data (Groves 2004). Some authors, however, understand them in a broader sense that covers not only measurement errors, but also some or all other survey errors related to the mode (e.g. sampling frame error, nonresponse error, and data processing errors). In the dissertation we used and elaborated the former aspect of mode effects.

The interest in mode effects became especially predominant with the increased use of mixed-mode surveys. Appropriate and carefully planned combinations of various modes within a single survey project can be an effective way of compensating for weaknesses of individual modes and assuring higher optimization of costs and data quality (de Leeuw 2005). Even large longitudinal cross-national surveys of high scientific reputation, like the *European Social Survey*, are increasingly considering a transition to mixed-mode data collection (P. Martin 2011). Mixed-mode approaches are particularly attractive with web surveys due to their cost-reduction potential. They can be used as an inexpensive mode for surveying parts of the target population with an Internet access, while more expensive modes, like telephone and face-to-face interviewing, are then reserved for reaching non-respondents to the web mode. However, mixed-mode data collection raises concerns about comparability of data obtained using different modes.

The research on mode effects in web surveys is predominantly based on experimental comparisons of web surveys to other modes. A majority of these studies focus on specific aspects of data quality, like the level of reporting on sensitive topics, deviations from the optimal question processing, length of answers to open-ended questions, and so on. The main problem of scientific research on mode effects in web surveys (and other modes) is the lacking theoretical elaboration of the problem. Because mode effects are not precisely defined, empirical comparisons often fail to establish sufficient and comprehensive relations between findings and theoretical foundations. It is therefore often unclear how observed differences between modes map to mode effects. This significantly limits the possibilities of *reducing* rather than just *observing* negative influences of a web survey mode on the accuracy of obtained data.

Objectives of the dissertation

The central goal of the dissertation is to offer a comprehensive elaboration of mode effects in web surveys of individuals from theoretical and empirical perspective. This will contribute to the better understanding of the mechanisms of survey errors that occur because a web mode is used for data collection.

The dissertation builds on the following central thesis:

> *Specific characteristics of the web mode influence the respondent's process of answering survey questions and potentially lead to mode effects. Whether or not, in what form, and to what extent these effects occur depends on a broad set of other factors related to specific implementations of a particular survey.*

This thesis guides our exploration of mode effects in web surveys in several important ways. First, it strictly distinguishes mode effects from between-mode differences in survey estimates. The reason for the occurrence of mode effects lies in the characteristics of the mode itself and not in the use of multiple modes. Although this reasoning is in

line with the common conceptualization of mode effects found in literature, its explicit application to research is often limited. The second implication of the thesis is that mode effects arise by affecting the response process through which the respondent derives answers to survey questions. Understanding the respondent's cognitive processes is therefore crucial to explain patterns of mode effects. Finally, the thesis denies a deterministic role of mode characteristics in emergence of mode effects. Inconsistent findings of studies on the presence and magnitude of mode effects indicate the existence of various moderating and mediating factors on which the emergence of mode effects is conditioned. One of the central contributions of the dissertation is a review of such factors in web surveys.

To accomplish the stated goal, we pursue the following main objectives:

1. **Establish a solid conceptual framework for understanding mode effects in web surveys and in other modes.** A substantial part of the dissertation is devoted to discussions and definitions of intuitively well understood, but formally vaguely defined basic concepts of *survey mode* and *mode effects*. We believe coherent definitions to be crucial for the adequate conceptual understanding of the problem and for the appropriate mapping of various sources and types of survey errors to mode effects. This also allows us to position mode effects into a broader context of survey errors.
2. **Provide a comprehensive and integrative review of forms and sources of mode effects in web surveys**. The main intended theoretical contribution of the dissertation is the systematic integration of existing empirical research on measurement errors in web surveys into the conceptual framework of mode effects. We paid special attention to relations between the characteristics of web mode and specific factors of survey implementation, influences of these relations on the survey response process, and their consequential role in the emergence of mode effects.
3. **Empirically demonstrate the complex relations between the different factors that contribute to the emergence of mode effects in web surveys.** To illustrate

the importance of relations between mode characteristics and other factors for the occurrence of mode-specific measurement errors, we analysed and compared effects on scale questions in an experimental application of the *Generations and Gender Survey* using web, telephone, and face-to-face modes. We focused on four common mechanisms of differences in answers to scale questions: impression management, extreme and midpoint answering, and non-differentiation between response categories.

4. **Expose implications for further research on mode effects and the importance of the problem for survey practice.** Based on all preceding objectives, we identified the limitations of current research on mode effects in web surveys and proposed some future research directions for the better explanation and treatment of the problem. We also summarized the key points of interest for survey practitioners who may especially benefit from understanding practices of web survey implementations that can increase the risk of mode effects.

In some ways, this dissertation may open more problems than it solves. It primarily focuses on conceptual issues of mode effects and less on the empirical investigation of the problem. However, its comprehensive elaboration of theoretical background and wide range of empirical studies helps foster understanding of potential causes of mode effects in web surveys, and establishes the framework for the more successful tackling of the problem in the future.

Structure of the dissertation

Our exploration of mode effects in web surveys begins by discussing the essential problems of defining mode effects and their placement into a broader context of survey errors (Chapter 1). We start with the clarification of basic terminology and the establishment of the required theoretical background. Firstly, we discuss the definition of *survey mode*. While the concept is intuitively well-understood, several important problems arise at the operational level, largely due to recent trends of the proliferation of survey

modes, their increasing complexity, and the rise of mixed-mode approaches (Couper 2011). Using a review of approaches proposed by various authors, we established an explicit separation between characteristics that define a survey mode and those depending on specific implementations and applications of survey data collection. In section 1.2, we make an overview of types of survey errors and the *Total Survey Error* paradigm (Anderson et al. 1979) which offers a well-established framework for a conceptual understanding of possible error sources in surveys. Since we primarily focus on the influences of mode effects on measurement errors, we devote some more attention to the relevant relations between concepts used by survey methodology and psychometrics. In the final part of the chapter, we use the established concepts to define mode effects. Rather than giving "the ultimate definition" of mode effects, the key purpose of this is to focus our further treatment of the problem concisely. We expose variations in the understanding of mode effects in the literature and draw an explicit separation between mode effects and between-mode measurement differences.

In Chapter 2, our focus turns to the survey response process, which can be affected by specific characteristics of the survey mode. We first review the general models of cognitive processes conducted by respondents when answering survey questions. A majority of this part is based on a comprehensive work by Tourangeau et al. (2000), but we also consider some alternative and complementary models. The second part of the chapter deals with deviations in the response process due to suboptimal response strategies (the satisficing principle by Krosnick and Alwin 1987) and distorted reporting caused by sensitive questions. Finally, we consider some more specific models to describe differences in the response process between interviewer-administered and self-administered modes.

These discussions and conceptualizations are used for the elaboration of mode effects in web surveys in Chapter 3. We systematically go through the characteristics of the web mode and comprehensively review studies comparing web to other survey modes as well as within-mode experimental manipulations to highlight possible sources and consequences of mode effects. At each stage of discussion, we show how different

implementations of a specific survey, characteristics of respondents, and other factors enter into the relationship with essential properties of the web mode and potentially cause mode-specific influences on measurement errors. Some common deviations in data quality resulting from these influences are summarized in the section 3.4. The chapter concludes with a synthesis of complex interrelations between causes of mode effects in web surveys to offer a conceptual model for the identification of critical factors that can produce mode-specific measurement errors.

Chapter 4 empirically demonstrates this complex role of contributing factors in the emergence of mode effects. The empirical analysis is not intended nor able to verify all aspects discussed in the preceding chapter empirically. It rather focuses on a subset of specific factors related to scale questions by observation of differences between web, telephone, and face-to-face modes. The main benefit of the experimental study is its ability to consider a large number of questions and items simultaneously, which is rarely done in empirical evaluations of mode effects. We observed different indicators of data quality, including substantive differences in questions, reporting of sensitive behaviours, and satisficing strategies.

In the final part of the dissertation we integrate all obtained theoretical and empirical findings into an overall picture of mode effects in web surveys. We emphasize the most important implications of derived conclusions for practical applications of web surveys, comparable mode implementations, and inclusions of the web mode into mixed-mode survey designs. Furthermore, we propose some approaches for the more effective future treatment of the problem.

Chapter 1
Defining mode effects and their context

Data collection based on standardized questionnaires is common in many social science disciplines, including sociology, psychology, economics, education, and communication studies. Survey methodology develops tools for all these disciplines and tries to offer solutions for a variety of their needs. Concern with the problem of data quality is common to all fields utilizing survey data collection; however, the consideration of error sources and terminology vary significantly among them.

The discussion by Groves (2004) offers an informative insight into the "language differences" in describing survey errors in (social) statistics, psychology and economics. The problem is not entirely different from the ancient story of the Tower of Babel depicting the confusion of tongues. According to Groves (2004, 7) "we thus face the unpleasant task of language lessons, reviewing words and their meanings". He adds that this confusion is unlikely to be due to substantial differences in the nature of errors between the disciplines, but rather due to variations in the perceived importance of different error factors.

Another set of problems arise from the insufficient definitions of key terms in the field of survey methodology itself. For example, both *survey mode* and *mode effects* are differently understood and thus inconsistently used across studies. More importantly, exact explanations about what these terms refer to are rarely offered by individual studies.

Therefore, it is important to begin our research on mode effects in web surveys by reviewing and defining the key concepts used to place the studied problem into the context of survey errors. Our aim is not to propose an ultimate definition of any of these concepts, but to establish the terminological framework which we deal with throughout

the dissertation. The relatively lengthy underlying discussions and justifications of these stipulative definitions are crucial for the further theoretical and empirical treatment of causes of differences between survey modes. In addition, they express the importance and current lack of more precise statements about commonly used terms that appear in the survey literature.

1.1 Survey mode

Although the interchangeable terms *mode of administration*, *mode of data collection* or simply *(survey) mode* are routinely used in the survey methodology, they remain surprisingly vaguely understood. They are commonly used as rather simple and intuitive descriptors of the approach used to collect data from respondents, e.g. mail, telephone, face-to-face, or web survey. Little effort has been made beyond this to provide a systematic answer to the question of what constitutes the mode and differentiate one mode from another (Couper 2011).

This intuitive understanding of survey modes is sufficient for a vast majority of survey applications and even mode comparison studies but becomes lacking when one aims to study the data quality implications of mode characteristics, and especially when integrating such findings across various studies. Surveys are by definition highly standardized and systematic data collection methods. The vast body of general survey literature shows that sometimes even very small procedural changes can lead to significantly different results (e.g. Biemer and Lyberg 2003; Dillman et al. 2008; Groves et al. 2009). In order to discuss which of these differences are caused by the mode itself and which by other factors (like sampling, nonresponse or questionnaire characteristics), it is necessary to explicitly state the defining components of a survey mode[1].

The vague understanding of the term has become especially problematic with the increasing complexity of survey data collection approaches over the last decades. This is

[1] Before formally establishing the notion of survey mode as used throughout the dissertation, we use the term in line with rather heuristic understanding, which is common in the survey literature.

illustrated by Table 1.1, which lists some modes commonly found in the literature, although the naming may vary somewhat between different authors.

Table 1.1: Examples of survey modes frequently encountered in the literature reviewed

paper-and-pencil face-to-face interviewing (PAPI)	The interviewer visits the respondent (face-to-face) in the field, reads the questions from the paper questionnaire and writes down the respondent's answers.
paper-and-pencil telephone interviewing	The interviewer calls the respondent by the telephone, reads the questions from the paper questionnaire and writes down the respondent's answers.
computer-assisted personal interviewing (CAPI)	Face-to-face interviewing in which the interviewer reads the questions from the computer screen and enters the respondent's answers into the computer.
computer-assisted telephone interviewing (CATI)	Telephone interviewing where the interviewer reads questions from their computer display and enters the respondent's answers into the computer.
interactive voice response (IVR) **touch-tone data entry (TDE)** **voice recognition entry (VRE)**	The respondent uses the telephone to listen to the recorded questions and provides answers using the telephone keypad (touch-tone data entry – IVR/TDE) or orally. In the latter case, voice recognition (VRE) technology can be used to record answers into the database (IVR/VRE).
CASI **audio-CASI (ACASI)** **video-CASI (VCASI)**	A range of methods in which the respondent answers questions by themselves using the interviewer's computer in a face-to-face situation. The questions can be presented textually, or using audio (ACASI) or video (VCASI).
mail survey	The paper questionnaire is sent by mail, answered by the respondent themselves and returned back to the research organization by mail.
fax survey	The paper questionnaire is answered by the respondent themselves and transferred back to the research organization using a fax machine.
disk-by-mail (DBM) survey	The respondent uses their computer to answer the electronic questionnaire on a floppy disk or another data storage device and sends the device back to the research organization by mail. This mode was largely replaced by web surveys.
e-mail survey	The questionnaire is sent to respondents by e-mail as a part of an e-mail message. Depending on the technology used, respondents provide answers by editing text of the message or by filling-in an integrated electronic form.
web survey	The questionnaire is displayed in the web browser of the respondent's computer or other Internet-enabled device (tablet, mobile phone). Data are transmitted to and from the respondent using the Internet.
virtual interviewer (VI) web survey	The questions in a web questionnaire are presented by a pre-recorded video of live interviewer or animated character. Respondents provide their answers similarly as in ordinary web survey.
SMS survey	Questions are presented to the respondent via a Short Message Service (SMS) on a mobile phone. Respondents enter their answers in the reply message.

According to Couper (2011), this complexity of modes arose from developments related to:

1. **The increased number of different modes**. At the beginning of survey research, face-to-face and mail surveys were the only widely used data collection modes (Groves et al. 2009). The widespread availability of telephones gave rise to telephone surveys in the 1960s. It presented the milestone of introducing information-communication technologies (ICTs) into survey data collection which then gained further momentum with developments in computer technology and the Internet during the 1980s and 1990s. This resulted in a whole range of new modes. Some of them emerged as a direct evolution of existing approaches: telephone surveys evolved to computer-assisted telephone interviewing (CATI) and interactive voice response surveys (IVR), and paper-based face-to-face surveys moved to computer-assisted personal interviewing (CAPI) and computer-assisted self-interviewing (CASI). Others, like e-mail and later web surveys, appeared as computerized versions of mail surveys, but with the utilization of numerous potentials of the Internet as a new data collection medium. The complexity of this variety of survey modes is increased even further by recent technological trends related to data collection on mobile devices (Couper 2005; Couper 2011).
2. **The increased complexity within modes.** Introduction of new technologies into survey research further blurred the border separating mode variations from completely different modes. For example, highly interactive web questionnaires can in some respects assume the role of the interviewer, video-CASI can now be administered either online or face-to-face, and mobile devices can collect location-based information in addition to survey responses (Couper 2005; Couper 2011).
3. **The rise of mixed-mode surveys.** It is becoming increasingly common that several modes are combined within a single survey project (Biemer and Lyberg 2003) in order to overcome the limitations of individual modes and reduce research costs (de Leeuw 2005). Some examples of mixed-modes during data

collection include mail invitation to web surveys, face-to-face follow-ups for web non-respondents, CASI for sensitive parts of an interviewer-administered questionnaire, and different sampling procedures for different parts of the target population. Mixed-mode approaches are now also an integral part of guidelines for conducting surveys, like Dillman's *Tailored Design Method* (Dillman et al. 2008).

1.1.1 Mode and system of data collection

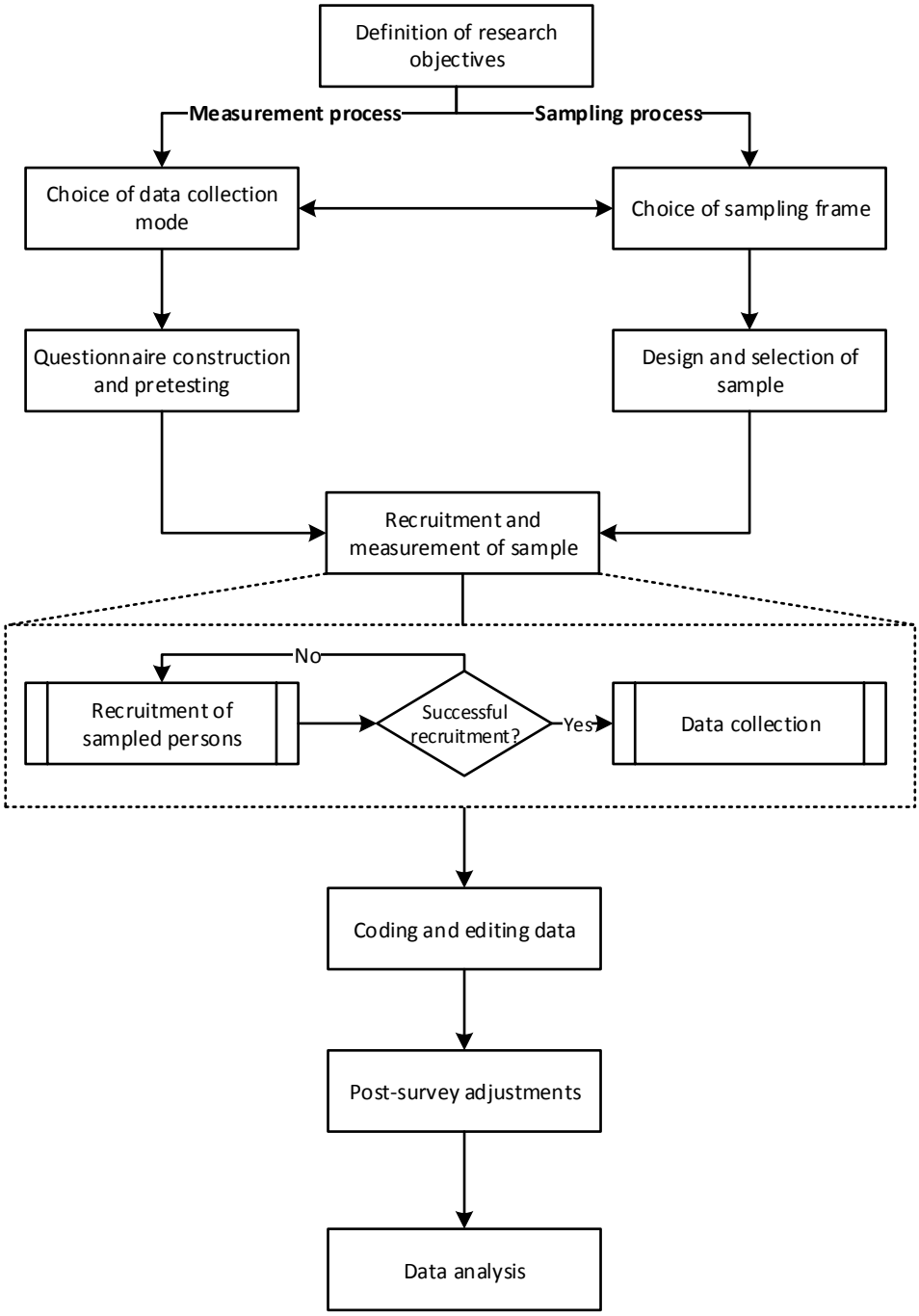
In order to offer a more explicit conceptualization of mode, it is first necessary to place the mode of data collection into a broader survey process, which consist of several phases carried out during the survey preparation and implementation. The phases are schematically presented in Figure 1.1 (adapted from Groves et al. 2009).

Biemer and Lyberg (2003) introduce the notion of *data collection system* to refer to all phases of the survey process and their implementation components (not shown in Figure 1.1). The system refers to all activities related to the operational implementation of survey data collection (Biemer and Lyberg 2003, 208):

> *A system might include design factors such as interviewer hiring, interviewer training, interviewer supervision, questionnaire contents, number of callback attempts, refusal conversion strategies, sampling system, and sampling frame coverage. /…/ In general, data collection systems do not consist of one mode only, since mixed-mode surveys are norm these days*.

The distinction of mode of data collection from the phases of system of data collection requires some important consideration. Firstly, the selected mode of data collection importantly constrains the implementation options of other phases of the survey process (Biemer and Lyberg 2003), suggesting a strong interdependence between the phases. If one decides, for example, to use a web survey for data collection, the questionnaire needs to be constructed and designed for web administration, sampling procedures need to be implemented to appropriately include those with Internet access, and the

Figure 1.1: Phases of the survey process with recruitment and measurement divided into separate steps, with a more detailed overview of the recruitment and measurement phase



Note: Adapted from Groves et al. (2009). Dashed area adapted from Vehovar and Batagelj (1996) and added to the original figure.

corresponding recruitment procedures need to be employed. The implementation of post-fielding phases (editing, adjusting and analysis) are also mode-specific to at least some degree. On the other hand, the research objectives and the target population on

which the sampling process depends also determine the mode itself. Continuing with the example above, it is not feasible to choose a web survey for measuring characteristics of the target population with low Internet coverage.

Secondly, to place the mode of data collection into the context of a data collection system, the phases of recruitment and measurement (data collection) need to be conceptually understood as two separate and to some degree independent steps (Vehovar and Batagelj 1996; Vehovar and Lozar Manfreda 2008). This is presented as the dashed area added to the figure above. For some modes, this separation is less obvious than for others. In telephone surveys, the recruitment is usually performed at the beginning of the telephone conversation. Similarly, the invitation is commonly sent in the same envelope as the survey questionnaire for a mail survey. In these cases, the mode of data collection is equal to the mode of recruitment and both are performed at the same time. However, increasingly often, the recruitment is made in a mode different from data collection, introducing a specific type of mixed-mode survey system (de Leeuw 2005). Mail invitation to web survey (e.g. Porter and Whitcomb 2007), telephone screening or recruitment for subsequent web surveying (e.g. Deutschmann and Faulbaum 2001), and mail pre-notification for face-to-face interview (e.g. Groves and Couper 1998) are only some examples of such designs.

Finally, as pointed out above, mixing modes during the data collection phase itself is becoming increasingly common. The data collection phase in Figure 1.1 therefore does not necessarily consist of one single mode but includes the set of all data collection modes used in a particular survey project.

Survey mode should thus be understood as a characteristic of data collection and thus distinguished from sampling procedures, mode of recruitment (solicitation), and other phases of the survey process. Yet, there is a strong interdependence between these phases. As we discuss in section 1.2.3, the procedures utilized in one phase can have a direct influence on the quality of implementation of another phase.

1.1.2 Dimensions of mode

Authors in survey methodology use various approaches to describe characteristics or dimensions of data collection procedures that distinguish one mode from another. To reduce the complexity of separating one mode from another, it is necessary to systematically disentangle the characteristics or dimensions of data collection procedures that are commonly regarded as different modes.

Couper (2011; similiar also Groves et al. 2009) discusses six dimensions of modes: the **degree of interviewer's involvement** (fully interviewer-administered, self-administered in presence of interviewer, administered in group of respondents, only delivered or supported by the interviewer and fully self-administered), the **degree of contact with the respondent** (more or less direct, indirect, or picture or video of the interviewer), the **channel of communication** (aural, visual, or both), **locus of control** (primarily respondent, primarily interviewer, software control), the **degree of privacy** (as the presence of other people during the interview), and the **degree of computer technology used** (paper-based questionnaires, technology used by the interviewer, organization-provided technology used by the respondent, or own technology used by the respondent).

De Leeuw (1992; 2005) describes the differentiating characteristics of modes by considering three groups of factors: **media-related factors** (medium used, familiarity with the medium and its use, locus of control, pace, and sincerity of purpose), **information transmission factors** (cognitive stimulus, communication channels, and the temporal order of presentation of questions), and the **interviewer impact.** The most obvious difference compared to the Couper's dimensions is the inclusion of socio-cultural factors (familiarity and use of the medium, sincerity of purpose and the respondent's perceptions of the appropriate pace of conversation).

Another approach to describing the dimensions of modes is provided by Tourangeau et al. (2000). According to them, modes of data collection differ in the **method of contacting the respondent and delivering the questionnaire** (telephone, mail, Internet/e-mail,

or in-person), the **medium of questionnaire presentation** (paper or electronic), the **person administering the questionnaire** (interviewer or respondent), the **channel of presentation** (aural or visual), and the **mode of responding** (oral, written, or electronic). An important addition to the other two classifications is the explicit separation of communication channels in a direction to the respondent (called channel of presentation) and from the respondent (mode of responding). The authors also list some other characteristics that are affected by these basic dimensions of modes: **impersonality**, **legitimacy**, **cognitive requirements, pace of the interview, question processing order**, and **mental models** (derived interpretations of the survey task from other situations)**.**

The final classification of modes we refer to is a three-dimensional classification presented by Biemer and Lyberg (2003). They classify survey modes according to the **data collector involvement** (low or high), the **degree of contact with the respondent** (direct, indirect or no contact), and the **computer assistance used** (paper or computer). In contrast to other approaches presented here, the "no contact" modes consist of the data collection methods beyond a common definition of survey research (e.g. direct observations, administrative records and electronic data interchange). In general, this reduction of differentiating characteristics to only three dimensions does not enable to uniquely distinguish between common survey modes (Biemer and Lyberg 2003, 189).

It is clear that different authors include somewhat different sets of dimensions in describing mode characteristics. There are also some differences in the naming of these dimensions. For example, Couper (2011) and Tourangeau et al. (2000) name the distinction between visual and aural question presentation as *channel of communication*. De Leeuw (1992; 2005), on the other hand, describes this distinction as *cognitive stimulus* and reserves the term *channel of communication* for describing the availability of verbal, non-verbal and paralinguistic communication.

Regardless of these differences, there is a considerable overlap between these authors. Putting all together, they address three broad groups of characteristics related to survey modes: *inherent characteristics*, *implementation-specific characteristics* and *contextual characteristics* of survey modes. We treat inherent mode characteristics as the basic

frame of the data collection approach. A selected combination of these characteristics is stable (do not change) across actual survey implementations, respondents, or situations. For example, aural question presentation and oral response are inherent mode characteristics of the CATI mode and do not depend on different ways in which the telephone interviewing can be implemented for a specific survey project.

Implementation-specific and contextual mode characteristics are more fluid. They are simultaneously influenced by combinations of different inherent characteristics and a broad set of other factors, related to the specific implementation of the survey and elements of the context in which surveying takes place. The contextual elements include interviewers, respondents, and properties of the environment of surveying situation.

Inherent mode characteristics

Based on the review of the four classifications of mode dimensions above, we regard the following dimensions as being inherent to every survey mode:

- **Information transmission medium** is the tool (or service) used to transfer the survey questionnaire and answers (data) between the respondent and the researcher. This can be done in person, via telephone line, mail or other form of physical delivery, e-mail, or web.
- **Main question presentation** (input) **channel** is used to present the questions to the respondent and can be auditory, visual, or both. It refers to the type of stimulus used to convey the questions in the direction to the respondent.
- **Response** (output) **channel** is used by the respondent to convey their answers, either orally or in a written form. Following Tourangeau et al. (2000), we also added the electronic response provision as a separate type of response channel. Electronic response can be done by selecting the response option by mouse, touchpad, touchscreen, keying-in the answer using the computer keyboard or telephone keypad etc. Note that the output channel refers to the way the respondent provides the answer and not how the answer is entered to the

questionnaire. This distinction is only important in interviewer-administered surveys, like CAPI, where the respondent provides answers orally and interviewer enters them electronically.

- **Interviewer involvement during data collection** refers to the presence (or absence) and the role of the interviewer in the data collection process. In interviewer-administered modes, the interviewer administers the questionnaire completely. In the group of modes commonly named *self-interviewing* the interviewer is merely present during the data collection, while the respondent completes the questionnaire by themselves[2]. Finally, in self-administered surveys no interviewer is involved in the data collection process at all.
- **Closeness of interaction with the respondent** is related to the interviewer's involvement, but focuses on the type of interaction between the respondent and the interviewer. The interviewer may interact with the respondent in person (face-to-face) or remotely (e.g. by the telephone or, though rarely, by video call). A special case is the use of a virtual interviewer in a form of audio or video recording of live or animated person. While no live interviewer is involved in the data collection here, the respondent may still experience such a situation as more interview-like than self-administered. Virtual interactions are common in IVR surveys where the respondent listens to a pre-recorded interviewer's voice. Video recordings of interviewers are sometimes used in VCASI (e.g. Krysan and Couper 2003) or on the web (e.g. Krysan and Couper 2006).
- **Computer technology for data collection** refers to whether any kind of computer technology is used at the time of data collection either by the interviewer or by the respondent[3].

The conceptual mapping of these characteristics to dimensions discussed by other authors is provided in Table 1.2.

[2] Couper (2011) mentions *group-administration* and *instrument delivery* as two additional roles of the interviewer. We regard the former as a special type of self-administered survey in the presence of interviewer (*self-interviewing*). We left out the latter role to separate data collection from solicitation.

[3] Here we do not make a further distinction (as does Couper 2011) whether the computer technology is interviewer-provided (like in CASI) or respondent's own (like in web surveys).

Table 1.2: Comparison between mode characteristics discussed in the dissertation and in the literature[a)]

	Couper (2011)	de Leeuw (1992; 2005)	Tourangeau et al. (2000)	Biemer and Lyberg (2003)
Inherent characteristics				
Response transmission medium	Medium of communication[b)]	Medium of communication[b)]	Method of contact[b)]	
Interviewer involvement	Interviewer involvement	Interviewer's presence	Method of administering the survey	Data collector involvement
Closeness of interaction	Contact with the respondent			Contact with the respondent
Question presentation channel	Channel of communication to the respondent	Cognitive stimulus	Channel of conveying the questions	
Response channel	Channel of communication from the respondent		Mode of responding	
Computer technology for data collection	Computer technology		Medium of question presentation	Degree of computer assistance
Implementation-specific characteristics	Locus of control	Temporal order of questions presentation Communication channel[c)] Locus of control Regulation of the communication flow	Sense of impersonality Order of question processing Cognitive requirements	
Contextual characteristics	Degree of privacy	Familiarity with the medium Use of medium Conveyed sincerity of purpose Social conventions regarding pace of communication	Mental models Conferred legitimacy Pace of the interview	

Notes:
a) Detailed definition of generally comparable dimensions may vary between different authors.
b) Does not explicitly distinguish the questionnaire delivery medium from the response delivery medium.
c) De Leeuw (1992; 2005) refers to the communication channel as the availability of verbal communication, non-verbal communication, and para-linguistic communication.

Implementation-specific characteristics

The implementation-specific characteristics depend on a specific way of mode implementation in a survey project. Relevant characteristics identified by the literature review above include locus of control, temporal order of question presentation, availability of verbal, nonverbal and paralinguistic communication channels, sense of impersonality, pace of the interview, cognitive requirements, and some others (Table 1.2).

Different combinations of inherent mode characteristics constrain and sometimes determine the implementation-specific characteristics. For example, the level of available verbal, nonverbal and paralinguistic channels, as distinguished by de Leeuw (1992), is constrained by the closeness of interaction with the respondent, the response transmission medium, and the question presentation channel used. Cognitive requirements like listening skills, numeracy, literacy, and the ability to follow routing instructions are determined by the combination of question presentation, response channels, the role of the interviewer, and the use of computer technology for data collection (Tourangeau et al. 2000).

Several implementation-specific characteristics depend on the way the mode is implemented in a specific survey project. Mail or web questionnaires can convey additional meanings to the text through graphical elements, depending whether or not such elements are used (de Leeuw 2005). Further, as Couper (2011) notes, a web questionnaire can be programmed completely statically, leaving the whole locus of control to the respondent. Alternatively, the inclusion of certain dynamic and interactive features into the questionnaire can, to certain degree, limit the respondent's locus of control, for example, by preventing free movement through the questionnaire, prompting for skipped answers, probing the respondent in case of too fast responding, and so on.

Implementation-specific characteristics are usually not rigid even within a specific survey implementation and may be influenced by individual interviewer or respondent. Continuing with the locus of control example, some respondents may try to exert more control over the temporal order of question presentation in the interview situation than others by asking the interviewer to go back and forth in the questionnaire in order to

revise their answers. Individual behaviour may also mediate other implementation-specific characteristics, like the level of available verbal, nonverbal and paralinguistic channels.

The implementation-specific characteristics thus depend not only on the mode itself, but may vary across the interviewers, respondents, and specific procedures used in the data collection phase.

Contextual characteristics

The contextual mode characteristics largely depend on specific social and individual contexts in which the mode is implemented, i.e. in which the surveying takes place. Factors like familiarity and use of the medium, sincerity of purpose, the respondent's perceptions of the appropriate pace of conversation (de Leeuw 1992; de Leeuw 2005), and the degree of privacy (Couper 2011) are only some examples of such characteristics. The inherent mode characteristics constrain the range of available surveying contexts; however, a considerable variation within these constraints can still be expected. For example, the degree of privacy in the face-to-face interview is inherently limited by the physical presence of the interviewer. Yet, the actual degree may still vary according to the context in which surveying of a specific respondent takes place (e.g. some respondents may be interviewed with other household members present and others alone).

The context does not refer only to the characteristics of the physical environment in which surveying takes place. Factors like social norms and values, respondent's characteristics and abilities and several other specifics of data collection situation may all affect the actual implementation of the data collection procedures. Examples of contextual characteristics from the literature review above include familiarity with the medium, conveyed sincerity of purpose, use of medium (de Leeuw 1992), conferred legitimacy, and mental models used by the respondent. How the survey mode is perceived regarding these aspects strongly depends on both social and individual factors.

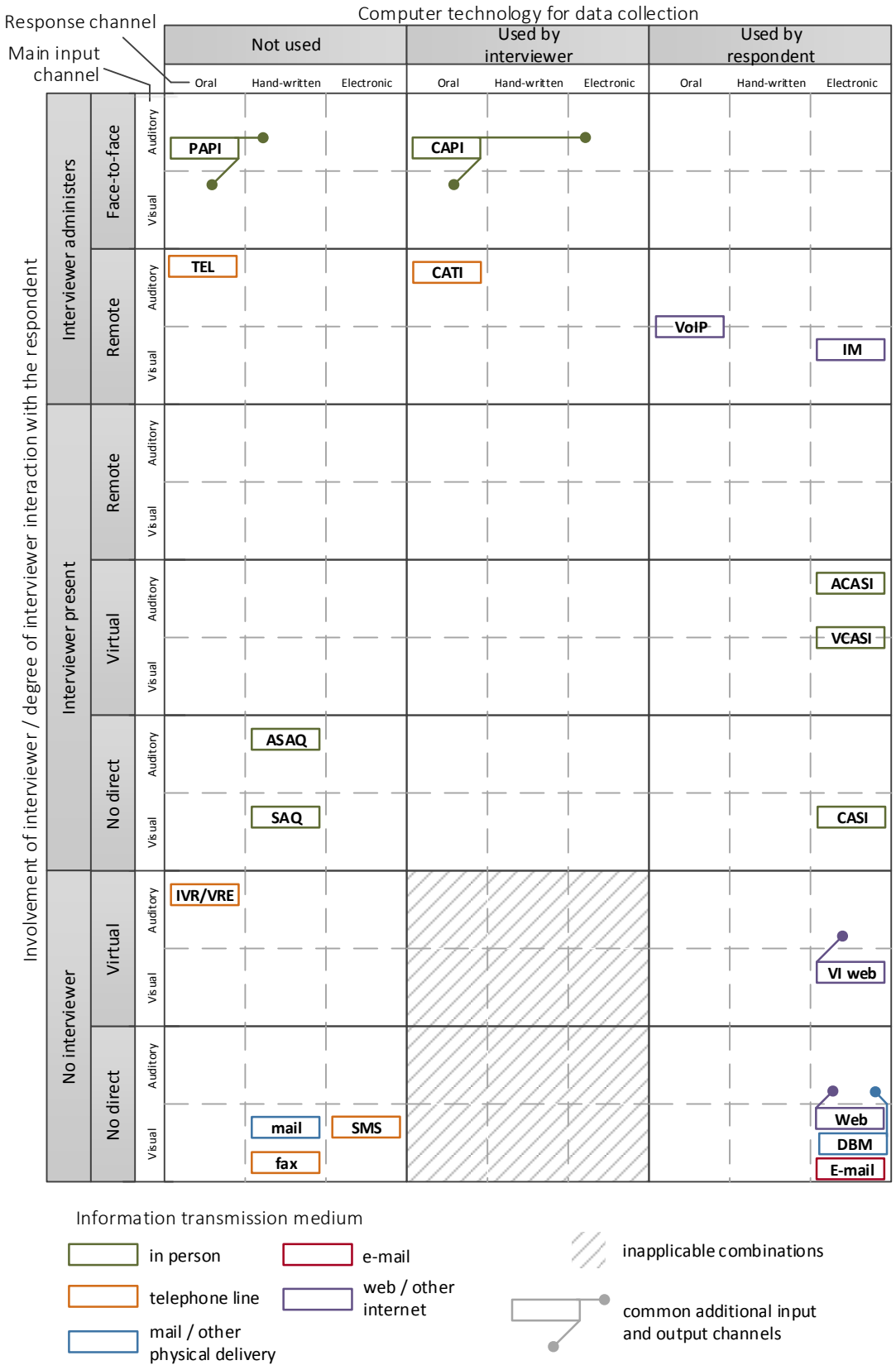
Another common contextual variation during survey application is the technology used by the respondent. Different devices and applications used by respondents in computerized self-administered questionnaires (CSAQ) can introduce differences in experience even if the same input and response channels are used. This is especially highlighted in web surveys where respondents may use different Internet-enabled devices, screen resolutions, and input methods (like a mouse, touchpad, touchscreen). All these may importantly influence the response process and the collected data (Couper 2008; Callegaro 2013).

1.1.3 Definition of mode and its problems

The following key points summarize our discussion of survey modes:

1. The term “mode” is the characteristic of the phase of data collection and is regarded as equivalent to the terms “mode of data collection” and “survey mode”. It is strictly distinguished from the mode of recruitment and other phases of the survey process. However, the actual implementation of a mode influences and is influenced by other phases.
2. The characteristics of mode that are stable and do not vary across implementations, respondents, interviewers and contextual factors are called *inherent mode characteristics*. The inherent mode characteristics include response transmission medium, interviewer involvement during data collection, closeness of interaction with the respondent, question presentation (input) channel, response (output) channel, and computer technology used for data collection.
3. A number of other mode characteristics may be, to certain degree, constrained by the inherent mode characteristics, but vary depending on a specific survey implementation, the behaviour and characteristics of involved actors (interviewers and respondents), and various factors of social and individual contexts in which the survey is applied.

Figure 1.2: The classification of modes according to the inherent mode characteristics



Definition of mode

On the basis of the above discussion on inherent, implementation-specific, and contextual mode characteristics we propose the following definition of mode and use it throughout the dissertation:

> *The survey mode is a set of data collection procedures that determine the basic principles of communication and information transmission between the respondent and the survey questionnaire. These basic principles are grounded on inherent mode characteristics that distinguish one mode from another and do not vary across specific implementation procedures, interviewer and respondent characteristics, or contextual factors.*

Basing the definition on inherent mode characteristics means that a change in any of these characteristics introduces a different mode. However, changing only implementation-specific or contextual characteristics does not mean a different mode. This is largely consistent with the current common implicit understanding of the term *mode* in the survey literature. The separation of modes according to these criteria presented in Figure 1.2 is implied also by early uses of the term (like in Groves 1979; O'Neil 1979), general books on survey research (Biemer and Lyberg 2003; Dillman et al. 2008; Groves et al. 2009), and publications addressing survey modes (Tourangeau et al. 2000; de Leeuw 2005; Couper 2011). The definition thus does not contradict the previous work, but exposes the key differentiating dimensions of modes more explicitly. However, we use this merely as an operational definition as many alternative definitions may be equally valid.

Alternative mode definitions

The limited number of defining inherent mode characteristics may be regarded as too restrictive and may preclude some data collection approaches to be identified as independent modes. Two highlighting examples that are often treated as independent modes, but would not be identified as such under the proposed definition, are mobile-CATI (e.g. Vehovar et al. 2010) versus ordinary CATI and mobile web surveys versus web

surveys on personal computers (e.g. Couper 2013). Alternative definitions, focusing on an extended range of defining mode characteristics, are therefore possible. Here we consider two of many such possibilities: a) inclusion of a type of device used for data collection for separation of different modes, and b) separation of modes with inclusion of all implementation-specific and contextual factors.

Type of device as a defining mode characteristic

One possible extension is to include the type of device used for answering the questionnaire as the defining mode characteristic. According to Couper (2005), the move from fixed to mobile technologies presents one of the key technological trends in survey research. Various studies suggest that the use of mobile devices like mobile phones, smart phones and tablet computers can lead to some differences in the data obtained by telephone interviewing (e.g. Kuusela and Notkola 2005; Kennedy and Everett 2011) or web surveying (e.g. Fuchs 2007; Peytchev and Hill 2010; Couper 2013)[4]. Differences in the nature of devices as well as the potential effects on the responses may be a sufficient argument to separate modes also according to this aspect.

We avoided adding the type of device (mobile or fixed) as a defining mode characteristic for several reasons. We treat it as a contextual factor since it depends on the respondent-level context of surveying situation, which is mostly beyond the researcher's control. While it is possible to limit a routine access to a web questionnaire to personal computers, this is rarely done in practice and can be easily overridden by technologically more savvy respondents. In the case of telephone surveys, the restriction to fixed-line can be achieved by calling only known fixed numbers, but there is a recognized need to include mobile phones in order to assure sufficient coverage of the target population (Vehovar et al. 2004; Mohorko et al. 2013a). When mobile numbers are included in the sample, the device used for survey participation depends on the individual respondent and is thus out of a direct researcher's control.

[4] As authors of these studies note, many differences are small or merely hypothetical. More research will have to be done in order to obtain a thorough understanding of the problem.

While the studies above suggest the possibilities of differences in responses between different types of devices, too little research has been done to understand whether these differences are caused by the mobile nature of such devices or are moderated by other characteristics. Peytchev (2010) claims that small screens and keyboards may be a key factor contributing to the differences in responses when a web questionnaire is completed on a mobile device. However, a similar limitation can be found on low-resolution screens of some portable computers (netbooks). Similar holds true for mobile telephone surveys compared to fixed-line telephone surveys. There is a growing body of evidence that the obtained estimates may differ if a person is interviewed on a mobile phone. The study by Kennedy (2011) summarizes that many of the measurement differences arise due to a variety of contextual factors (like multitasking, sound quality, environmental distractions, and lower sense of privacy), which may also appear when a fixed phone is used.

Finally, there is a large variation between different mobile devices themselves that can result in unstandardized web questionnaire presentation among different respondents (Callegaro 2010a; Buskirk and Andrus 2012). It is thus unlikely that a simple separation of mobile from fixed devices would decrease within-mode differences compared to the classification based only on the inherent mode characteristics.

Implementation-specific and contextual factors as defining mode characteristic

A further extension of defining mode characteristics could include other dimensions in addition to a type of device used for answering the questionnaire. In principle, all implementation-specific, contextual or other mode characteristic could be used to separate one mode from another. This can be justified by the fact that specific characteristics of mode implementation and all external circumstances in which surveying takes place might have significant influence on the response process and consequentially cause differences in the obtained data.

The survey mode under this view is thus a complex conglomerate of a large number of characteristics related to the data collection procedures and may vary across respondents, interviewers, implementation-specific designs and contexts. Under such an extended definition of mode, it is questionable whether any two surveys could be claimed to be conducted using the same mode due to differences in implementation-specific characteristics. Furthermore, virtually each survey could be labelled as mixed-mode due to variations in contextual characteristics.

In contrast, the definition we use does not consider each variation in mode implementation as a different mode unless one or more of the inherent mode characteristics change. However, it does not preclude the possibilities of within-mode variations, occurring due to implementation-specific and contextual mode characteristics, including different types of devices used by respondents. These variations can be understood as *sub-modes* or *contextual and implementation versions* to bring the proposed conceptualization closer to the alternative ones.

A definition based on a limited number of defining mode characteristics that do not change across implementations and contexts thus offers some important advantages. It makes classification of survey modes more manageable and preserves consistency with the common intuitive understanding of the term "mode". In addition, and most importantly for our dissertation, it explicitly separates fixed characteristics from those that can be varied by researcher or vary beyond the researcher's control. In Chapter 3 we develop a conceptual model of relations between these and other survey characteristics and their complex influence on the response process.

Definition of the web survey mode

Figure 1.2 helps us define web surveys as the mode on which we focus in the dissertation:

- Web surveys are self-administered as no interviewer is involved in the data collection phase of the survey process.

- Self-administration, of course, precludes any presence of the interviewer. In addition, web surveys as defined in our dissertation also do not contain any form of virtual interaction with the interviewer. Web surveys utilizing virtual interviewer approaches are thus considered as another web-based mode (labelled *VI Web* in Figure 1.2).
- All information from and to the respondent during data collection is transmitted through the World Wide Web.
- The respondent interacts with the questionnaire using a personal computer or another type of web-capable device.
- The main question presentation channel is visual. Although additional auditory and multimedia elements can be used to extend the range of available stimuli, the questionnaire is essentially delivered visually[5].
- Answers are entered electronically into a computer, using a keyboard, mouse or other pointing device, or touchscreen.

Self-administration and the respondent's use of computer technology for data collection result in highly flexible influences of these defining properties of the mode on implementation-specific and contextual characteristics. The questionnaire can be programmed in many different ways, include highly interactive features, and respondents have a great deal of freedom during survey administration. We address these aspects into details in Chapter 3.

1.2 Survey errors

A general and straightforward definition provided by Hansen et al. (1951, 147) states that a survey error is "the difference between a survey estimate and the value which is

[5] Couper (2008) reports on attempts to present web questionnaires using only an auditory input channel, which proved to be unsuccessful presumably due to the web being a primarily text-based medium. In line with our discussion of defining mode characteristic, such an approach would essentially mean a new mode that departs from the common definition of web surveys as a mode grounded on the visual input channel.

estimated". Since the objective of survey data collection is usually to estimate an aggregated parameter of interest for the target population (e.g. average, proportion or some measure of association between variables), the survey errors can be observed at two levels: the level of individual respondent and the level of population. Both levels include inference, the former from the respondent's answer to the individual's survey measure and the latter from an estimate based on a sample of respondents to the target population (Groves et al. 2009).

1.2.1 Survey error at the level of the individual respondent

A survey error at the level of the individual respondent is a failure to obtain the true value of the variable of interest from a respondent. This assumes that each respondent has a true value on the variable (questionnaire item), which could be theoretically obtained under a completely error-free survey administration[6]. Since this is in practice almost always impossible, the measured value (survey response) is the sum of the true value and survey errors at the respondent's level:

$$y_{git} = \tau_{gi} + \epsilon_{git} \qquad \textit{Eq. 1.1}$$

where y_{git} is the measured value on the variable Y_g for the i-th respondent in the data collection trial t, τ_{gi} is the true respondent's value of the variable and e_{git} is the result of all errors that affect this respondent's answer to the item corresponding to this variable in the trial t. The errors stem from different sources we discuss later, e.g. due to erroneous reporting, refusal to answer the whole questionnaire or a particular item, inappropriate recoding during post-survey activities and so on.

[6] We avoid the discussion whether or not the existence of the true value of a variable is a reasonable assumption and we treat the true value as an underlying theoretical concept. Further discussion of the problem of true value is provided, for example, by Hansen (1951), Deming (1960), Novick (1966), Allen and Yen (2002) and several others.

These errors are either systematic or random; the former specifically affect individual data collection trials and result in the variance at the level of the respondent[7], while the latter are common to all trials and thus causes the response bias. The error term ϵ_{git} is the sum of i-th respondent's bias b_{gi} that is common to all measurement trials of the variable Y_g, and the corresponding variance that is specific to the trial t, e_{git}. The above equation can be then rewritten as:

$$y_{git} = \tau_{gi} + b_{gi} + e_{git} \qquad \textit{Eq. 1.2}$$

This very important distinction between random and systematic errors relies on the largely conceptual assumption of *replicability* (Groves 2004). If we could take a large number of independent data collection trials (measurements) on the same respondent and take the expectation over the individual's response distribution, we would obtain:

$$E_t(y_{git}) = E_t(\tau_{gi}) + E_t(b_{gi}) + E_t(e_{git}) = \tau_{gi} + b_{gi} \qquad \textit{Eq. 1.3}$$

The variance component is not present in this term any more, since the expectation of random errors is zero, $E_t(e_{git}) = 0$. If $E_t(b_{gi}) = b_{gi} = 0$, there are no systematic errors for the i-th respondent over trials. By taking the variance over trials, we obtain information about the variation of responses between trials for the i-th respondent:

$$Var_t(y_{git}) = Var_t(\tau_{gi}) + Var_t(b_{gi}) + Var_t(e_{git}) = \sigma^2_{gi} \qquad \textit{Eq. 1.4}$$

Since the bias component b_{gi} and true value τ_{gi} are equal in all data collection trials, replications allow us to discuss the variance of repeated measurements on the same respondent. This requires that certain survey characteristics are fixed, i.e. they do not change over trials. In the literature these characteristics are variously called *essential survey conditions* (Hansen et al. 1951), *survey design* (Groves 2004), *adopted system of work* (Zarkovich 1966), etc. In general, we can regard them as constant ways of implementing the survey design phases, presented in Figure 1.1 (page 30). Some examples of

[7] We deliberately avoid the use of the term *response variance*, which commonly refers to a random *measurement* error. According to Groves (2004, 8–9), the term denotes "the variation in answers to the same question if repeatedly administered to the same person over different trials or replications".

such characteristics include sampling techniques, data collection mode, recruitment and solicitation approaches, questionnaire content, design and application, interviewers involved in the data collection, supervision of interviewers, fielding period and so on.

The variation is also affected by a large number of external factors beyond the control of survey design that may include, for example, specific interviewer's behaviour, temporal respondent's distractions, respondent's mood or even weather at the time of data collection (Hansen et al. 1951; Lessler 1984). In order to pinpoint a specific cause of variation, all survey design and external factors would need to be held constant or controlled for, except those for which variation is to be measured. While the complete elimination of sources of variations cannot be achieved in practice, it is possible to restrict the range of variations by assuring as much control over the surveying process as possible (Hansen et al. 1951).

The notion of essential survey conditions has important implications for studying differences between modes. To ascribe all obtained variations to the data collection mode, it would be necessary to vary only the mode itself while keeping all other factors equal or modelling them out. However, as we discussed in the previous chapter, each mode has inherent characteristics that constrain but not determine other factors influencing the implementation of the data collection procedures. Some essential survey conditions, in addition to the mode itself, will thus necessarily change between measurement trials.

The estimation of bias is conceptually more complex since it requires the known true value. For some variables, the true value can be obtained from another, presumably accurate source known as *the gold standard*. This is the most commonly used method of estimating the deviations from the true value in survey research and relies on relatively simple statistical procedures (Groves and Lyberg 2010). However, for many variables (like attitude questions) it is not possible to obtain the true value directly.

1.2.2 Survey error at the aggregate level

The concepts of survey errors at the individual level directly apply to the level of parameter estimation based on a sample of survey respondents. The parameter θ_g represents the true value of some population characteristic (like mean, proportion, correlation coefficient, etc.) that would be obtained if the variable g was measured error-free and over all members of the population. In a survey, this parameter is estimated on a sample of respondents using the statistic $\hat{\theta}_g$, calculated for respondents i $(i = 1,2, \dots, n)$ in the sample s and the trial t. The obtained estimate is the sum of the parameter value and influences of errors that can be systematic (bias B) or random (ϵ):

$$\hat{\theta}_{gsit} = \theta_g + B_{gsit} + \epsilon_{gsit} \qquad \textit{Eq. 1.5}$$

For conceptual replications of surveying trials it is usually assumed that all essential survey characteristics are fixed, but different individuals are selected into different samples (Groves 2004). Expectation is then taken over all samples of respondents and trials, i.e. over sampling and response distribution:

$$E_{sit}\left(\hat{\theta}_{gsit}\right) = E\left(\theta_g\right) + E\left(B_{gsit}\right) + E\left(\epsilon_{gsit}\right) = \theta_g + B \qquad \textit{Eq. 1.6}$$

Again, the expectation of variable errors is zero. Following the logic in Eq. 1.4, we obtain the variance of the statistic $\hat{\theta}$, σ_g^2.

The survey statistics will be an unbiased estimator of the population parameter if there are no systematic errors present, which is the case when positive and negative errors cancel each other out.

1.2.3 The total survey error paradigm

The total survey error (TSE) paradigm importantly enhances the conceptual understanding of survey errors by systematically disentangling them into components in order to

identify error sources and their causes (Biemer 2010). The idea of survey error decomposition can be traced back to early works in survey statistics by Deming (1944) and was further developed in later classical works in the field (e.g. Cochran 1953; Kish 1965). According to Biemer (2010), the first use of the term *Total Survey Error* can be attributed to the work by Anderson et al. (1979), who provided the typology of survey errors which is, with some variations, still used today.

The central aim of the TSE approach is to consider all error sources affecting the survey data in any phase of the survey process presented in Figure 1.1 (page 30). According to Biemer (2010, 817), the TSE is the "accumulation of all errors that may arise in the design, collection, processing, and analysis of survey data". The approach thus attempts to address the problem of survey errors from a broader perspective than sampling statistics, psychometric theories, or econometrics (Groves 2004).

There is often a perceived difference between those interested in an overall effect of the total survey error and those attempting to disassemble the total survey error into error components and estimate them as separately as possible. This can be viewed as a reflection of interests of survey practitioners versus interests of more theoretically oriented survey methodologists. To some degree, this view is justified when focusing on the errors of estimation itself: the practitioners may be more interested in the overall accuracy of the obtained estimate and less in individual error sources that contributed to the deviations. As soon as the focus moves to the *error reduction* rather than the *error estimation*, understanding the individual error sources becomes crucial from both theoretical and practical perspectives. As Biemer (2010) states, parsing the error into smaller components makes the error sources more manageable and thus provides a better strategy for coping with survey errors.

The TSE paradigm found its way into many organizations as a conceptual tool for planning, understanding and learning the approaches for increasing the quality of survey data (Weisberg 2005). It is also a useful tool for qualitative assessment of procedures for survey quality assurance (Biemer et al. 2012). Unfortunately, it has had only limited

Table 1.3: Basic typology and descriptions of the TSE components by Biemer and Lyberg (2003)

TSE	Bias / Variance	Sampling errors	Sampling error	Difference between the estimate and the parameter due to observing only a subset of the target population.
		Non-sampling errors	Specification error	Difference between the concept implied by the survey question and the concept that should be measured in the survey.
			Frame error	Failure of the frame to provide a required probability of selection for some units of the target population.
			Nonresponse error	Failure to conduct the data collection on a unit selected in the sample.
			Measurement error	Difference between true and reported value of the variable.
			Processing error	Inappropriate processing of the collected data during the data preparation.

success in the empirical estimation of individual error sources[8]. Platek (2001) attributes this to the disproportionately high attention devoted to sampling errors at the expense of studying other error components, lacking collaborative efforts and complexity of statistical models that would enable disentangling individual error components.

Typologies of error components and sources in the TSE

Although the general outline of error components is relatively consistent between different authors, Groves and Lyberg (2010) expose a large difference in typologies and the labelling of errors. They caution that any listing of errors is likely to be lacking due to the continuous emergence of new errors with developments in survey methodology.

At the very heart of the TSE lies a distinction between random errors (variance) and systematic error (bias). Each error component can consist of both types of errors (Groves 2004). Table 1.3 depicts a very general typology of TSE components as presented by

[8] There are, however, promising efforts of many researchers to accomplish this complex task. The annual International Total Survey Error Workshop (ITSEW), for example, is one of the central events that deals with a whole range of topics related to various survey error sources and their interrelations.

Biemer and Lyberg (2003). A similar typology is provided by Groves (2004) who omits processing errors, uses the term *coverage* instead of *frame* error and groups components into errors of non-observation (coverage, nonresponse, and sampling) and errors of observation (measurement error).

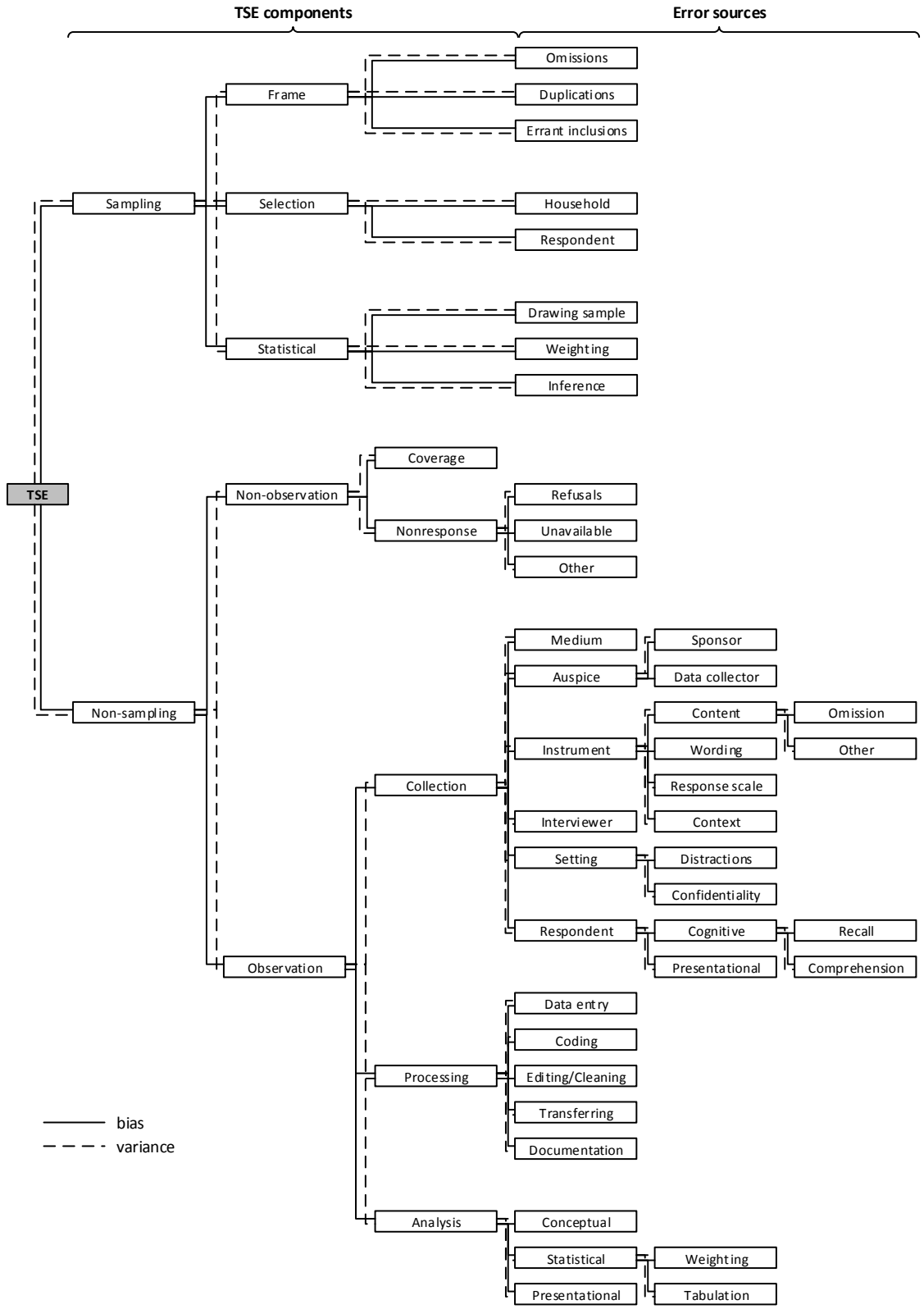
This basic typology can be extended in various ways. For example, Biemer (2010) adds sampling schemes, sample size and estimator choice as additional sub-components of sampling errors. Another common extension is the addition of sources of measurement error, usually the interviewer, respondent, instrument (questionnaire) and mode of data collection (e.g. Groves 2004).

One of the most comprehensive recent extensions, proposed by Smith (2011), is presented in Figure 1.3. The author lists a large number of error components and corresponding error sources. In the paper, he additionally lists several other elements that could be added to the scheme, for example, the detailed two-dimensional categorization of nonresponse according to the level (unit, module, and item nonresponse) and the reason for nonresponse (refusal, unavailable, other). He also acknowledges the possibilities of adding aspects of survey administration (similar to mode characteristics we discussed in the previous chapter), the incorporation of response process elements, open-ended and closed ended wording effects and several others. While Smith does not provide a detailed definition of each component and source presented in the scheme, his work offers a valuable insight into the variety and complexity of survey errors.

It is important to bear in mind that every conceptual scheme of the total survey error components presents an (over)simplification of the problem. In addition to a large number of components that can be included into a typology, the complex interplay between these components and error sources needs to be considered.

Error components can contribute to the TSE in different directions, reducing or even cancelling each other out. This can result in low TSE regardless of possibly high errors at the level of individual components. The same can occur at levels of error sources and error causes. For instance, low data collection error can be the result of sources contributing to it in different directions, reducing the overall measurement error.

Figure 1.3: The extended typology of TSE components by Smith (2011)



Note: Adapted from Smith (2011). Distinction between error components and error sources added to the original figure.

An important conceptual issue emphasized by Smith (2011) is that there are many alternative and largely arbitrary ways of categorizing lower levels of the error taxonomy. The key problem is that one error source can contribute to several error components. Probably the most typical case is the common placement of interviewer as a source of measurement error. However, different interviewers achieve different response rates and sometimes perform sampling of units, thus potentially contributing also to nonresponse error[9] and sampling error. Another common case is that specific recruitment procedures influence not only nonresponse errors but also measurement errors due to, for example, lower perceived survey legitimacy or privacy. The study by Heerwegh et al. (2004) found that personalized e-mail invitations to a web survey increase response rates, but at the same time also increase measurement errors, likely due to a lower sense of perceived privacy. Therefore, it may be worth noting that schematic representations of the TSE typologies include only some key sources contributing to individual error components, but not precluding the possibility that the same error sources affecting other components as well. Figure 1.3 acknowledges this by (non-comprehensively) including some error sources under multiple error components.

The complexity of interrelations of error components due to common error sources is further increased because of interactions between them. Although the studies of relationships between errors in the TSE are significantly underexplored (Groves 2005), it is acknowledged that one error source can influence another in either direction. Smith (2011) illustrates this with a number of cases, like interaction between the interviewer's and respondent's race for race-related questions, interactions between respondent abilities and mode of responding and so on. One of the consequences is that simple indicators of error magnitude are often inadequate. As shown by Vehovar et al. (2010) there are also complex relations between recruitment procedures and characteristics of respondents. In their study, the use of incentives increased response rates but also nonresponse error, likely because incentives were more effective for specific respondents.

[9] A comprehensive overview of interviewer effects on nonresponse is provided by Groves and Couper (1998).

This multifaceted nature of survey error reminds us that the investigation of any error component requires its placement into the context of other components as well as careful consideration of error sources and their interrelations at different levels of the survey process. The TSE perspective here offers a promising framework for a conceptual understanding of the problem of different survey errors, although it is unlikely that it would be soon possible to provide a complete empirical verification of the complex picture.

Measurement errors and relations to psychometrics

A vast majority of our discussion of mode effects is going to take place in the context of measurement (response) errors. Groves et al. (2009) describe the measurement error as the observational gap between the ideal measurement that would provide true values of the observed variables and the responses obtained. For a formal description of measurement errors in surveys, the Eq. 1.1 (page 46) is largely applicable if we restrict the set of contributing error factors to those arising from measurement itself, and not due to the item or unit nonresponse, post-processing mistakes, and so on.

Biemer and Lyberg (2003) note that measurement errors may be the most damaging error component for many surveys. As with other error components, measurement errors can be either systematic or random. A common example of a systematic measurement error is underreporting of socially undesirable behaviour. On the other hand, random measurement errors often occur because respondents use various contextual cues in the measurement environment (Groves et al. 2009).

The concept of measurement errors is the junction point of survey statistics and psychometrics, yet not without some important terminological and substantial differences that are comprehensively elaborated by Groves (2004). Probably the most influential psychometric theory is classical test theory, CTT (Novick 1966), with a key assumption that the expected value of independent measurements of the same individual equals to that individual's true value[10]. That is, no measurement bias is present:

[10] For a complete overview and formalization of other assumptions of CTT see for example Allen and Yen (2002).

$$E_t(y_{git}) = \tau_{gi} \qquad \textit{Eq. 1.7}$$

As Groves (2004) notes, the CTT rules out the possibility of biased measures by stating that potential biasing effects do not add a systematic error to an indicator measuring the true score, but rather change the true score. This means that the indicator becomes an unbiased measure of some other construct.

Despite opposing views on the existence of the measurement bias, there is a conceptual equivalence between the simple response variance in survey statistics and reliability in psychometrics: both are the measures of random response variations over all respondents and measurement trials (Groves et al. 2009). There are several alternative interpretations of the reliability, but it is generally defined as the proportion of the observed variance that is accounted for by true-score variance (Alwin 2007):

$$\rho_g = \frac{\sigma^2_{T_g}}{\sigma^2_{Y_g}}, \qquad \textit{Eq. 1.8}$$

where T_g and Y_g are true and observed scores of a variable in the population, respectively.

However, a consequence of the different treatments of a measurement bias between the two disciplines is that there is a complex correspondence between the psychometric notion of validity and the survey statistics notion of bias. Validity can be defined in numerous ways. Some of the most common validity concepts are: *content validity* which refers to the extent to which the selected indicators represent different domains of the measured construct, *criterion validity* as the correlation between the test score and the criterion score, and *construct validity* which is the correlation between the true score and responses (measurements) over trials (Allen and Yen 2002; Groves 2004). Systematic errors that constitute bias in survey statistics do not reflect in the change of validity coefficient as a correlation between the true score (or criterion) and the obtained response. This follows directly from the definition of the correlation coefficient and

properties of variance, stating that the added constant value does not change the variance or covariance[11].

It should be noted that some extensions of the CTT with relaxed assumptions acknowledge the possibility of non-random (B_g) in addition to random measurement errors (E_g):

$$Y_g = T_g^* + B_g + E_g \qquad \textit{Eq. 1.9}$$

Under the classical model non-random errors are not separated from the true value, thus $T_g = T_g^* + B_g$. While such separation is theoretically possible under the extended model, it introduces more complex definitions of response variances, reliability and validity (Groves 2004). Furthermore, empirical separation of random and non-random sources of errors can be only partially done using the currently available advanced statistical modelling, like multitrait-multimethod designs and confirmatory factor analyses (Alwin 2007).

1.3 The concept of mode effects

The rather lengthy discussion of basic and background concepts serves as a foundation for a conceptualization of mode effects that we use in the dissertation. The most common variations in the meaning of the term arise because some mode comparison studies focus on the differences between the two mode systems (i.e. the set of all phases of the survey process presented in Figure 1.1 on page 30) and others on differences between the two modes of data collection. Biemer and Lyberg (2003) name this broader and narrower meaning of mode effects as *mode system effects* and *pure mode effects*, respectively. The former include the set of effects that stem from all error sources in a specific implementation of the survey process and may affect all components of the TSE, including errors due to sampling and solicitation procedures. The latter, on the contrary,

[11] The derivation and proof can be found in virtually any advanced statistical textbox, for example in Rice (2007).

are limited to effects attributable to a specific mode of data collection used (Biemer and Lyberg 2003). In line with this understanding, (pure) mode effects are the result of influence of various factors arising from one or more data collection modes used in the survey process. In the remainder of the dissertation we use the term *mode effect* as a synonym of *pure mode effect*. However, some additional considerations are needed in order to provide more concise definition and place mode effects into the framework of TSE.

In section 1.1 we based our definition of the survey mode on the inherent mode characteristics. Consistently, we consider mode effects as the influences of these characteristics on the survey estimates. However, this influence is not limited to a direct relationship between inherent mode characteristics and the estimate. As our elaboration of mode characteristics showed, there are many complex interdependences between inherent mode characteristics, implementation-specific and contextual characteristics. Because of this, various mode-related variables are expected to mediate or moderate the relationship between the inherent mode characteristics and survey errors. Similar causal relations occur with variables that are not directly mode-related. For example, the influence of a specific question presentation and response channels may be moderated by a specific question wording or type (Dillman et al. 2008), which is related more to the instrument design than the mode itself.

We thus understand mode effects as:

> *All direct and indirect influences of the inherent mode characteristics on the obtained survey estimates. Sources and causes that might increase or decrease mode effects include all factors that mediate, moderate or in any other way change the nature of a relationship between the inherent mode characteristics and the survey estimate.*

Under such a definition, mode itself can be placed into the typology of TSE as an error source, further subdivided into six sources, each corresponding to one of the inherent mode characteristics. Yet the complex interrelation of error components, sources, and causes within the TSE paradigm should be taken into account. While mode effects are

commonly identified as a source of measurement error (Biemer and Lyberg 2003; Groves 2004), they can influence other error components as well. For example, several studies (reviewed in the meta-analysis by Lozar Manfreda et al. 2008) suggest that lower response rates in web surveys may be influenced also by the web as a medium and mediated through mode-related factors like familiarity with the medium and perceived legitimacy.

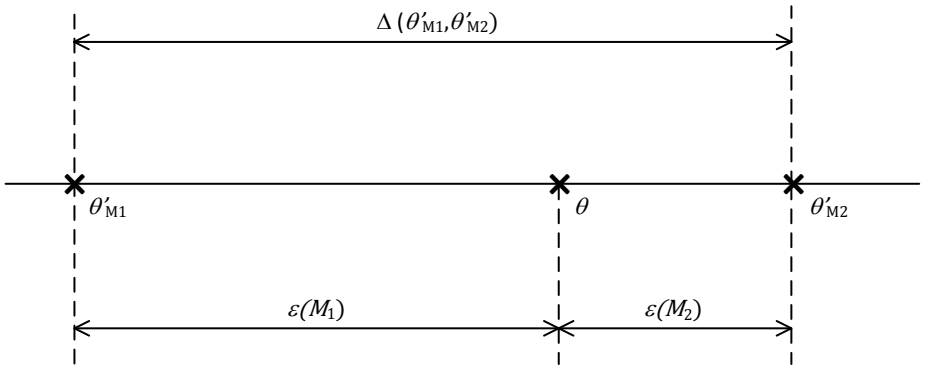
In sum, while we define mode effects as the influence of inherent mode characteristics on the obtained data, their underlying causes might lay not only in specific characteristics of the survey mode, but also in various other factors at different stages of the survey process. Similarly, while they mainly influence the measurement process, they can also affect other error components like item and unit nonresponse. Which causes and influences of mode effects are taken into account depends on the scope of the study. However, potential effects of the mode on unit nonresponse should be strictly separated from effects of solicitation procedures and other factors unrelated to the mode itself. In practice, this is often impossible to do. In our dissertation, we focus on the impact of mode effects on the response process and thus primarily on their contribution to measurement errors.

Mode effects are sometimes regarded as the difference in the obtained estimate between two or more survey modes. While such differences present a common indicator of the presence of mode effects, assuming that other error sources do not affect any of the compared modes or are modelled out, it should not imply that mode effects do not exist when only one mode is used.

If we assume that the estimate is affected only by mode effects, we can represent the between-mode difference Δ between two modes M_1 and M_2 and mode effects of each mode $\epsilon(M_1), \epsilon(M_2)$ as shown in Figure 1.4.

Vannieuwenhuyze and Loosveldt (2012) name the between mode difference, assuming that measurement error is the only error source, as a marginal measurement effect on a function of the observed set of target variables, X. Assuming that all respondents

Figure 1.4: Between-mode difference and mode effects



would respond to both modes and that there is no variability in responses given a particular mode, the marginal measurement effect $M\big(f(X)\big)$ is then:

$$M\big(f(X)\big) = f(X|M_1) - f(X|M_2) \qquad \textit{Eq. 1.10}$$

The between-mode difference for two modes under the above assumptions can be written in terms of survey errors we used before:

$$\Delta\big(\theta'_{M_1}, \theta'_{M_2}\big) = \theta'_{M_1} - \theta'_{M_2} = \big(\theta + B_{M_1} + \epsilon_{M_1}\big) - \big(\theta + B_{M_2} + \epsilon_{M_2}\big) = \qquad \textit{Eq. 1.11}$$
$$= B_{M_1} - B_{M_2} + \epsilon_{M_1} - \epsilon_{M_2}$$

Considering a variety of factors that can affect the influence of inherent mode characteristics on the survey estimate and as implied by Eq. 1.11, mode effects can be expected to have systematic or random effects on the survey estimates. Considering the parallels with the CTT this means that they can also decrease reliability and validity of the measurement.

When mode effects are observed as a cumulative effect, which is also the case in overall measurement difference, the complexity of relations between errors in the TSE should be again taken into account. Different sources of mode effects may contribute to the error in different directions, potentially cancelling each other out. While this is certainly positive from the practical point of view, the thorough understanding of mode effect causes still require the investigation of effects at the level of individual sources. This, to some degree, reflects the difference between practically oriented interests in the overall

TSE estimation and more theoretically driven efforts to disentangle the TSE into individual components. Another issue is that two modes can be subjected to relatively high but similar mode effects, resulting in low measurement difference despite substantial influence of mode effects on survey error.

This chapter established the framework for the elaboration of a variety of influences of mode effects on survey estimates. In the next chapter we cover the survey response process and the corresponding deviations that lead to measurement errors. In Chapter 3 we then study how specific characteristics of the web mode influence the response process and contribute to the emergence of mode effects.

Chapter 2
Survey response process

Surveying is essentially a form of conversation, although a very specific standardized one. Tourangeau and Rasinski (1988, 302) draw parallels between survey interviews and everyday conversations:

> *Two people – the interviewer and the respondent – take part, and the interview consists of conversational units involving connected questions and answers on a given topic. Although the interview is clearly a very specialized form of conversation, it may still follow many of the principles that guide more ordinary and less structured conversations*.

Communication is the foundation of every survey, even when it is not conducted by the interviewer; in self-administered modes, the communication between the respondent and the researcher is mediated by the questionnaire itself (Jenkins and Dillman 1997). Survey communication is thus importantly determined by the survey mode and its inherent characteristics. How information is transmitted to and from the respondent, to what degree interviewers are involved in communication, and what medium and tools are used for this purpose, does not only determine the nature of communication itself, but also the respondent's cognitive tasks and behaviour in survey situation.

Effects of mode characteristics can alter the way in which the respondent conducts the task of answering survey questions (Jäckle et al. 2006). Mode effects occur when this leads to increased measurement error. We therefore based our exploration of mode effects in web surveys on understanding the processes that respondents perform in order to derive accurate answers to survey questions.

The theories, or models, of the survey response process fall into three wide groups. The first group includes models of cognitive processes that take place when respondents answer survey questions. In the second group are models with a focus on reasons for

deviations in the response process and their impact on accuracy of survey data. Finally, some models attempt to extend their views by explicitly including the aspects of interviewers, interactions between respondents and interviewers, or interactions between respondents and self-administered questionnaires. We do not go in-depth into the cognitive foundations of the response process theories, but rather provide an overview of their key elements to which we refer later during our discussion of mode effects in web surveys.

2.1 Information processing models for respondents

The models of cognitive processes through which respondents answer questions are sometimes labelled as *information processing models* (Jobe 2003). Different models share many common aspects and, at least generally, agree about cognitive steps taken by respondents to come up with an accurate answer to the survey question (Schwarz 2007): the respondent is expected to understand the question, retrieve relevant information from memory, evaluate and integrate the retrieved information, and form an answer.

We limit our discussion to the models, describing the response process at the level of individual respondent. However, especially for business surveys, it becomes important to consider additional business-level aspects of the response process, including organisation of the survey participation, retrieval of information from the business information systems and data authorisation (Bavdaž 2007).

2.1.1 The model by Tourangeau and colleagues

A well-known information processing model of the response process was proposed by Tourangeau (1984). A comprehensive extension in the *Psychology of Survey Response* (Tourangeau et al. 2000) offered definitely the most elaborated and widely used approach to explaining psychological foundations of answering survey questions. Because

Table 2.1: Cognitive stages of the response process by Tourangeau et al. (2000)

Stage	Specific cognitive processes
Comprehension	a) Attend to questions and instructions
	b) Represent logical form of question
	c) Identify question focus (information sought)
	d) Link key terms to relevant concepts
Retrieval	e) Generate retrieval strategy and cues
	f) Retrieve specific, generic memories
	g) Fill in missing details
Judgment	h) Assess completeness and relevance of memories
	i) Draw inferences based on accessibility
	j) Integrate material retrieved
	k) Make estimate based on partial retrieval
Response	l) Map judgment onto response category
	m) Edit response

of a detailed and sufficiently general treatment of underlying cognitive operations, a majority of other models can be consistently integrated with this work (Jobe 2003).

The authors describe a respondent's task consisting of four key cognitive stages or components: comprehension, retrieval, judgment, and response. Each of them encompass several specific processes, listed in Table 2.1.

During the **question comprehension** stage, the respondent establishes the first contact with a question. The process of question comprehension is divided into two main parts: 1) the obligatory representation *of* the question, and 2) the optional representation *about* the question. The former consists of cognitive operations related to the understanding of the formal grammatical and logical structure of the sentence and is thus expected to be relatively stable across the respondents. The latter representation requires the respondent to infer the question's intention, i.e. its pragmatic meaning (Schwarz 2007). This process incorporates specific inferences made by the respondent on the basis of their own viewpoint, knowledge of the topic, relation to the interviewer, context of the question, and many other factors.

The main role of **retrieval** is to obtain relevant information from the long-term memory to use it in ongoing cognitive processes. The respondent's memory assumes the central

role of recalling relevant memories and attempting to fill-in potentially missing details. Retrieval processes can be aided by memory cues generated by the respondent themselves, or provided by the survey context (Tourangeau et al. 2000).

During **judgment,** the respondent performs cognitive processes of reviewing, combining and supplementing the retrieved information. Although the retrieval processes can sometimes result in an accurate direct answer, and thus dismiss a separate judgment processes, this is usually not the case (Tourangeau and Rasinski 1988). Similarly, judgment may completely supersede retrieval processes if respondents base their answers on a general plausibility of response (Tourangeau et al. 2000).

Finally, the respondent forms and reports their **response**. The mapping process is responsible for the translation of judgmental outcomes onto the expected response format (e.g. selection of the most appropriate response category). However, the respondents may conduct further editing prior to reporting the response. This is most commonly done to assure consistency or due to social pressure. According to Tourangeau and Rasinski (1988), the provided answer can therefore be regarded as a compromise between an accurate answer based on cognitive processing and final editing for other purposes, like social desirability.

Two general remarks are worth making regarding the presented model. First, although the authors illustrate the model in a sequential way, its detailed elaboration accounts for possible re-routings. Some instances discussed by Tourangeau et al. (2000) include superseding between judgment and retrieval, looping between stages in case of identified missing details (particularly between retrieval and judgment stages), and selective partial or complete omissions of individual cognitive processes. We return to the last situation in our discussion of two-track theories of the response process (section 2.2).

Secondly, a short description of the response stages and corresponding cognitive processes should not create an impression that the response process is actually performed in a standardized way, free from influences of external factors. Throughout this chapter we will highlight many such factors that may change the response process in various

directions in different modes. How much this contributes to mode effects depends on the degree to which these factors are related to specific characteristics of the mode.

2.1.2 Alternative information processing models

Although the approach by Tourangeau and colleagues presents our main theoretical foundation for discussing factors that influence mode effects in web surveys, several other models have evolved in survey methodology (for a review, see Jobe and Herrmann 1996; Bavdaž 2007). We consider some of those to highlight different views on the response process, and list their key differences to the model presented above.

Cannell, Miller, and Oksenberg

The response process model by Cannell et al. (1981) is an updated version of probably the first multistage model of the response process (Cannell et al. 1977). The initial version outlined the connections between some key concepts of the respondent's information processing, including interaction with the question stimulus, activation of cognition, and information retrieval. The revision introduced the schematics of cognitive stages, which serves as a basis for a majority of later models. However, compared to later approaches there are some specific differences. The authors include a separate stage of response evaluation in terms of accuracy, which later models commonly integrate into the judgment and response components. In addition, the authors do not explicitly address mapping of the answer onto response categories of closed-ended questions.

The model is far from a merely historical prototype value. It includes a clear distinction between processes carried out by conscientious respondents who answer the questions carefully and deviations that lead to skewed responses. Deviations can occur at any stage if the respondent starts modifying a response based on cues from interviewer, questionnaire, beliefs, values, attitudes, or goals. This offers an integration of information processing models with two-track theories. However, while the authors do stress

the important role of the respondent's memory, they treat inaccurate responses largely as a result of the respondent's conscious unwillingness to devote sufficient cognitive effort. Failure to address other factors of inaccurate reporting is one of the key disadvantages of this model exposed by Tourangeau et al. (2000), in addition to lacking elaboration of the underlying cognitive processes.

Strack and Martin

The model by Strack and Martin (1987) use a somewhat different categorization of cognitive processes by merging retrieval and judgment stages into opinion generation component, and splitting the response stage into response formatting and response editing. Bavdaž (2007) comments this as a reflection of Strack's and Martin's focus on attitudinal questions. Their approach suggests strong interrelations between cognitive processes related to retrieval and judgment, which is also acknowledged by Tourangeau et al. (2000).

The model introduces a second, suboptimal response track within the opinion generation stage. Respondents following the suboptimal track base their answers on retrieval of readily available judgments from previous questions instead of comprehensive retrieval of the most relevant information. In their critique of the model, Tourangeau et al. (2000) emphasize the importance of treating the two tracks strictly as ideal types. Respondents are more likely to use a combination of tracks and thus retrieve both, some specific information and some readily available judgments.

Willis, Royston, and Bercini

Willis et al. (1991) significantly extended the role of judgment processes. Some judgments are needed even before the retrieval of relevant information can start. This pre-retrieval judgment serves to the respondent to decide whether to conduct search for relevant information in memory. When such information is retrieved, the respondent again judges whether further retrieval is needed, or whether the answer should be altered for other purposes (like social desirability). The judgment component is thus not

limited to the evaluation of retrieved information, but rather presents a central decision-making mechanism for the whole response process (Bavdaž 2007). It is also responsible for the final mapping of the answer onto response categories. The response stage itself is thus limited to a simple provision (output) of the answer. In contrast to a majority of other models, the authors thus stress a non-sequential nature of the response process, where respondents essentially and continuously move back and forth between cognitive components.

Most other models acknowledge the possibility of alternative paths and at least implicitly describe the decisions respondents need to make at each stage before proceeding further. The more explicit inclusion of these aspects into the schematic representation of the response process is a welcoming addition to underline the high flexibility and complexity of respondent's cognitive operations during surveying.

Forsyth and Hubbard

Forsyth and Hubbard (1992) presented a model to be used for expert coding of question properties. They split the comprehension component of the Tourangeau's model into comprehension and interpretive processes. The former set of processes help the respondent to understand the question and the latter is responsible for the generation of a specific representation of the task. This can be roughly paralleled with the representation *of* the question and representation *about* the question described by Tourangeau et al. (2000).

Schwarz and Oyserman

Although the response process model by Schwarz and Oyserman (2001) focuses on behavioural (factual) questions, its basic conceptualization follows other more general approaches. Their inference and estimation component resembles a somewhat extended judgment stage of the Tourangeau's model. It additionally embraces a minor part of retrieval processes responsible for filling-in partial memories through inference, and some aspects of response formulation, like the rounding of numbers to open-ended

questions. Schwarz and Oyserman stress the essential interconnection between the formulation of response, judgment, and comprehension of the question. Finally, they treat mapping and editing of answers as two separate stages, similar to Strack and Martin (1987).

Summary of information processing models for respondents

A detailed comparison of different models is difficult to perform due to highly variable elaboration by their authors. However, by accepting some assumptions about missing details, the models exhibit a high consistency in the identification of the main cognitive stages of the response process (Figure 2.1). On the other hand, specific emphases and often subtle variations between these theoretical approaches, substantially driven by particular interests of their authors, offer valuable insight into different parts of cognitive processes that govern the quality of provided answers.

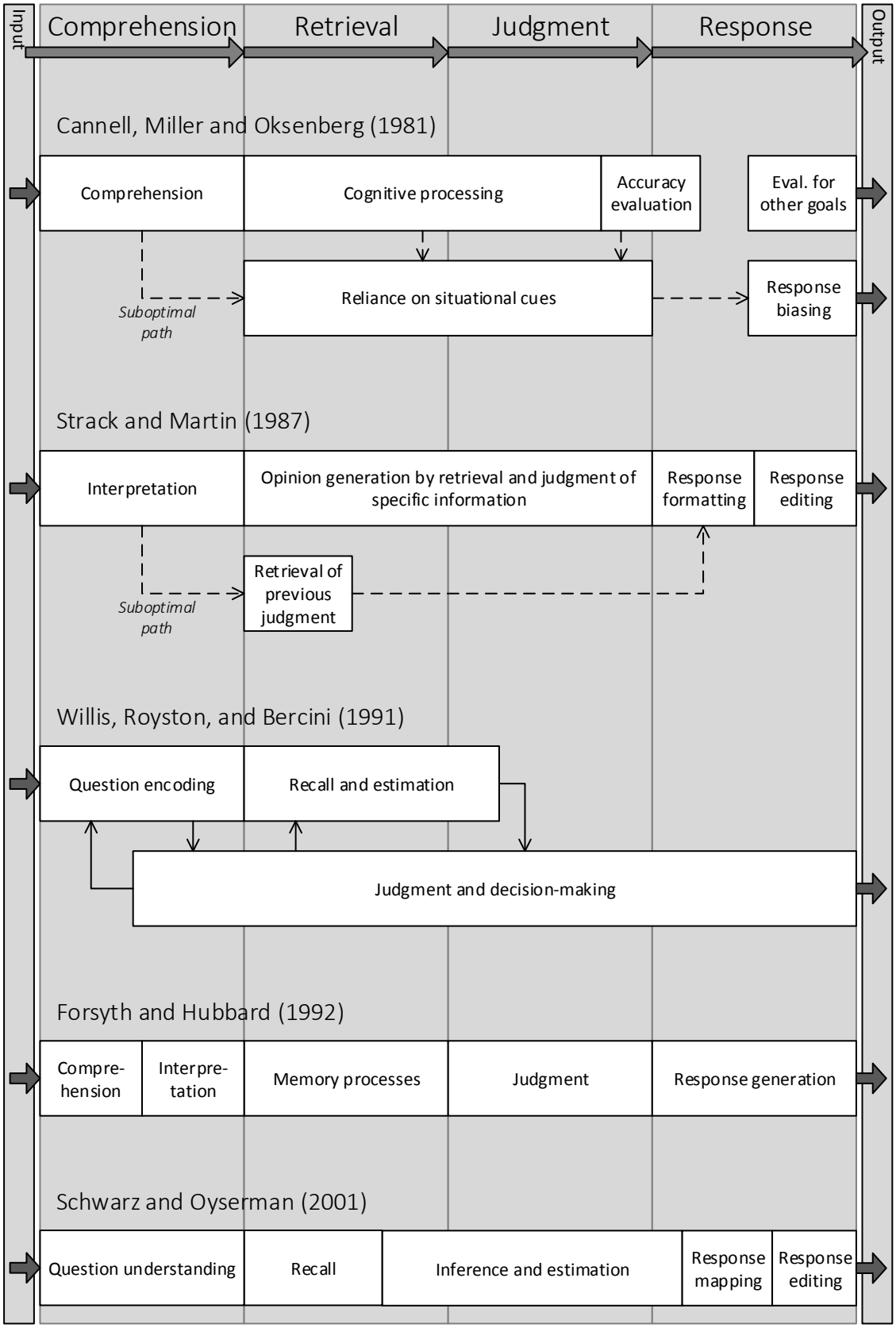
2.2 Models of response process deviations

The complexity of cognitive processes in the response process reveals the burden imposed on respondents. They are expected to perform a large number of cognitive operations in order to provide an accurate answer. Unsurprisingly, they are often not motivated to or even incapable of performing this task thoroughly. The motivational aspect of answering survey questions is described by Tourangeau et al. (2000, 254):

> *There's no reason why respondents should work hard to answer the difficult questions posed in many surveys. The evidence indicates that many respondents may choose to take it easy instead.*

Deviations from the expected full cognitive processing of survey questions is the focus of two-track theories of the response process. One track presents an optimal way of question processing and results in an accurate answer. The second track leads through some form of mental shortcutting used by the respondent to reduce cognitive resources at the expense of accuracy.

Figure 2.1: Schematic comparison of key components of information processing model by Tourangeau et al. (2000)*, and alternative models*



Some authors include the two tracks directly into their proposed information processing models. The two examples we presented above are the models by Strack and Martin (1987) and Cannell et al. (1981). The former focuses on non-optimal processing of attitude questions when respondents recall previous judgments instead of creating new ones. The latter is more general and allows for the occurrence of deviations at each stage of the response process, although it does not comprehensively elaborate the underlying causes.

2.2.1 The satisficing model

Krosnick and Alwin (1987) introduced a two-track model of the response process based on Herbert A. Simon's concept of satisficing in decision-making. When making decisions, individuals often try to reduce the psychological resources required by searching for acceptable, but not necessarily optimal solution. According to Russo and Dosher (1983), they balance cognitive efforts and accuracy by lowering their efforts to a degree where they believe it causes a relatively small increase in error.

Survey respondents similarly search for a balance between the effort they need to put into the question processing and the answer accuracy they deem sufficient (Krosnick and Alwin 1987). Krosnick (1991) linked the satisficing principle to Tourangeau's response model. Respondents who are optimizing go through all four stages of the response process carefully and thoroughly. Respondents who are satisficing, on the other hand, perform some cognitive processes superficially or even entirely dismiss them.

Until the respondent performs all stages with somewhat reduced comprehensiveness, they will be able to form an answer that meets their criterion of *sufficient* accuracy. However, once the respondent abandons a larger part of cognitive processes or entirely omits some stages (usually retrieval and judgment), they will give an answer for which they believe to merely *seem reasonable* to the interviewer or researcher. Krosnick labels the two levels of satisficing as weak and strong satisficing, respectively.

A likelihood that the respondent resorts to satisficing is, according to Krosnick (1991), the function of three variables:

- **Task difficulty** inherent in a question or influenced by other factors, like pace of communication between the interviewer and respondent, or distractions from a wider environment of survey situation.
- **Respondent's ability** in the sense of cognitive sophistication[12]. Respondents with greater abilities find required cognitive processes easier to perform. In addition, those who are more informed about the topic of a question and have a well-established opinion on it will need less cognitive effort to process the question.
- **Respondent's motivation** to perform a comprehensive question processing in order to provide an accurate answer.

Krosnick presents the role of task difficulty (d), ability (a), and motivation (m) with the equation for probability of satisficing, $p(s)$:

$$p(s) = \frac{a}{d \cdot m} \qquad \textit{Eq. 2.1}$$

Of course, this equation should be regarded as merely a schematic representation, reminding that the likelihood of satisficing can be decreased or increased by variations of these factors. For example, even if a task is difficult and the respondent's ability is low, their high motivation decreases the likelihood of resorting to satisficing.

A general consequence of satisficing is decreased accuracy of survey data. While the direction of such effects is hard to predict, a number of typical strategies can be used by respondents to reduce cognitive burden. The initial satisficing approach by Krosnick and Alwin (1987) was used to explain response order effects. Later it was extended to account for a variety of other strategies as well (Krosnick 1991):

[12] Krosnick (1991) understands this sophistication primarily as an ability to perform retrieval operations and to integrate the retrieved information into overall judgment. Yet, there is no reason to exclude the potential impact of the respondent's ability on performance of question comprehension and response mapping.

a) **Response order effects** occur when response categories at the beginning or at the end of a list are more likely to be selected. Selection of initial items (primacy effects) usually occur due to the extended cognitive processing of initial items in a confirmatory direction. Selection of last items (recency effects) is usually the result of highest cognitive attention paid to items toward the end of the list.

b) **Acquiescence** is a tendency of respondents to agree with statements in question. As a satisficing strategy, it occurs due to limited cognitive processing of the statement in a confirmatory direction. Acquiescence can also be a result of specific personality traits or deferential behaviour.

c) **Endorsing *status quo***, which is the selection of a response that expresses appropriateness of the existing status of the object in question. Such answers are commonly a consequence of severely limited or completely skipped retrieval and judgment stages of the response process.

d) **Scale non-differentiation** is manifested as a provision of the same answer to all items of a scale question. A satisficing respondent may select any seemingly reasonable scale value with little cognitive processing and then simply use the same value for other items, regardless of their content.

e) **Non-substantive responses**, like “don’t know”, are a result of satisficing when respondents select them without performing required retrieval or judgment processes.

f) **Random responding**, or mental coin-flipping, is the satisficing strategy with the lowest amount of needed cognitive effort as a respondent simply picks one of the answer categories at random.

Respondents relying on the first two strategies need to perform at least some cognitive processing across all stages of the response process. Response order effects

and acquiescence therefore result in weak satisficing[13]. The remaining strategies are based on the complete elimination of some cognitive stages and therefore result in strong satisficing.

2.2.2 Deviations due to sensitive questions

Specific forms of response process deviations often occur when questions are sensitive. Surveys are delving into more and more sensitive topics that the respondent may feel uncomfortable to answer (Tourangeau et al. 2000). Common examples include questions about income, sexual lifestyle, health issues, use of illegal drugs and alcohol consumption, study grades, and many others (Bradburn et al. 1978; Kreuter et al. 2008).

Although there is no universal agreement on what a sensitive question is, Tourangeau et al. (2000) distinguish three aspects of question sensitivity (see also Tourangeau and Yan 2007). First, the question can be **intrusive** in the sense of invasion into the respondent's privacy. Respondents regard topics of intrusive questions as inappropriate for everyday conversation with a stranger or even a casual acquaintance. Second, respondents may perceive questions as sensitive due to the **fear of disclosure** of answers to third-party individuals or agencies, like governmental organizations and employers. Finally, questions can be sensitive from the perspective of **social desirability** if some of possible answers are more acceptable than others according to social norms that are important to the respondent.

Social desirability is the most complex aspect of question sensitivity. Nederhof (1985, 264) describes it as a reflection of "the tendency on behalf of the subjects to deny socially undesirable traits and to claim socially desirable ones, and the tendency to say things which place the speaker in a favourable light". Paulhus (2002) further explains

[13] Response order effects are sometimes regarded as strong satisficing because the first response category is often selected without performing any cognitive processing at all. We believe that this is not in line with the reasoning by Krosnick and Alwin (1987) who understand response order effects as the selection of the first *reasonably accurate* option. Among Krosnick's satisficing strategies, a selection of the first item, regardless of its content, would best fit into the random responding.

that respondents may distort answers either due to purposive impression management or because of unrealistic self-deception. Social desirability can result in the under-reporting of undesirable behaviours or over-reporting of desirable ones, if the respondent does not rather decline to answer the question (Bradburn et al. 1978). Deviations in answer due to social desirability are the result of over-editing answers in the reporting stage of the response process (Cannell et al. 1981; Tourangeau et al. 2000). Holtgraves (2004) empirically indicated this editing effect by observing longer response times when social desirability was presumably affecting the response process.

The level of distorted reporting importantly depends on the respondent's characteristics. While some topics are intrusive by nature (e.g. income questions), others may be perceived sensitive only by respondents who treat their specific characteristics as undesirable from the perspective of formal or informal social norms (Bradburn et al. 1978). Furthermore, some respondents are more inclined to socially desirable responding because of their personality characteristics, like conformity or the need for social approval (Tourangeau et al. 2000). What is deemed to be a desirable response to a specific question will thus most likely vary across different respondents, their social environment (Näher and Krumpal 2012), and even contextual factors of surveying situation. This is illustrated by Schwarz and Oyserman (2001, 152):

> *Whereas admitting that one has tried drugs may seem threatening to some teenagers when interviewed by an adult, admitting that one has never tried drugs may seem as threatening to some teens when interviewed by a peer.*

According to Cannell et al. (1981) social desirability in surveys is thus a product of perceived appropriateness of the answer in a specific context of survey situation.

2.3 Extensions to survey administration

Schwarz (2007) describes an interviewer-administered survey as a collaboration between the interviewer and the respondent. Self-administered modes, on the other hand,

rely on a direct interaction between the respondent and the survey questionnaire. A systematic consideration of these different forms of interactions is largely missing in the models discussed above, although their elaborations consider some interviewer- or questionnaire-related contextual factors of survey administration. Some authors proposed extended models to incorporate these aspects into the response process more explicitly. Here we reviewed three of these models. While we are predominantly interested in self-administration as an inherent mode characteristic of web surveys, a brief consideration of interviewer-administered response process offered by two of these models is beneficial to highlight specifics of web surveys against other modes.

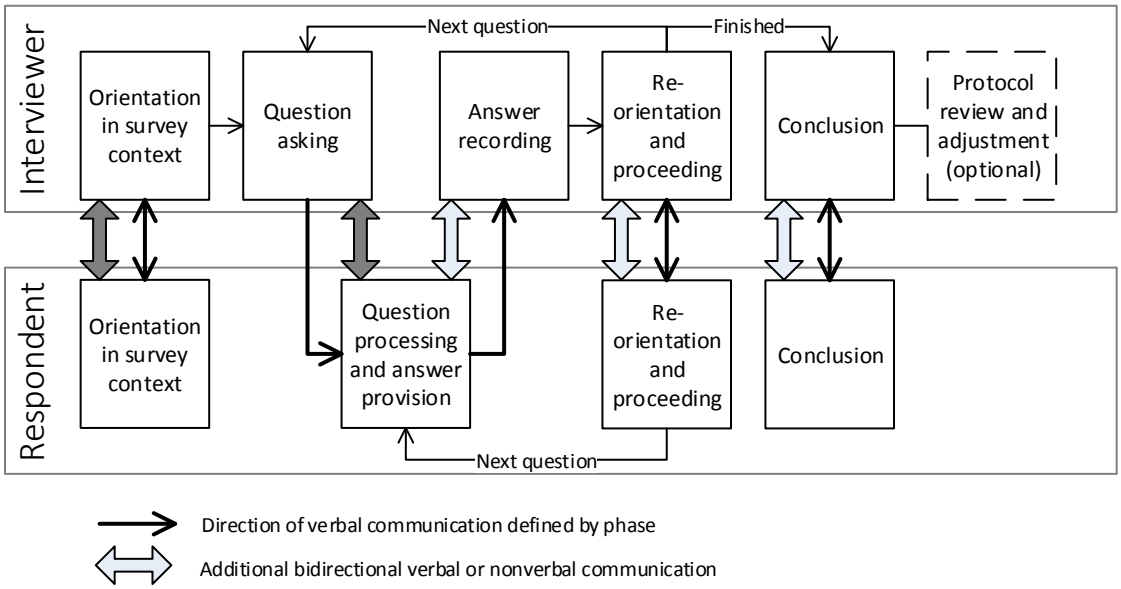
2.3.1 Models of interviewer-respondent interaction

Esposito and Jobe (1991) proposed several extensions to the models discussed above to explain the importance of various survey contexts in which the response process takes place. These contexts include a range of variables related to interview setting, timing of contact, respondent and interviewer characteristics, survey publicity method, survey sponsorship, data collection mode, degree of privacy, attributes of survey instrument, and incentives. The authors most comprehensively elaborate the interaction between interviewers and respondents.

Their survey interaction model specifies the following sequence of the response process:

1. **Interviewer and respondent orient themselves within the survey context** by establishing a preliminary interaction before the surveying begins. An example of a task in this phase is the presentation of a purpose of data collection.
2. **Interviewer asks a question** by reading an item on the questionnaire. Variables affecting this stage include question wording, item characteristics, questionnaire-provided context, and interviewer's reading style.
3. **Respondent processes question and provides answer** – the respondent performs the question answering process which is covered by the information processing models presented above.

Figure 2.2: Schematic representation of the survey interaction model (Esposito and Jobe 1991)



Note: Adapted from Esposito and Jobe (1991).

4. **Interviewer processes and records respondent's answer** by performing the response-categorization process to appropriately interpret the respondent's answer and categorize it in line with expected procedures. The response-categorization processes are similar to information processing on the respondent's side and include comprehension of the response, its cognitive processing, evaluation of appropriateness, and response categorization.
5. **Interviewer and respondent reorient themselves and proceed to next question**, returning to the phase 2. However, both parties in the interaction process are affected to a certain degree by the context established through previous questions.
6. **Interview is concluded.**
7. **Interviewer optionally reviews and adjusts the questionnaire protocol** by performing tasks such as post data collection coding of open-ended answers or filling-in incomplete answers on the basic of inferences.

Figure 2.2 gives an adapted schematic representation of the survey interaction model. Esposito and Jobe (1991) acknowledge that directed verbal communication defined by

individual phases (black arrows in the figure) is in practice always accompanied by continuous bidirectional verbal communication. This occurs, for example, when the respondent asks for clarification or the interviewer uses additional probes. With face-to-face modes, there is additional nonverbal communication between both interacting parties.

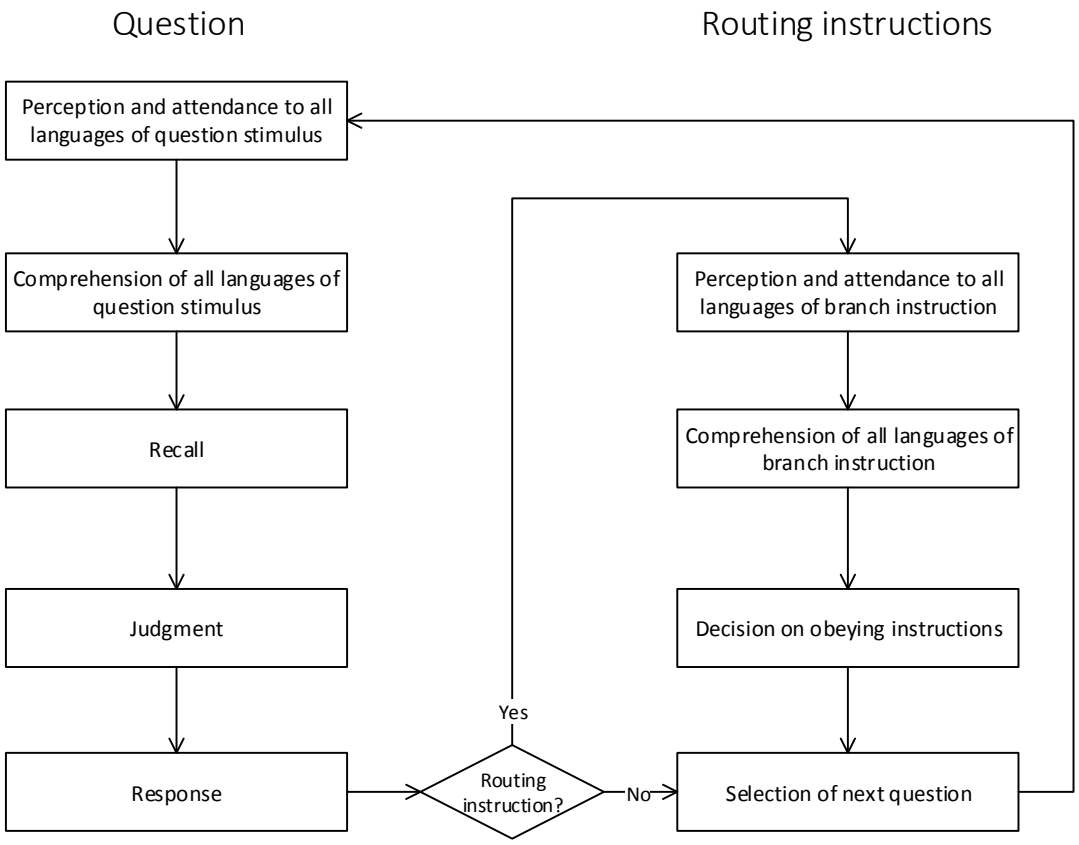
Another approach, the information exchange theory of standardized interviewing (Sander et al. 1992), embraces four cognitive models: the interviewer's question generation and question clarification models, the respondent's question answering model, and the model of interviewer-respondent interaction. Both interviewer's models consist of complex pathways from the interviewer's reading of the question to the actual presentation of the question or its clarification to the respondent.

The question generation model focuses on the interviewer's decisions about how to present the question, either by following the exact wording or paraphrasing it. Paraphrasing occurs when the interviewer decides not to commit to verbatim speech and formulates the question phrasing on the basis of interpreted meaning instead of the exact wording. The process of question clarification is more complex and requires the interviewer to actively monitor the respondent's words, gestures, eye movements and other body language that may indicate question comprehension problems. Furthermore, they have to evaluate the meaning of the respondent's answer and its appropriateness to the question. In case of identified problems, the interviewer needs to formulate clarification probes. The key added value of the information exchange theory of standardized interviewing is thus its contribution to understanding of the complexity of the surveying situation from the interviewer's perspective.

2.3.2 The model of respondent–questionnaire interaction in visual self-administered surveys

The process of answering individual survey questions in self-administered modes is well-covered by general information processing models for respondents. However, they do

Figure 2.3: Schematic presentation of the response model for self-administered surveys by Redline and Dillman (2001)



Note: Adapted from Redline and Dillman (2001). Some paths were excluded from the original figure.

not specifically address additional cognitive processes that respondents perform to accomplish specific tasks of self-administration. In visually presented self-administered questionnaires[14], respondents must perceive the information by themselves, comprehend layout and design of the questionnaire in addition to question wordings, and process guiding directions (Jenkins and Dillman 1997).

Redline and Dillman (2001) devote attention to mental processes required for routing through the questionnaire. They emphasize the questionnaire's role in providing appropriate verbal language, non-verbal language and graphic paralanguage to guide the

[14] Although the title of work by Jenkins and Dillman (1997) *Towards a theory of self-administered questionnaire design* suggests focus on self-administration, it mainly address visual presentation of questions. It does not apply to auditory self-administered modes, like IVR surveys.

respondent through the questionnaire. The respondent's task is then to appropriately identify and follow this guidance (Figure 2.3).

To account for this interaction between the questionnaire and the respondent, Redline and Dillman (2001) add three key additions to the information processing components of the model by Tourangeau et al. (2000): 1) attendance to the question stimulus as the first step of the respondent's question processing, 2) an explicit notion of the role of paralinguistic, symbolic, and numeric language elements in the comprehension stage, and 3) a sequence of stages related to processing of routing instructions.

The model has some important implications for understanding mode effects in web surveys. It shows the increased number of tasks to be performed by the respondent in order to successfully route through the questionnaire and process applicable questions. It also explicitly notes possible intentional deviations from the expected path due to the respondent's control over decisions whether to obey or disobey questionnaire instructions. However, an appropriately implemented web survey questionnaire can substantially relieve the burden imposed on the respondent by providing automated routing, clear visual guidance, and many other assistive elements.

In sum, the survey response process, addressed by various models presented in this chapter, importantly determines the accuracy of answers to survey questions. Deviations in the response process can prevent the respondent to produce an accurate answer and result in the measurement error. Review of the models in this chapter show that inherent mode characteristics, particularly a degree of interviewer's involvement and closeness of interaction with the respondent, introduce some important specifics in the cognitive tasks that the respondent is expected to perform. Additional implementation-specific and contextual mode characteristics can further affect the respondent's performance of the response process. In the next chapter we elaborate a wide range of such influences that can stimulate or supress the occurrence of mode-specific measurement errors in web surveys.

Chapter 3
Factors of mode effects in web surveys

In the previous two chapters we have established conceptual grounds for the elaboration of mode effects in web surveys. In this chapter we explore how inherent characteristics of the web mode – a visual channel of question presentation, self-administration, respondent's use of computer technology during data collection, electronic response channel, and web as information transmission medium – influence response process and measurement error in web surveys.

Discussions in this chapter are guided by the thesis that mode effects are caused through complex relations between inherent mode characteristics and other factors. The key conceptual question to answer in order to identify factors of mode effects on survey error is whether an observed measurement error would be eliminated or reduced if another mode with different inherent characteristics was used. It is therefore necessary to take into account a variety of factors contributing to emergence of mode effects through their relations to the inherent characteristics of the web mode. For example, not each question nor each respondent contribute to mode-specific measurement error in the same way.

Following the distinction of inherent mode characteristics from implementation-specific and contextual dimensions of a survey mode in section 1.1.2, mode-related properties also depend on how the survey is implemented and in what context it is administered. In this chapter we further elaborate and apply this scheme to web surveys. The role of distinctions between fixed and variable dimensions of the mode is further strengthened by the high flexibility of web questionnaires, which can be programmed and presented to the respondent in many different ways and may introduce a high level of variation in the surveying experience.

We explore factors of mode effects by reviewing existing mode comparison studies and experimental manipulations of implementation-specific characteristics of the web mode. Because mode characteristics are often in close interrelation and jointly impact the response process, the discussion in this chapter is not structured strictly by the inherent dimensions of the web mode. We also do not separately address interviewer involvement and closeness of interaction with the respondent as the distinction does not apply to the self-administered nature of web surveys.

3.1 Visual presentation of the questionnaire

We begin exploring the factors of mode effects in web surveys by discussing the role of visually presented questions. We defined web surveys as being inherently a visual mode (Figure 1.2 on page 40), although they can be extended by auditory communication in the form of audio clips or other multimedia elements. The effects of visual presentation on the response process are, of course, predominantly related to the visual input channel. At the same time they are also indispensably influenced by computerization and self-administration. Many features of visual presentation essentially rely on computer technology, therefore analysis of the visual channel of presentation in web surveys cannot be separated from the computerization of the questionnaire. In addition, the visual input channel in web surveys does not only serve to deliver questions. The absence of interviewers, which we more comprehensively cover in section 3.2, requires visual presentation to take over the role of guiding respondents through the questionnaire (Smyth et al. 2006a). Several mode effects may therefore stem from interactions between visual presentation, self-administration, and computerization.

The auditory and visual input channels significantly differ in the way of information transmission. Words and numbers, as the essential component of both, are perceived aurally in the former and visually in the latter. Differences between spoken and written language determine the types of communication available for transmission through

each of the two channels. The auditory channel extends verbal language with paralinguistic communication, such as emotional tone, emphasis, and timing; with physical presence of actors, like in face-to-face interviewing, an additional non-verbal communication is available (de Leeuw 1992). Visual modes rely on other communication languages: words and numbers, non-verbal symbols, and graphic paralanguage. In their presentation of these languages, Redline and Dillman (2001) treat the graphic paralanguage as the fundamental precondition of visual perception, essential for transmission of other languages through the visual channel. The model proposed by the authors requires respondents to attend to all three languages before proceeding with the question comprehension stage.

The combination of visual input channel and the respondent's use of computer technology establishes an environment of numerous possible alternative questionnaire presentations. The interest in visual features is especially large in web surveys, because their inclusion is simple, straightforward, and available at largely no additional costs – but at the same time very important for data quality. Each of these possibilities, at least to some degree, influence the type and amount of information transmitted through the visual channel.

As with other mode characteristics, the key conceptual question in studying mode effects attributable to the visual input channel is whether or not a specific factor causes measurement errors that would be absent or reduced if the auditory presentation was used. This is somewhat difficult to answer because each type of the channel is based on a very specific set of communication types. For example, a measurement error arising from a specific implementation of graphic paralanguage cannot occur in auditory modes where this paralanguage does not exist. Such an error is therefore attributed, by definition, to the mode effect. To understand the role of the visual input channel in the emergence of mode effects it is therefore necessary to understand how different visual implementations of the questionnaire contribute to measurement errors. In other words, the key issue for mode effects is not the inherent visual presentation, but specific implementation of this presentation (Tourangeau et al. 2000).

3.1.1 General design principles: visibility and clarity of presentation

There are several similarities between the practice of web questionnaire design and human-computer interaction studies (Kaczmirek 2009). The most important underlying principle of visual input channel is *visibility*. Conrad et al. (2006) draw parallels to web usability studies according to which users are less likely to search for information if this requires them to invest even minor additional effort. Translated to a surveying situation and following the Krosnick's equation for the probability of satisficing (Eq. 2.1, page 73), less visible information increase the burden of respondents and produce more satisficing behaviours. Respondents therefore expectedly pay more attention and processing effort when information is visible immediately, without requiring further actions like mouse clicks, scrolling, or even eye movements (Tourangeau et al. 2013). Even the basic elements of questionnaire design can thus importantly increase or decrease survey measurement errors.

Visibility positions the visual design and layout of a questionnaire into the foundation of the response process in visual modes. The successful transmission of information from the questionnaire to the respondent critically depends on the proper visual perception of the stimuli. Jenkins and Dillman (1997) use a large body of knowledge on cognition and visual perception to derive principles of assuring clarity of presentation, unambiguous identification of individual questionnaire elements (such as question text, response options, additional instructions, and input fields), organization of information, and so on. Later work by Dillman et al. (2008) and by Couper (2008) extended these principles into a comprehensive set of guidelines and considerations for appropriate visual design. Consistency of visual presentation, visually distinguished questionnaire elements, contrasting text colour against the background, easily readable typography, careful use of emphasis, and uncluttered pages are only some ways of ensuring standardized visual

stimuli across respondents. Although we do not elaborate on these general aspects of the questionnaire design[15], their contribution to mode effects can be important.

An increase in mode effects caused by a specific web questionnaire design is demonstrated by studies comparing different distribution of questions across pages. Mail and other paper-based questionnaires always include several questions on one page for practical reasons. Web surveys allow for much higher flexibility as the computerization enables the complete customization of a number of questions per page without added costs. Couper (2008) offers a detailed review of advantages and disadvantages of different layouts, but here we focus only on the clarity of presentation. When all questions are presented on the same page, visual clutter can lead to accidentally skipped questions (Dillman et al. 2008) or heighten the likelihood of satisficing due to increased task difficulty. Both these factors help explaining higher item nonresponse found in single-page presentation of questions compared to the layout with one question per page (Lozar Manfreda et al. 2002; Peytchev et al. 2006; Toepoel et al. 2009b). Mode effects occur here because the reduced clarity of presentation prevents the adequate transmission of information through the visual channel, but does not apply to the auditory question presentation channel. This can be approximately paralleled with measurement errors due to unclear interviewer's speech in auditory modes.

3.1.2 Question wording and format

Question wording is the main tool for transmitting the request for information to respondents. It consists of a statement that specifies the question focus, response categories, and additional instructions. All these elements are used by the respondent to construct their representation about the question (Schwarz and Oyserman 2001).

Modes with different question presentation channels require different, although importantly related skills from respondents in order to successfully comprehend the

[15] Couper (2008) gives an extensive overview of very detailed visual features of web surveys and their implications for data quality. Informative insight into the application of principles of human-computer interaction for web survey design is also offered by Kaczmirek (2009).

question (Perfetti et al. 2005). The auditory input poses a larger demand on the respondent's working memory to memorize the question during information processing. The respondent needs to perform a dual task of simultaneously keeping the interpretation of the question in their memory and formulating an answer (Burton and Blair 1991). Memory demands are substantially relieved with visual presentation, since respondents can recheck the question wording any time. However, they need to possess sufficient reading skills for its comprehension (Tourangeau et al. 2000).

Complexity of wording

In line with Grice's maxim of clarity, respondents expect that the question is asked in an obvious way (Schwarz and Oyserman 2001). However, complex wordings are sometimes required to sufficiently specify the question's focus. Longer questions require a respondent to keep more information in their working memory, potentially overloading it and resulting in terminated or slowed-down cognitive processing (Tourangeau and Smith 1996; Tourangeau et al. 2000). Comprehension difficulties further arise with factors such as technical, vague, or rarely used terms, low predictability of grammatical structure, complex syntax, and required inferences (Graesser et al. 2006; Lenzner et al. 2010)[16].

As the question complexity increases, the task of comprehension becomes more burdensome and may result in the respondent's objective inability to comprehend the question, or encourage a reliance on satisficing strategies to reduce the cognitive burden (Krosnick 1991). Studies report on a number of significant effects of complex question wordings in web surveys, including prolonged response times (Yan and Tourangeau 2008; Lenzner et al. 2010), increased drop-outs (found by Ganassali 2008, but not by Lenzner et al. 2010), and the tendency to select middle values (Lenzner et al. 2010). Similar effects were found in other modes: higher endorsement of midpoint scale values (Velez and Ashworth 2007), decreased reliability (Saris and Gallhofer 2007), higher reliance on interpretative cues provided by response categories (Bless et al.

[16] Lenzner et al. (2010) point out that question complexity is not uniformly defined. With the term "complex question" we thus very generally refer to a question with some of linguistic characteristics listed here.

1992), more non-substantive responses (Knäuper et al. 1997), and lower accuracy of answers (Schober and Conrad 1997).

Expected mode differences in question complexity, while lacking sufficient empirical support, can be partially derived from research in cognitive psychology that confirms higher memory demands of auditory processing[17]. As the eye-tracking study of visually presented questions by Lenzner et al. (2011) show, respondents commonly reread complex questions and pause at difficult words for further processing. Although in interviewer-administered auditory modes respondents can ask the interviewer to repeat the question or slow down the reading pace, they rarely make such requests (Conrad and Schober 2000; Dijkstra and Ongena 2006). We can thus expect web surveys to be generally less prone to negative effects of complex question wordings than auditory modes. The same expectation can be applied to wordings of additional instructions and clarifications of unknown or ambiguous concepts.

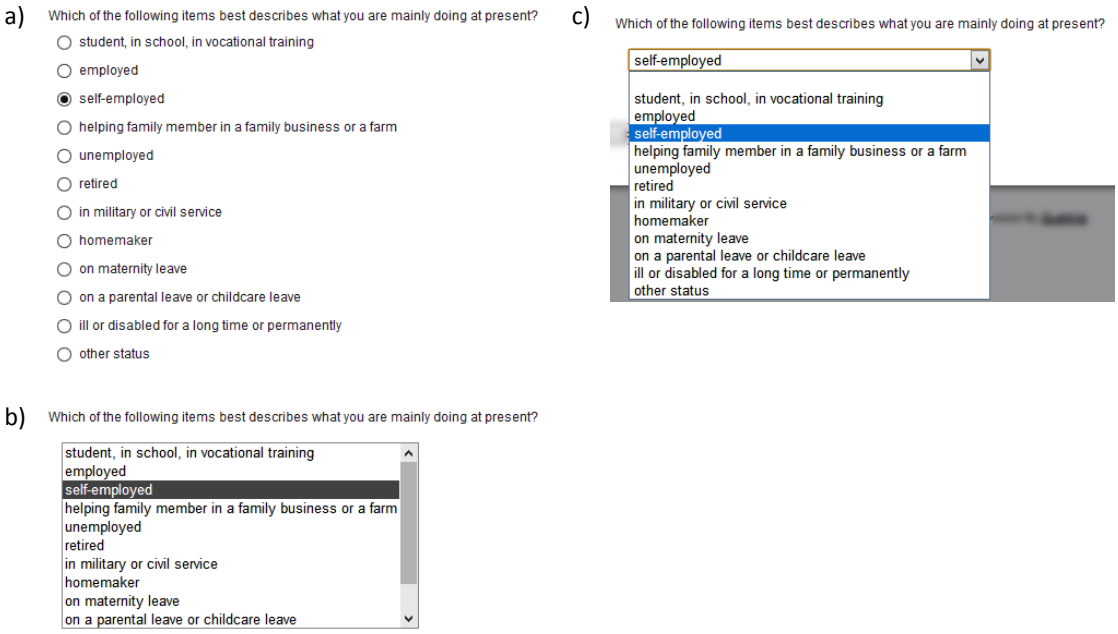
However, comprehension difficulties and corresponding mode differences are importantly moderated by the respondent's motivation, cognitive performance and reading ability (Knäuper et al. 1997; Yan and Tourangeau 2008; Chang and Krosnick 2010). Under certain circumstances, we can therefore expect a visual channel of presentation, combined with other factors, to increase mode effects related to complex question wordings in web surveys. Especially respondents with limited reading skills may find the auditory presentation easier and would be subjected to higher cognitive burden when they have to read complex wordings (Dillman 1991). If the burden exceeds their reading capacity or the amount of effort they are ready to invest into the question comprehension, a resulting measurement error would be attributable to mode effect.

Question formats

Computerization enables web surveys to utilize a wide range of question types and formats with high flexibility. Although some formats are often considered as

[17] A general overview is given, for example, by Eysenck and Keane (2010).

Figure 3.1: Examples of common interface control elements in web surveys to present close-ended single-answer questions: a) radio buttons, b) list box, and c) drop-down menu



Note: Questions adapted from Aasve et al. (2011).

interchangeable, alternative presentations can produce significant differences in responses. Mode effects can be most clearly related to formats that cause measurement errors because they violate the visibility principle.

This can be well illustrated by the most basic single-answer type of question (Figure 3.1). Web surveys offer several alternative user interface elements to format such questions: radio buttons, list boxes, drop-down menus, and a variety of visually customized fields. These various formats were found to significantly impact the response process. List boxes and drop-down menus restrict the amount of information (response options) available to the respondent without further action (clicking and scrolling). According to observations by Galesic et al. (2008), and in line with the visibility principle, immediately visible response options receive higher attention in term of fixation times than options visible only after scrolling. Unsurprisingly, response order effects have been shown to increase with these question formats compared to radio buttons, due to higher task difficulty or the lacking motivation of respondents to access the remaining response

options (Couper et al. 2004a; Galesic et al. 2008). Such restriction of accessibility of response options is specific to visual computerized modes and cannot occur with the aural presentation of questions neither with paper questionnaires.

Despite the potentially damaging impact of web-specific question formats on the quality of obtained responses, their use can be beneficial in some situations. Drop-down menus, for example, allow for the use of close-ended questions with a large amount of response categories without extending the length of the questionnaire page (Couper 2008). Compared to paper-based and auditory modes, where such questions can only be asked in an open-ended form, this enables the higher standardization of answers and assures more successful mapping of responses to the expected content (Dillman and Christian 2005).

Effects of many other alternative question formats and layouts available in web surveys are less well understood, especially in comparison to other modes. Reviews by Couper (2008) and Tourangeau et al. (2013) suggest that many alternatives, like visual analogue scales and customized graphical input fields, at best do not provide any considerable advantage in terms of data quality. On the other hand, minor enhancements in question presentation, such as the addition of labels next to answer spaces, can improve the accuracy of the information provided (Christian et al. 2007b; Couper et al. 2011). In this way, a more explicit transmission of information through the visual channel can be beneficial over the auditory modes.

3.1.3 Questionnaire-provided context

Like any other form of conversation, surveying takes place in a context. Survey context is somewhat broadly defined and can include a large number of questionnaire-related, respondent-related, and environmental factors (Uhan 1998; Tourangeau et al. 2000; Smyth et al. 2009a). The term is most commonly used to describe effects arising from the order of the questions in the questionnaire, i.e. question order effects. Yet, many other important questionnaire features can provide adverse interpretational cues which

respondents might include into the response process as acknowledged in the model by Cannell et al. (1981). This is even more emphasized in web surveys with extended possibilities of using a whole range of graphical elements.

What we call *questionnaire-provided context* embraces a much larger set of contextual factors than only question order, but still limits the discussion to those based on a specific (visual) question presentation channel of web surveys.

Assimilation, contrast, and question order effects

First observations of question order effects were published early on, for example in the works by Sayre (1939) and Cantril (1944). In their evaluation of the problem, Schuman and Presser (1981) distinguished between consistency and contrast effects, later labelled as carryover and backfire effects by Tourangeau and Rasinski (1988), and now commonly known as assimilation and contrast effects, respectively[18]. **Assimilation effects** occur when the respondent includes information from one or more of preceding items into the formation of a response to the target item, resulting in the judgment being consistent with the primed question or category. **Contrast effects** are the result of a standard of comparison established on the basis of preceding items, leading the respondent to judge the target item by contrasting it to the previous ones. They may also occur because the respondent tries to keep the conversation in line with Grice's maxims of increasing informative value by avoiding redundancy (Strack et al. 1988).

According to Tourangeau and Rasinski (1988; also in Tourangeau et al. 2000), assimilation and contrast effects can affect each stage of the survey response process by 1) providing specific guidance for interpretation, 2) affecting the accessibility of relevant considerations for retrieval, 3) altering the establishment of judgment criteria, or 4) changing the way in which the answers are mapped to the response categories and reported. Table 3.1 briefly summarizes these effects on the basis of various literature (Herr

[18] As we shall see later, assimilation and contrast effects can arise not only from the question order but also from other questionnaire features.

Table 3.1: Key variables influencing assimilation and contrast effects in each stage of the response process

Response process stage	Assimilation	Contrast
Comprehension Context enters the respondent's interpretation *about* the question.	Preceding items provide the interpretational cues for the target item. Increased if the target item is unfamiliar or ambiguous, attitudes are inaccessible, or the questionnaire design implies the relation between questions.	Preceding items cause the exclusion of information from interpretation of the target item. Increased when items are perceived as repeating or undefined in scope.
Retrieval Context affects which information is retrieved and decided to be used in the response process.	Preceding items increase the probability of retrieving specific information in processing of the target item. Increased if topics of items are perceived as related, are administered in close temporal succession, or appear related by the questionnaire design.	Preceding items trigger exclusion (subtraction) of relevant material. Increased if respondents are aware of priming items due to obvious context, they perceive the item as unrepresentative, biased, or repetitive.
Judgment Context guides establishment of dimensions and criteria of comparison used to evaluate the retrieved material and form a judgment.	Preceding items cause implication of a norm to the target item, such as typical value or even-handedness.	Previous items or own previous experiences establish an extreme standard of comparison (anchor) for the target item. Increased when items relate to the same dimension and the context item is perceived as an extreme case, or when comparative judgments are required.
Response Context influences mapping of responses to the question scale or editing the chosen response.	Preceding items impose the pressure for a consistent response to the target item. Increased when items appear related.	Anchors from preceding questions alter mapping of the target response. Increased when items are judged on the same dimension.

et al. 1983; Tourangeau and Rasinski 1988; Strack et al. 1991; Judd et al. 1991; Tourangeau et al. 2000; Dillman et al. 2008).

Extremity and amount of contextual information increase the influential potential of these effects on the processing of subsequent items (Schwarz and Bless 2007). The context established by previous questions and answers is significantly more explicit when a questionnaire is presented visually (Dillman and Christian 2005; Smyth et al. 2009a) and depends less on the working memory capacity of respondents than in auditory modes.

More explicit questionnaire-provided context of visual modes may lead us to expect more pronounced assimilation and contrast effects in web surveys than in auditory

modes. This reasoning has an important caveat. Although empirical research on this topic is lacking for web surveys, several studies comparing telephone and mail surveys consistently show lower or even absent effects of preceding questions in the mail mode (Bishop et al. 1988; Schwarz et al. 1991; Schwarz and Hippler 1995; Lorenz and Ryan 1996). The hypothesized reason lies in the interrelation between the channel of question presentation and the implications of self-administration, particularly the locus of control discussed later in this chapter (section 3.2.1, page 101). This allows respondents to see several questions at a time, move back and forth in the questionnaire, and consider the context more deeply because they are under a lower time pressure (Schwarz et al. 1991). This does not eliminate the effects of surrounding questions, but makes them less conditioned to the order. Respondents are more likely to interpret the target question on the basis of a broader context of preceding *and subsequent* questions, potentially resulting in an increased consistency due to perceived relations between them (Schwarz et al. 1991; Schwarz and Hippler 1995) [19].

Interpretational cues provided by the questionnaire layout and design

Previously, we discussed the importance of question layout and design from the perspective of visibility and accessibility of information. In addition to this, specific layout and design of questions can convey additional interpretative cues to respondents. Tourangeau et al. (2004) list five interpretative heuristics that respondents might use when responding to scale questions in a web questionnaire:

- **middle means central**: the middle item on a list is regarded as the central value and establishes the meaning of other values;
- **left and top means first**: such item is regarded as the first in some conceptual sense[20];

[19] These effects should not be understood simply as a form of satisficing. While the satisficing behaviour may partially contribute to their occurrence, Tourangeau and Rasinski (1988) expect contrast effects to occur even more frequently among thoughtful respondents who devote more time to making a decision about the relevance of a retrieved material.

[20] It is important to bear in mind that the second and the fourth heuristic on this list may be strongly culturally dependent and applicable mostly to Western languages because of left-to-right and top-to-bottom reading (Tourangeau et al. 2004).

- **near means related**: items appearing near each other are treated as conceptually related;
- **up means good**: the top item in a vertically oriented list is perceived as the most desirable one, especially when the content of items explicitly vary in desirability;
- **like means close**: visually similar options are regarded as similar in content.

The incorporation of these heuristics into the response process guides the respondent's expectations according to which they interpret the questions.

Question layout

Tourangeau et al. (2004) conducted a series of experiments on scale questions to verify how their visual presentation affect responses. They manipulated visual presentation by reducing the spacing between scale points on the left side of the scale, thus shifting a visual position of the middle category to the left. This manipulation increased the mean values by heightened the probability of selecting *apparently* central values, although they actually belonged to the right side of the scale. A similar shift in the perceived scale centre occurred when non-substantive responses "don't know" and "no opinion" were added without a visual separation from substantive values. The authors conclude that visual presentation plays an important role in visualizing the middle scale point which, according to the "middle means central" heuristic, also determines the respondent's interpretation of other scale values.

With the "left and top means first" heuristic principle respondents expect the response options to be presented in a logical order. When a specific visual layout violates this expectation, respondents have to devote further cognitive effort to resolve the inconsistency. Experimental studies show significant changes in response distributions and increased response times if scale points are not presented linearly across multiple rows and columns or their order is inconsistent (Tourangeau et al. 2004; Christian and Dillman 2004; Toepoel et al. 2009a).

Evidence of a different type of interpretational cue derived from the question format comes from research on the size of answer spaces in open-ended questions. For mail

surveys, larger answer spaces were shown to result in longer descriptive responses (Smith 1995; Christian and Dillman 2004), and results were, although to a lesser degree, replicated in the web mode (Smyth et al. 2009b). While availability of such additional contextual information may often be beneficial to respondents, unsuitably sized boxes can lead to interpretational difficulties and errors. For example, unnecessarily large input fields for numeric answers were found to increase the proportion of inappropriately formatted answers, like textual instead of a numeric input (Couper et al. 2001; Fuchs 2009). Although later evidence on this was mixed (Couper et al. 2011), the results caution against the inclusion of contextual information (size of the input box) that conflicts the question instructions. In auditory modes such contextual indication of the expected response length is, of course, completely absent.

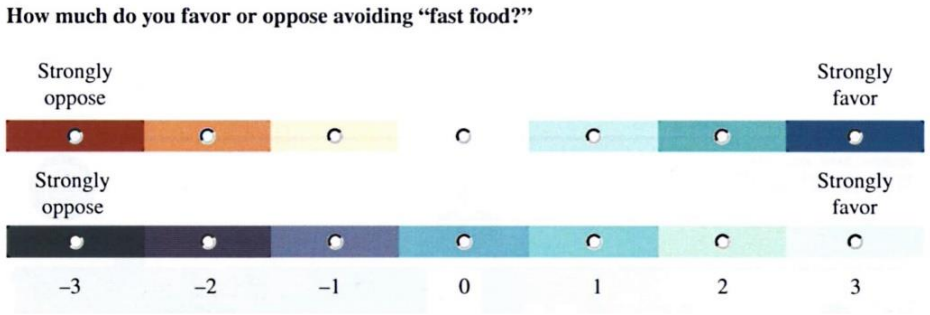
Visibility of other questions

Accessibility of contextual information implied by other questions is affected by the different distributions of questions across pages. This allows bringing the presentation of a questionnaire closer to mail surveys (several questions per page) or telephone interviewing (one question per page). Unfortunately, there is no simultaneous comparisons of telephone and alternative web designs to explain how this changes between-mode differences in estimates.

Inclusion of several questions on a single page increases the availability of contextual information and thus more likely produces context effects (Schwarz et al. 1991). Studies by Couper et al. (2001) and Tourangeau et al. (2004) compared correlations between items presented on a single page and the same items distributed across several pages. Consistent with the “near means related” heuristic, correlations were higher in the single-page condition. The authors of the latter study do not regard this as a necessarily desirable result, since increased correlations may well be a sign of lower differentiation between items due to satisficing.

The presentation of questions on a single page also gives respondents immediate information about the questionnaire length (Couper 2008), similar to mail surveys but

Figure 3.2: Different colours for different scale values used by Tourangeau et al. (2007)



Note: Reproduced from Tourangeau et al. (2007). Colours may deviate from the original.

contrasting auditory modes. While certainly positive from the perspective of a respondent's informed decision about survey participation, this may lead to a higher drop out, stimulate satisficing, and also introduce other sources of measurement errors. For example, some respondents may start searching for the fastest route through the questionnaire by avoiding answers that conditionally display additional questions. Further research is needed to empirically verify this form of shortcutting, although Peytchev et al. (2006) found some effects in this direction.

Colours and graphics

Colours are a powerful tool of visual modes to attract the attention of respondents to specific elements and can be effectively used to guide them through the survey task. However, special meanings associated to specific colours may alter the response process, elicit specific emotional responses, and vary across respondents and cultures (Couper 2008).

How colours can govern the respondent's processing of questions is illustrated by Tourangeau et al. (2007). The authors compared responses to two differently coloured scales (Figure 3.2). Each side of the first scale was presented in different colours, with stronger shades representing more extreme values. For the second scale, varying shades of single colour were used. The former condition produced significantly higher (more positive) responses with a substantially lower proportion of midpoint answers. These differences confirm that respondents treat visually similar objects as similar in meaning

(the “like means close” heuristic), where the interpreted distance between objects is increased if they are presented using different colours. The effect, however, disappeared if every scale point was verbally labelled, suggesting the existence of a hierarchy of contextual interpretive cues in which verbal labels seemingly take precedence over colours (Tourangeau et al. 2013).

The potential effects of colours are a strong argument against their use for merely decorative purposes, which extends to other graphical elements as well. Smith (1995) reported an interesting example from a visual presentation of scales on showcards in the ISSP 1987 survey. In the majority of countries the scale on social stratification was presented in the form of a ladder consisting of equal-sized stacked boxes. One of the countries changed the design to reflect a truncated pyramid with wider boxes at the bottom. This substantially increased the respondents’ perception of the typical population value of the observed phenomenon, leading to a higher endorsement of bottom categories.

Web surveys brought an increased use of more sophisticated graphics, including images. These can present an essential part of the question, for example to measure a customer’s preference of different product packages. On the other hand, they can be used merely as an incidental element with the purpose of making the overall questionnaire design more attractive. According to Couper (2008), images can introduce various contextual effects, including contrasting judgments, primed question interpretation, and judgment formation based on a concrete information conveyed by the image. They may also influence mood or emotional state of the respondent.

Several studies were conducted regarding the influence of images on the response process in web surveys. Couper et al. (2007) found significant contrast effects produced by the inclusion of an image next to a question on self-rated health. Respondents judged their state in contrast to the information provided by the image, rating own health significantly better when presented with the image of a sick woman than with the image of a fit woman jogging. Couper et al. (2004b) similarly observed significantly more shopping reports when respondents were exposed to the image of general shopping

compared to the image of clothes shopping, as well as more eating out reports with the image of a fast food restaurant than with the image of a luxury restaurant. These latter examples indicate assimilation effects, where images establish an interpretational framework within which respondents judge relevance of recalled events.

Other contributing factors

We only reviewed a few typical elements of the visual questionnaire presentation that sometimes impact how information is transferred through the visual input channel in web surveys, how it enters, and potentially alters the response process. There are many other specific factors with potential similar effects. For example, Smyth et al. (2006a) show how visual grouping of response options conveys the sense of higher proximity within groups than between groups. Another example is given by studies suggesting that the persistent visibility of the researcher's affiliation, like when a logo is included on each questionnaire page, may cause respondents to start providing answers in line with their perceived interest of the researcher (Schwarz and Oyserman 2001; Galesic and Tourangeau 2007). All these are enabled or emphasized by the visual channel of presentation and are absent or reduced if an auditory presentation is used.

There is an important deficit of research that would help explain how mode effects due to visual context provision in web surveys are moderated by the characteristics of respondents. Toepoel et al. (2009b) report that verbal, graphical, and numerical language elements exhibit a larger impact on older respondents, suggesting that contextual effects increase with reduced cognitive functioning. Also, Schwarz and Oyserman (2001) claim that the impact of response categories is higher among respondents that cannot retrieve relevant information from memory because of their poor representation or due to lower memory capacity.

It is unlikely that we will soon be able to consider, understand, or at least detect this great variability of different sources that influence the effects of visual questionnaire presentation on the measurement process. It is, however, clear that the visual input channel transmits a significantly higher amount of explicit and implicit information than

auditory modes. Because the role of a respondent's working memory is of a smaller importance, this information is also likely to be more comprehensively included into the response process. This is advantageous when information contributes to a higher accuracy of survey measurement, but results in mode effects when the influence is negative. An implementation-specific presentation of the questionnaire can thus stimulate or reduce measurement errors specific to the visual input channel and related computerized features of web questionnaires.

3.2 Self-administration and computerization of the questionnaire

The second main inherent characteristic of the web mode is self-administration. Absence of interviewers has many important implications discussed in a large body of general survey literature (e.g. Biemer and Lyberg 2003; Weisberg 2005; Groves et al. 2009). It eliminates random and systematic errors caused by interviewers, including the impact of specific interviewer's characteristics and behaviours on responses, variations in individual interviewing techniques, recording errors, and data falsifications. It also decreases the reluctance of respondents to provide answers to questions they find sensitive, and gives them more control over surveying location, pace, and time.

On the other hand, self-administration can have a negative influence on responses in many ways. Possibilities to motivate respondents are severely limited, there are no interviewer-delivered explanations and probes, respondents take all the burden of navigating through the questionnaire, and so on. We deal with these and other issues in this chapter. Furthermore, data quality may be compromised by lack of more general control over the respondent's state. This is vividly illustrated by Stanton (1998, 712): "WWW respondents might be sleepy, angry, bored, intoxicated, or otherwise in an unsuitable frame of mind to provide honest, accurate responses to items".

It would not make much sense to discuss self-administration without considering other closely related inherent characteristics of the web mode: visual presentation, the respondent's use of a computer technology, and the closely related electronic response channel. In the previous chapter we have mentioned the role of visual presentation for guiding respondents through the questionnaire. In web surveys the guidance is further controlled by computer administration. This offers dynamic features for a continuous interaction between the questionnaire and the respondent (Lozar Manfreda and Vehovar 2002a). Interactivity of modern web questionnaires can even resemble several aspects of interviewer-administration – including some of its drawbacks.

3.2.1 Interaction with the questionnaire

The model by Esposito and Jobe (1991), presented in Figure 2.2 (page 78), characterizes the interaction process in interviewer-administered surveys as an ongoing verbal and nonverbal exchange of information between the interviewer and the respondent. The interviewer acts as a proxy for a two-way transmission of information between the questionnaire and the respondent, guides the respondent through the questionnaire, and checks the appropriateness of provided answers.

In self-administered modes, respondents need to perform all these tasks by themselves. A larger burden imposed on them by increased task difficulty increases the importance of their motivation and abilities for the successful conduction of the response process. This direct potential to increase satisficing positions self-administration as the most general source of mode effects among all inherent mode characteristics. On the other hand, self-administration offers respondents a greater flexibility of interaction with the questionnaire.

Locus of control

De Leeuw (1992) uses the psychological concept of *locus of control* to describe differences between modes in the regulation of survey communication flow. According to the

author, the type of interviewer involvement and closeness of interaction with the respondent play a major role in determining the locus of control: in a direct personal interaction of face-to-face interviews the locus of control is shared between the interviewer and the respondent, while in telephone interviewing the interviewer takes over a larger part of the control. These relations between actors are substantially governed by social norms of interpersonal interactions. With self-administration, the respondent assumes locus completely, giving them the freedom to decide when and how to answer the survey, and how quickly to move through the questionnaire (Christian et al. 2007a).

Pace of responding

De Leeuw (1992) pays special attention to advantages of respondent-controlled pace in self-administered questionnaires. Of course, the mode itself does not inherently determine the pace of responding, even when the questionnaire is administered by an interviewer. Normally, it would be highly unusual for the interviewer to demand from their respondents to provide answers swiftly and within strictly bounded time limit. However, pace of communication is influenced by social conventions applied to different survey settings (de Leeuw 1992). Especially in telephone interviewing both, the respondent and the interviewer, may feel uncomfortable with too slow pace of communication caused by silences during cognitive processing of questions (Schwarz et al. 1991; de Leeuw 1992; Dillman et al. 1996). In line with this, respondents to telephone surveys were found to provide more impromptu responses (Dillman 1991; Hippler and Schwarz 1998). In contrast, nonverbal communication in face-to-face interviews helps to bridge the awkward silences (Schwarz et al. 1991), while in web and other self-administered surveys there is no such social pressure at all.

Time pressure especially affects recall and judgment components of the response process (Schwarz et al. 1991). Availability of additional time is generally expected to improve the accuracy of answers as long as respondents use it to perform more thorough cognitive processing. For factual questions additional processing time enable respondents to retrieve more relevant episodes from the long-term memory and reduce misplacement of episodes in time (Burton and Blair 1991; Schwarz and Oyserman 2001).

For attitudinal questions, under the assumptions of the belief-sampling model, respondents are able to retrieve a larger samples of relevant considerations and enhance the stability of reported attitudes (Tourangeau et al. 2000; Schwarz and Bless 2007)[21].

How many respondents utilize this freedom of custom response pace in web surveys and other self-administered modes is difficult to empirically estimate. The comparison of response times between modes is not a sufficient indicator, because times are affected by other inherent mode characteristics, especially the question presentation channel and response channel. Commonly measured faster completion times in web surveys against face-to-face interviews (Heerwegh and Loosveldt 2008), telephone interviews (Chang and Krosnick 2010), and paper-based self-administered questionnaires (Mangunkusumo et al. 2006) can therefore be attributable to different reasons: speed differences between reading and speaking, or lower effort put into the processing of questions.

Some indirect evidence of the beneficial effects of custom response pace in self-administered questionnaires comes from comparisons of open-ended questions. Longer and more comprehensive answers can be generally regarded as an indicator of more thorough question processing. De Leeuw (1992), for example, found somewhat longer answers in the mail mode compared to face-to-face and telephone interviews. The same patterns were observed in web surveys (Shin et al. 2012; Gravlee et al. 2013). Fricker et al. (2005) also report a higher proportion of correct answers and longer response times to open-ended knowledge questions in the web than in the telephone mode. However, the authors allow for an equally probable explanation that differences occur because

[21] The belief-sampling model of answering attitudinal questions follows the position that attitudes are more *constructed on the spot* than *stored in memory* (Bohner and Dickel 2011), although some stable representations may be used to construct the attitude evaluation (Cunningham et al. 2007). During the processing of attitudinal questions, the respondent retrieves and processes only a small part (a sample) of information from a range of relevant beliefs, feelings, impressions, general values and prior judgments about the topic (Tourangeau et al. 2000). This is in sharp contrast to the more traditional view of attitudes as stable and based on the pre-existing availability of attitude-related evaluations (e.g. Wilson and Hodges 1992). Bohner and Dickel (2011) discuss this debate among social psychologists and survey methodologists in more detail.

visual presentation decreases task difficulty of question comprehension, while keyboard-based entry of answers prolongs response times.

The positive potential of respondent's locus of control due to the flexibility of response pace have a powerful negative counterpart: speeding. Absence of interviewers makes it easier for respondents to employ weak or strong satisficing strategies. This allows them to reduce the time spent completing the questionnaire, but to still appear as cooperative respondents. Studies have confirmed relations between short response times and response order effects (Malhotra 2008), non-differentiation (Gutierrez et al. 2011), and other satisficing strategies (Callegaro et al. 2009). A higher likelihood of speeding may help to explain higher occurrences of satisficing patterns in web surveys observed by some studies (e.g. Fricker et al. 2005; Heerwegh and Loosveldt 2008), but current inconclusive research does not allow for the drawing of definitive causal inferences on this. We further elaborate on satisficing and other response patterns in web surveys in section 3.4.

Similar to other issues related to satisficing, the characteristics of respondents and questions can increase or decrease such effects. Speeding in web surveys seem to contribute to measurement error less among respondents with higher cognitive abilities and for cognitively less demanding questions. Research on speeding in web surveys by Malhorta (2008) indicated more pronounced primacy effects among fast responders with lower education, but largely absent effects among higher educated fast responders. In addition, the time needed for cognitive processing depends on the degree of a respondent's existing articulation of the topic in question. Respondents with well-established pre-existing attitudes need to put less time and effort into the response process than those with only some general background idea or no opinion at all (Tourangeau et al. 2000).

Further aspects of locus of control

Locus of control at the respondent's side has a very general influence on the nature of survey interaction in self-administered surveys. The respondent takes over complete control over the questionnaire. They can check questions ahead by moving back and

forth in the questionnaire. The order of question presentation and related context effects may thus vary according to individual's respondent behaviour (Schwarz and Hippler 1995; Smyth et al. 2009a). Furthermore, respondents can easily terminate answering questions at some point and continue later (Couper 2008), or even decide to cruise through the questionnaire without providing any answers at all (Bosnjak et al. 2001). It is also their decision how thorough to read the questions prior to providing an answer, whether or not to obey various instructions, and so on. In sum, a vast majority of topics addressed in this chapter is to some degree influenced by the locus of control.

In web surveys, however, locus of control can be partially restricted by the computerization of the questionnaire (Couper 2011). Depending on the programming of the questionnaire, respondents have a varying degree of freedom in interaction with it. In one example, automated questionnaire routing does not allow respondents to move completely freely through the questionnaire. For this reason we treat locus of control as an implementation-specific characteristic (Table 1.2 on page 36) with important implications for mode effects in web surveys.

Interactive and dynamic features of web questionnaires

Routing through the questionnaire, the absence of external help with question comprehension, and lack of immediate feedback on appropriateness of provided answers are some of the major issues of a respondent's interaction with a self-administered questionnaire. Mail surveys can rely almost exclusively on the appropriate questionnaire design to tackle these problems (Jenkins and Dillman 1997; Redline and Dillman 2001), but computerization of the questionnaire offers additional possibilities in web surveys.

User interface controls elements, constraining the input type to the expected answer format even in the most basic web questionnaires without any additional programming (Couper et al. 2004a). The increased interaction between the questionnaire and the respondent in dynamic web surveys extends this by bringing some features of human dialogue to self-administered surveys (Conrad et al. 2007). Appropriate programming of

the questionnaire can thus inhibit a part of mode effects stimulated by self-administration.

Questionnaire routing

The response model by Redline and Dillman (Figure 2.3, page 80) indicates additional cognitive burden imposed on respondents when a self-administered questionnaire contains additional navigational (branching) instructions. A direct influence on measurement error occurs when respondents incorrectly answer inapplicable questions, or incorrectly skip applicable ones and thus produce item nonresponse (Redline et al. 2005; Dillman and Christian 2005). Errors of commission can often be corrected by the simple dismissal of the inapplicable answer, but little can be done when questions are erroneously skipped. Computerized questionnaires allow complex skips to be performed automatically, without being even noticed by the respondent. This significantly reduces the burden imposed on respondents and virtually eliminates routing errors (J. Martin et al. 1993; Denniston et al. 2010). Consistently, Lorenc (2010) found item nonresponse to be higher on conditional items in a mail survey, but not in the web mode with automated routing. Automation of skips also reduces response times, which are an indirect indicator of the respondent's burden (Peytchev et al. 2006).

Answer controls and interventions

Another key advantage of computerization is the possibility to implement real-time answer validations. Web survey software can immediately detect whether an answer to a question is not provided (item nonresponse), is inconsistent compared to preceding answers, has an invalid format or numeric range, or does not match other predefined criteria[22]. When such event is identified, a prompt (error message) requesting a correction can be displayed to the respondent. The computerization here again helps to reduce mode effects related to absence of an interviewer to resolve problematic answers.

[22] A systematic typology of different validations in web surveys is provided by Peytchev and Crawford (2005)

Several studies confirm the effectiveness of prompts for the reduction of item nonresponse and non-substantive responses (Derouvray and Couper 2002; Wine et al. 2006; de Leeuw et al. 2010a). Conrad et al. (2005) used validations based on the sum of numeric answers and obtained more valid responses when interactive feedback was provided to respondents. Validation prompts may also encourage respondents to become more careful on subsequent questions (Peytchev and Crawford 2005).

However, real-time validations do not necessarily contribute to higher data quality. There is a general concern, although with limited and mixed empirical evidence, that too many validation prompts may lead to decreased quality of responses (Derouvray and Couper 2002; Couper 2008; Dillman et al. 2008). While some studies did not detect any negative effects of prompts (e.g. Mooney et al. 2003), some others indicated the possibility of increased respondent's frustration. Some observed manifestations of this included break-offs (Kerwin et al. 2006), lower quality of responses to open-ended questions (Peytchev and Crawford 2005), and input of invalid information (Grondin and Sun 2008).

With the advancement of web technologies, new validation mechanisms which focus on the response process are being used and evaluated. One recent example pointed out by Tourangeau et al. (2013) includes interventions to reduce unreasonably fast responding (Conrad 2011). Although initial results are encouraging, more careful evaluations of data quality are needed.

Presentation of definitions

Surveys sometimes deal with terms and concepts that are unknown to some respondents or have a too broad meaning in everyday use (Gerber et al. 1996; Schober and Conrad 1997). In such cases, additional definitions or clarifications become essential to bridge the gap between the respondent's understanding and the intended meaning of the concept. Of course definitions make little added value if the respondent does not appropriately incorporate them into the comprehension stage of the response process.

Differences between modes in the respondent's considerations of definitions are a function of three inherent mode characteristics: the presence of an interviewer, the channel of question presentation, and the respondent's use of computer technology for data collection.

Extended communication channels of auditory question presentation and oral response help facilitate the survey interaction in interviewer-administered surveys (de Leeuw 2005). On the basis of respondent's verbal, nonverbal, and paralinguistic expressions, the interviewer is able to identify comprehension difficulties and offer additional explanations or paraphrasing of the question (Conrad and Schober 2000; Biemer and Lyberg 2003). In self-administered modes, the identification of the need for clarification and the decision to consult it is completely up to the respondent. Self-administration is, however, not completely disadvantageous with regard to definitions. Conrad et al. (see also Conrad and Schober 2000; 2006) report that respondents are less likely to request further clarifications from interviewers than in computerized self-administered questionnaires. The authors explain this by the additional burden of formulating the request and reluctance to admit comprehension problems to another person.

The computerization of web questionnaires offers an additional layer of possibilities to provide definitions. They can be presented by default or upon request, for example by click on a hyperlink, mouse rollover, automatically after a period of the respondent's inactivity, and so on (Conrad et al. 2006; Conrad et al. 2007). Research shows a higher likelihood of considering definitions if less effort is needed to access them: when one rather than two mouse clicks are needed (Conrad et al. 2006), when they are displayed on mouse rollover instead by clicking (Conrad et al. 2006), and when they are displayed by default compared to all other alternatives (Galesic et al. 2008; Peytchev and Hill 2010). In general, these results are consistent with the visibility principle of higher use of immediately accessible information and lower burden (Tourangeau et al. 2013). However, a large number of immediately displayed definitions may lead respondents to start ignoring them or process them less carefully (Peytchev et al. 2010). Several experiments

by Conrad and Schober also indicate beneficial effects of automatically displayed clarifications when a respondent's uncertainty, measured by time of inactivity, is detected (Conrad and Schober 1998; Conrad and Schober 1999; Conrad et al. 2007).

The presented studies expose interesting implications of computerized self-administered questionnaires for mode effects. Following Conrad et al. (2006), interviewer-administration seems to discourage requests for clarification and so does a computerization-enabled interactive display in self-administered modes. The presence of an interviewer, on the other hand, has the advantage of the interviewer being able identify the need for clarifications, which can be at least to some degree replicated by advanced questionnaire programming. Depending on questionnaire implementation, computerization therefore allows the use of definitions more similar to interviewer-administered modes (presented upon request or in case of detected comprehension difficulties), or more similar to paper questionnaires (immediately visible). This can reduce between-mode differences, but on the other hand also introduce some of the problems regarding the definitions typical for these modes.

Other interactive and dynamic features

Interactive web questionnaires offer many other interactive and dynamic features with the potential of improving data quality by decreasing task difficulty and the burden of respondents. For example, burdensome matrix questions can be replaced with automated loops of question groups (Dillman et al. 2008), answers to preceding questions can be included into the text of subsequent questions to aid comprehension (Couper 2008), and probes for open-ended answers can to some degree resemble those used in interviewer-administered surveys (Holland and Christian 2009; Gravlee et al. 2013).

Finally, although we have by no means mentioned all the advanced features related to the interaction with web questionnaires, multimedia technologies can introduce virtual interviewers to ask questions and react to respondent's answers[23]. Virtual interviewers

[23] Although this is regarded as a separate mode according to our mode definition (Figure 1.2 on page 40), we include this example here to illustrate the interactive capabilities of modern web questionnaires.

are intended to increase respondent's engagement, improve the comprehension of questions, and allow for a more customized surveying experience (Tourangeau et al. 2013). Spoken question texts and the interactive provision of necessary clarifications by virtual interviewers seem to improve comprehension in the way live interviewers do (Conrad et al. 2008), although the benefits over highly interactive textual presentations are unclear (Tourangeau et al. 2013).

The response models for interviewer administered surveys (Esposito and Jobe 1991; Sander et al. 1992) explain complex cognitive processes performed by interviewers to process and record answers, and especially to identify the need of using additional clarifications and probes to tailor surveying procedures to the respondent's abilities and motivation. Computerized questionnaires will probably never be able to supplement all these processes, but they can bring several aspects of similar interaction to self-administered modes. This is in many cases beneficial over paper surveys, but sometimes these specific features of computerization can negatively influence data quality. Answer validations are a prime example of the complex relations between the factors of mode effects. Here, one inherent mode characteristic (computerization) enables the establishment of mechanisms to reduce mode effects attributable to another (self-administration). However, if these features of computerized questionnaires start producing other forms of low-quality responding, like satisficing, the result is a mere replacement of one measurement error due to mode effects with another.

Computerization and characteristics of respondents

Participation in web surveys requires respondents to have at least some basic computer literacy. Characteristics of respondents may therefore interact with the effects of computer administration. Buchanan and Smith (1999) expressed concern that responses in computerized psychological tests are affected by a computer anxiety among those not confident or familiar with the use of computers. In addition, specifics of electronic response channel could pose a challenge for some respondents. Mice and other pointing devices used to interact with the computer are conditioned by eye-hand coordination (e.g. Card et al. 1977), which can be especially troublesome for beginners or disabled

persons. Kaczmirek (2009) shows that certain default questionnaire elements (like radio buttons) are often missed when respondents attempt to click them, and this may be even more pronounced with less-experienced computer users. Dillman and Bowker (2001) also report increased frustration with a questionnaire among some inexperienced computer users, because they did not know how to use various interface control elements.

Lozar Manfreda and Vehovar (2002b) conducted a test-retest comparison of web mode with telephone and mail questionnaires. Their results show a smaller difference between modes among more frequent Internet users regarding some variables. This suggests that less computer experience may increase mode effects related to the use of computer technology in web surveys, although not generally and probably depending on specific characteristics of the question. Less experienced respondents may also face an increased burden of learning to use specific questionnaire elements, like hyperlinked skip instructions (Peytchev et al. 2006), and need more time to answer the questionnaire in general (Yan and Tourangeau 2008).

The use of electronic response channel (usually keyboard and mouse) is assumed to offer higher convenience of entering answers, reflected in longer open-ended responses compared to paper questionnaires (Wygant and Lindorf 1999; Kwak and Radler 2002). Less experienced users may, however, find the use of a keyboard burdensome. Consistently, more frequent computer users in the study by Denscombe (2008) tended to provide longer answers to open-ended questions. However, a simultaneous comparison with a paper-based mode would be beneficial to obtain a clearer picture on the role of interaction between a respondent's experience and computerization for the emergence of mode effects.

There is no definitive conclusion to what degree these results are a direct consequence of actual ability or willingness to interact with the computer. However, they tend to support a hypothesis of increased task difficulty due to computerization among less computer-savvy respondents. The corresponding mode effects are expected to manifest as satisficing to reduce the burden of survey participation.

3.2.2 Motivation

How respondents will use or abuse the available locus of control in self-administered modes strongly depends on their motivation. This is a complex construct and refers to processes that contribute to goal-directed behaviour (Esposito and Jobe 1991). In order to answer a question accurately, respondents sometimes need to loop between stages of the response process several times to retrieve and judge all the required information (Tourangeau et al. 2000). This is most clearly exposed in the response model by Willis et al. (1991), stressing the centrality of respondent's decision-making in governing the response process (Figure 2.1 on page 71). Motivation can be regarded as a strong factor of the respondent's decisions about the amount of effort they invest into question processing and therefore essentially leads to variations in the quality of answers.

Respondents participate in surveys for various reasons: self-expression, interpersonal response, intellectual challenge, self-understanding, feeling of altruism or emotional catharsis, or the desire to contribute to things they find important (Krosnick 1991). Additional extrinsic motivation can be equally or sometimes even more important to stimulate respondents for the thorough conduction of the response process. Inherent mode characteristics largely determine the capability of individual modes to contribute to extrinsic motivation. Clearly, this is one of the most pronounced disadvantages of self-administration as well-performing interviewers motivate the respondent's optimization of the response process (Groves 2004; Weisberg 2005). For example, by reviewing various mode comparisons, Groves et al. (2009) draw a direct relation between the interviewer's presence and lower percentage of unanswered questions. However, self-administered surveys are advantageous for motivation due to a higher locus of control at the respondent's end, allowing them to choose the most convenient time and place of survey participation.

We have several times referred the contributing role of motivation for many response effects in surveys. It has a very general effect of increasing or reducing the probability of satisficing (Eq. 2.1, page 73), and may also be influenced by task difficulty as another satisficing factor. Even when respondents are motivated enough to start participating,

their readiness to invest efforts into providing accurate answers may start declining due to fatigue, loss of interest, or impatience (Krosnick 1991). The increased burden of self-administered surveys can thus reduce the respondent's motivation. Lower motivation and lack of interviewers to provide extrinsic stimulations then continue to interact and can further increase the likelihood of mode effects.

Motivation and questionnaire length

By reducing the motivation of respondents due to increased burden, questionnaire length is expected to directly contribute to increased satisficing and other suboptimal strategies (Helgeson and Ursic 1994; Deutskens et al. 2004; Backor et al. 2007; Galesic and Bosnjak 2009). These deviations are expected to be more pronounced with self-administered modes where options for extrinsic motivation are far more limited (Dillman and Christian 2005).

Despite strong theoretical grounds for mode effects, very few direct comparisons of web surveys and other modes take into account the effects of questionnaire length. One rare exception, Lozar Manfreda and Vehovar (2002b) found a more pronounced decrease in the quality of answers to later items in web survey than in telephone surveys. The comparison of web and mail as two self-administered modes by Wolfe et al. (2008) also show somewhat higher item nonresponse for later items on the web. The latter result indicates the potential negative effects of computerization, but it is not possible to draw any reasonable explanations without further research of these issues.

Mechanisms for increasing the motivation of web respondents

The potential negative influence of self-administration on the motivation of respondents can be to some degree overcome using various mechanisms. Dillman (most recently in Dillman et al. 2008) proposed a sophisticated and comprehensively developed *Tailored Design Method*, largely based on social exchange theory, to motivate participation in self-administered modes. The key idea of the method is to establish trust, increase benefits, and decrease costs of participation. This includes specific design procedures in

each step of the survey process, such as pre-paid incentives, emphases of task importance, polite and motivating communication, and careful implementation of questionnaire design to increase the convenience of responding. Two specific factors we address here are incentives and interactivity of the questionnaire interface.

Incentives

Several experiments confirmed the potential of small-value incentives to increase participation of respondents, especially when incentives are monetary and pre-paid (Göritz 2006; Dillman et al. 2008; Vehovar et al. 2010). There is also some evidence of decreased item nonresponse with the use of incentives (Tuten et al. 2004; Deutskens et al. 2004; Heerwegh 2006), but very little is known about the effects on measurement errors and related mode effects. Deutskens et al. (2004) compared the accuracy of estimates obtained using different types of incentives and found no significant difference between groups.

Although we can speculatively reflect the findings of increased motivation for participation to the expectation of higher quality of responses, there are equally valid fears that incentives could increase measurement errors due to satisficing. Some respondents may just strive to reach the end of the questionnaire to collect a post-paid incentive (Göritz 2006), or invest a minimal effort to satisfy their need of social exchange. Much more research on these topics will therefore need to be done in order to learn how incentives can contribute to changes in mode effects.

Questionnaire design and interactivity

The possibilities offered by a visual input channel and computerization are sometimes regarded as a partial solution for limited possibilities of providing sufficient extrinsic motivation in web surveys. In section 3.1 we emphasized the importance of appropriate visual design for decreasing task difficulty. Deutskens et al. (2004) believe that visually appealing design can effectively increase respondents' motivation to provide better-quality responses, as long as it does not interfere with their task.

Deutskens et al. (2006) regard respondent's satisfaction as one of the strongest driving factors of motivation. Mode comparisons often show higher respondent's satisfaction with surveying experience in web surveys than paper self-administered questionnaires (Layne et al. 1999; Grandcolas et al. 2003; Mangunkusumo et al. 2006), but it remains unclear whether satisfaction is mainly influenced by the convenience of a computerized questionnaire or some other factors.

Visual design and interactive capabilities of computerized questionnaires are often exploited in the hope of positive influence on the respondent's experience and motivation. For example, Dillman et al. (2008) show how an attractive and relevant image on the introductory page helps to stimulate respondents' motivation to participate in the survey. Kunz and Fuchs (2013) also successfully used dynamic motivating instructions to reduce response-order effects in web surveys. However, Couper (2008) notes the lack of sufficient proof about the generally positive impact of web questionnaire design on the motivation of respondents. Justifying Couper's reluctance, Downes-Le Guin (2012) reports various studies that indicate a preference of respondents for simple and standard questionnaire elements (like input boxes) over visually more appealing customized ones.

A well-researched example of using interactive features with the intention to increase motivation of respondents is progress indicators. These dynamic elements offer respondents feedback about their advancement through the questionnaire. Although the initial studies by Couper et al. (2001) were positive about the use of a progress indicator to reduce drop-out, a majority of later studies exposed the opposite effects (e.g. Crawford et al. 2001; Matzat et al. 2009; Yan et al. 2011). This was further confirmed in a meta-analysis by Callegaro et al. (2011) and a literature review by Tourangeau et al. (2013). Progress indicators are thus more likely to *decrease* than increase the motivation of respondents. Because such feedback is largely mode-specific and not continuously available in auditory modes, the potential negative influences on decreased motivation and measurement error are directly related to mode effects.

Figure 3.3: An example of interactive and highly gamified web survey by Ghoneim (2013)

Note: Screenshot obtained from Ghoneim (2013).

New possibilities of motivating respondents through the attractiveness of the surveying task are being continuously investigated. One of the most notable and controversial recent approaches in web surveys is the gamification of the questionnaire. In the context of web surveys, this somewhat buzzword and inconsistently-defined term refers to the incorporation of game-like elements into the respondent-computer interaction during the surveying process (Downes-Le Guin et al. 2012). According to Bälter (2005), the underlying idea is to equip surveys with the same elements that keep people playing computer games for several hours at a time. Actual implementation can vary significantly, from relatively simple but a visually appealing feedback to respondents' answer (Bälter 2005) to a completely game-like questionnaire interface (Figure 3.3). Whether or not such innovative approaches can improve the participation of respondents and motivate them to perform the response process more thoroughly will become clearer in the future. The research will also have to take into account performance over different groups of respondents, especially those who do not find the idea of computer games interesting at all, or are not sufficiently experienced with computers in such highly interactive way.

In a recent study, Casey and Poropat (2014) found that the use of creative and highly expressive questionnaire design can simultaneously increase both the state positive and state negative affect of respondents. A more classical aesthetic design, on the other hand, decreased the state negative affect and increased the perceived ease of use. These findings, along with potential contextual effects of visual design that we discussed in section 3.1.3, caution against radical web-specific design and interactive approaches without a sufficient understanding of their methodological consequences for survey outcomes.

3.2.3 Perceptions of privacy

As described at the beginning of this chapter (section 2.2), sensitive questions are often prone to response editing in a socially desirable direction. Respondents are well aware of the disclosure of their answers to research, but may also fear of their disclosure to third-parties. Sensitive questions can therefore be expected to stimulate some impression management in any survey mode. However, variations in reports to sensitive questions between modes are well-recognized and show that inherent mode characteristics differ in the level of privacy perception they offer to respondents (Richman et al. 1999; Tourangeau and Yan 2007).

Privacy and self-administration

If one was to point out a single typical example of mode effects, the answer would very likely be related to the differences in reporting on sensitive topics due to influences of the interviewer involvement. This would hardly be surprising, as the issue is probably the most researched, comprehensively understood, and consistently empirically supported consequence of mode effects.

Impersonality of interaction in self-administered modes reduces the respondent's sense of disclosure to someone with whom they personally interact (Tourangeau et al. 2000;

Tourangeau and Yan 2007). The first measurement-related advantage of this is the complete elimination of effects occurring due to politeness to interviewers: sometimes respondents do not misreport their position because they would regard it as a violation of social norms, but alter their response to be more in line with specific interviewer characteristics, like race (Schaeffer 1980). These effects can occur even if the interviewer is merely present and does not administers the survey, like in self-interviewing modes (Krysan and Couper 2003).

When questions are sensitive, respondents may be concerned with interviewer's approval or disapproval, even if the response process is unaffected by the interviewer's characteristics (Tourangeau et al. 2000). An early study by Hochstim (1967) found that mail surveys produce higher reporting on sensitive questions than face-to-face and telephone surveys. This general pattern was later confirmed by numerous other experimental verifications, recently summarized in a meta-analysis by Tourangeau and Yan (2007).

Consistent with expectations, web surveys were shown to elicit better performance on sensitive issues than telephone (Lozar Manfreda and Vehovar 2002b; Jäckle et al. 2006; Kreuter et al. 2008; Chang and Krosnick 2009) and face-to-face interviewing (Jäckle et al. 2006). Tourangeau et al. (2013) conducted a meta-analysis of ten studies comparing a web survey to another mode and further confirmed the advantages of the web mode.

Impact of computerization on privacy perceptions

It is less clear how privacy perceptions are affected by the computerization of self-administered questionnaires. According to de Leeuw (2008), some respondents may experience a lower degree of privacy with the use of computer, for example due to the "big brother" effect, while others may perceive computerized data more secure against third-party access. Comparisons between web and mail surveys could help revealing these effects in web surveys, but do not offer a clear picture. Two studies observed significant higher reporting on the web (Lozar Manfreda and Vehovar 2002b; Kypri and

Gallagher 2003), but a majority of comparisons found no significant differences and inconsistent directions of effects (Wygant and Lindorf 1999; Pealer et al. 2001; Knapp and Kirk 2003; McCabe et al. 2006). A meta-analysis by Tourangeau et al. (2013) summarized the results of ten comparative studies, and concluded that potential advantages of web surveys over mail mode survey are only very minor.

Generally inconclusive results also hold true for studies comparing other computerized and paper-based self-administered modes. Although some early studies demonstrated an increased sense of privacy with computerized modes (e.g. Evan and Miller 1965; Kiesler and Sproull 1986), meta-analyses by Richman et al. (1999) and Tourangeau and Yan (2007) again show only a subtle and non-significant tendency in favour of this expectation.

However, there is some indication of a lower willingness of respondents to disclose sensitive information to the computerized questionnaire when they are aware of lack of anonymity. Richman et al. (1999) found significantly lower reporting in a non-anonymous computerized self-administered questionnaire than in a non-anonymous paper questionnaire. The possibility of the "big brother" effect is also tentatively supported by two studies that yielded slightly more positive answers in web than in mail surveys. Since both were inter-organizational surveys, one on work satisfaction within a company (Smither et al. 2004) and the other on study issues (Carini et al. 2003), respondents probably felt less confident about the anonymity of computerized questionnaires.

It is also important to consider a potential decrease of perceived privacy due to specific implementations of highly interactive features in web questionnaires. Tourangau and Yan (2007) caution on the effect of *media presence*. Increased interactive capabilities of computerized questionnaires may create the illusion of presence and trigger effects commonly found with interviewer-administration. For example, experimental inclusions of virtual interviewers into self-administered questionnaires produced some gender and race effects common to interviewer-administered surveys (Krysan and Couper 2003; Fuchs 2009). Overexploiting computerization to humanize web questionnaires can thus

reduce the benefits of higher reporting on sensitive topics by introducing mode effects similar to those caused by the presence of interviewer.

3.3 Web as information transmission medium

The last inherent mode characteristic of web surveys is the most obvious of all: the World Wide Web as a medium through which all information to and from the respondent is transmitted. It enables some of the major advantages of web surveys, including cost-independent widespread distribution, fast data collection, and surpassing of physical boundaries. It is inseparably entangled with other inherent mode characteristics. The web is predominantly visual, an essentially computerized medium, and interaction through it is based almost exclusively on the use of electronic devices.

Web surveys, as defined in this dissertation and commonly understood in survey methodology, are not the only possible mode, based on the web as an information transmission medium. There is, for example, no technical limitations in conducting interviewer-administered surveys via the web, present questions using an auditory rather than visual input channel, or relying on voice recognition systems to enable oral responding instead of electronic response output. Couper (2008) reports attempts to present web questionnaires using only the auditory input channel, which proved to be unsuccessful, presumably due to the web being regarded as primarily text-based medium. Currently, there is no substantial methodological research on such alternative web-based modes.

Because of the close relation to other inherent mode characteristics and a lack of practice with other web-based modes, relatively little can be said about the medium's own implications for mode effects. In this chapter we expose three aspects relevant for survey data collection that seem to be most strongly influenced by the medium's specific social and technical characteristics: definition (or restriction) of a physical environment in which surveying takes place, the ability to convey the legitimacy of a survey request,

and technical issues related to a variety of hardware and software systems used to access the questionnaire. The first two aspects cannot be addressed much beyond the theoretical expectations based on general properties of the web and some implications drawn from various studies discussed above. On the other hand, rapid developments and the widespread adoption of mobile web-enabled devices started drawing more empirical attention to the technical issues of web surveys.

3.3.1 Environment of the survey situation

The web as a medium for information transmission defines the range of physical environments in which surveying can take place. While self-administration allows respondents to choose the most preferred location of completing the questionnaire, the web medium somewhat restricts these possibilities by requiring access to computers or other web-capable devices and the Internet. On the other hand, remote access to the questionnaire allows us to regard it as the least restrictive of all surveying media. Respondents can access a web questionnaire at any time and from virtually anywhere as long as technical preconditions are fulfilled. A respondent may, for example, remember to complete the questionnaire on an airplane. This is an advantage over paper-based modes where the respondent has to have the paper questionnaire at hand. However, this interaction of self-administration and web-enabled freedom of choice can also have some negative effects.

The first set of negative effects can arise from the immediate environment in which the questionnaire is completed. Research has shown that the presence of others decreases the perception of privacy and increases false reporting even when the survey is self-administered (Beebe et al. 1998; Aquilino et al. 2000; Castelli and Tomelleri 2008). The respondent may, for example, feel uncomfortable answering sensitive questions if other people around could see answers on the respondent's computer screen (Denniston et al. 2010). Similar effects of decreased privacy were found in group-administered web surveys (Epstein et al. 2001; S. C. Bates and Cox 2008; Eaton et al. 2010). It is unclear to what extent these results can be attributable to mode effects as the presence of other

persons seems to work largely consistently across modes with different inherent mode characteristics (Tourangeau and Yan 2007; Tourangeau et al. 2013). However, due to the high flexibility of settings in which the survey can be completed using the web as a medium of information transmission, we can speculate that web surveys produce a higher variability of effects related to the surveying environment.

More specific effects of the data collection environment in web surveys may arise due to the distracting factors of using the web. During participation in a web survey, the respondent might have a large number of concurrently opened websites and other programs. Continuous availability of new information from these sources can interfere with a thorough performance of the response process (de Leeuw 2005; Heerwegh and Loosveldt 2008). While there is no direct empirical evaluation of such distractions in web surveys, human-interaction studies confirm high prevalence of multitasking and task switching during retrieval and processing of information on the web (e.g. Wang and Chang 2010). Distractions substantially increase the difficulty of optimizing and increase the likelihood of resorting to satisficing strategies (Krosnick 1991).

3.3.2 Legitimacy

Legitimacy and the perceived importance of the survey often govern the respondent's decision to participate in a survey and are one of the key mediating factors of the obtained data quality (Tourangeau et al. 2000). Characteristics of the medium used for surveying, as well as social and personal attitudes toward it, importantly affect the mode's ability to convey a legitimacy of the survey request, i.e. the sincerity of purpose (de Leeuw 1992).

Close personal interaction is a powerful tool for presenting survey credentials and undoubtedly contributes to a generally higher response rate in face-to-face surveys (Tourangeau et al. 2000). An absence of interviewers in web surveys severely limit their capacity to establish legitimacy among respondents. In addition, specific perceptions of the web as medium can further and profoundly contribute to this issue. A large amount

of spam e-mail messages, fraudulent websites, media-fostered privacy concerns and reports on security breaches, and many other potential perils continuously undermine any general trustworthiness of the web. This is most directly reflected in lower response rates in web surveys compared to other modes (Lozar Manfreda et al. 2008).

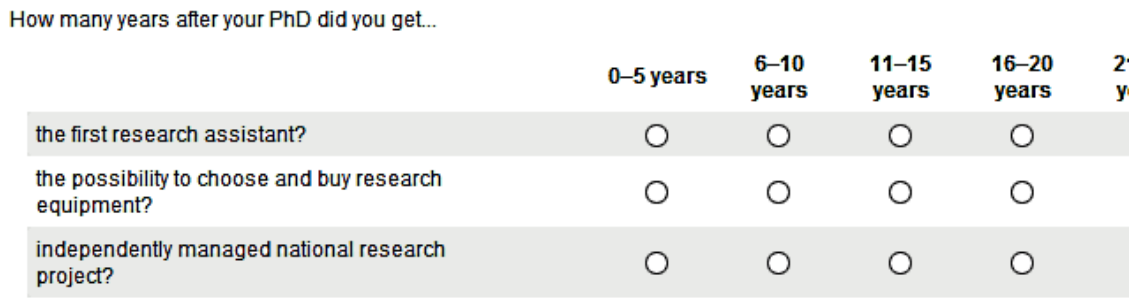
Tourangeau et al. (2013) add the negative side of web-fostered democratization of the survey process. User-friendly and low-cost software tools boost a large amount of low-quality surveys, making it hard for respondents to distinguish between good and bad ones. In the past, the use of computer technology for data collection was believed to increase legitimacy and perceived importance of the survey (Kiesler and Sproull 1986; Tourangeau et al. 2000). With web surveys this effect apparently faded away as the use of computers and other devices became everyday practice.

Krosnick (1991) closely relates legitimacy and importance to the motivation of respondents in investing their efforts into optimizing. Tourangeau et al. (2000) also link the perceived legitimacy to the increased reporting of sensitive behaviours. Researchers must therefore make good use of other mode characteristics of web surveys to establish the trust of respondents. A professional look and feel of the questionnaire, inclusion of logos, privacy assurances, motivating interactivity, and stressed importance of the task are only some elements of self-administration and computerization that can help achieve this goal (de Leeuw 1992; Dillman et al. 2008; Couper 2008). Important improvements can be achieved also by the appropriate implementation of survey design phases preceding data collection, for example, by providing incentives and using mixed-mode contacts, like ordinary mail invitations (de Leeuw 2005; Lozar Manfreda et al. 2008).

3.3.3 Technical issues

The methodological importance of the technical aspects of web surveys lies in the combination of computerization of the questionnaire and the web as the medium through which this questionnaire is brought to respondents. As a result, both impact how information is transferred through the visual question presentation channel. Technical errors

Figure 3.4: An example of scale questions where horizontal scrolling is needed to see all response categories due to low screen resolution



can lead to the incomplete presentation of information, failures in interaction with the questionnaire, the increased burden of respondents, and other unwanted scenarios with a potential to negatively affect the accuracy of data.

The technical foundation of any web questionnaire is an HTML form, optionally supported by additional client-side web technologies such as JavaScript, Silverlight, Flash, and Java. These additional technologies are a prerequisite for many interactive questionnaire features, beginning with real-time validations, within-page question skips, and some more complex questions types. The distribution of the questionnaire through the web medium means its operation in largely uncontrollable technical environments, consisting of a wide variety of different software and hardware systems used by respondents. Different devices, browsers, operating systems, screen resolutions, internet connection speeds, and many other factors can introduce variations into the display and performance of the questionnaire (Buchanan and Smith 1999). Because of this, Couper (2008) regards the respondent's experience with a web questionnaire as less standardized than with a paper-based or offline computerized questionnaire.

Variations between personal computers

The general performance of a web questionnaire in different web browsers became less troublesome over the recent years due to their increased if still less-than-perfect compliance to standards (Zeldman and Marcotte 2009). Also, JavaScript as the widespread client-side scripting technology seem to remain unavailable only on a minor proportion

of desktop web browsers. Funke and Reips (2012) report approximately 1% of respondents who were unable to use special question types due to disabled JavaScript.

A more common problem arising from variations between computer systems is related to different screen resolutions, which can affect the amount of information conveyed through the visual input channel. While vertical scrolling is common in everyday use of the web, the need for horizontal scrolling can be easily overlooked by the respondent and is thus much more problematic (Couper 2008). An example in Figure 3.4 shows how the presentation of a grid question on a low-resolution screen can critically violate the visibility principle and might increase selection of more visible categories, either due to satisficing (response order effects) or due to overlooked categories.

Many other variations may be even less under the control of the researcher and survey software tool. Some examples include reduced clarity of questionnaire presentation due to inconsistent colours across displays, prevention of prompts by browser-level blocking of dialogue boxes, long loading times of multimedia elements on slow internet connections that increase the burden, and so on. The burden of respondents can also vary across devices used for electronic answer output. For example, laptop touchpads are for many users more burdensome than an ordinary computer mouse (Kelaher et al. 2001).

How much these and others technical issues due to variations in the performance of the questionnaire across different technical environments bias the results depends on whether the users of problematic technologies differ from others. This issue is illustrated by Buchanan and Reips (2001) who found personality differences between PC and Mac users. If this was a topic of interest and the questionnaire was not compatible with both platforms, the obtained estimates would be biased.

Mobile and other web-capable devices

With the growing popularity of smartphones, tablet computers, and other devices with web access, the problems outlined above have already became more pronounced and

more widely considered by survey methodologists[24]. As for now, only a minority of respondents seem to opt for using a mobile phone to participate in web surveys, even when they are explicitly encouraged to do so (Millar and Dillman 2012). However, this is likely to change with the development of a more user-friendly web browsing experience on mobile devices.

The most obvious and highlighting specific of mobile devices from the perspective of surveying are (very) small screens, less convenient methods of text entry via small on-screen or hardware keyboard, and substantial differences in software support of web technologies (Peytchev and Hill 2010). The negative effects arising from these factors are already supported by some methodological studies. Three non-experimental comparisons between respondents using mobile devices and desktop or laptop computers by Callegaro (2010b) found an increased drop-out among the mobile group. Peytchev and Hill (2010) report several indications of an increased burden with the use of mobile devices, including item nonresponse on even short open-ended questions, overlooked information, and satisficing. Higher satisficing in the form of non-differentiation was also found by Guidry (2012).

Buskirk and Andrus (2012) expose that a lower implementation of standards into different mobile web browsers can introduce substantial differences in the presentation of the same questionnaire across devices. In their study, users of some devices and operating systems were also more likely to leave a question unanswered or terminate participation (similar was also found by Guidry 2012). Finally, demographic differences between users of different devices again introduce concerns of biasing results due to technical incompatibilities, expressed for personal computers over a decade ago by Buchanan and Reips (2001).

Survey methodologists still face a long journey to be able to understand and tackle mode effects in web surveys caused by mobile devices. New developments are constantly

[24] Under the term *mobile device* we consider mobile phones (smartphones) and tablet computers, but not ordinary laptop computers with desktop operating systems.

changing their nature, and even blurring boundaries between them. Some laptop computers (subnotebooks) are becoming increasingly smaller, more powerful, and with touch-screen capacities, while the size of tablets and mobile phones is increasing. Internet access is increasingly available on other every-day devices (like TV sets), which can also have specific impacts on the response process (Chang and Krosnick 2009). Callegaro (2013) therefore questions whether respondents have overtaken us when it comes to answering surveys. The answer may well be affirmative.

3.4 Mode-specific response sets in web surveys

Throughout this chapter we have noted several times how mode effects encourage specific forms of response behaviour that results in biased estimates. Most obvious forms are various satisficing strategies and social desirability bias, already elaborated in section 3.2.3. Evan and Miller (1965), following the work by Cronbach, label the tendency of providing consistently directed answers due to specific presentation of content as *response sets*. In this part we review findings of studies dealing with the occurrence of some typical response sets in web surveys compared to other modes. Comparative observations enable further insight into the consequences of mode effects in web surveys.

A majority of mode-specific response patterns cannot be attributable to a single inherent mode characteristic and its mediating factors, nor to one specific interaction of them. In many cases they are a result of several independent and complexly interrelated influences, which may contribute to deviations in different directions. Furthermore, their occurrence is almost universally moderated by implementation-specific and contextual factors, including questionnaire implementation approaches and specific respondent's characteristics. Research shows a generally heightened reliance on satisficing strategies with increased task difficulty and lower performance on different indicators of cognitive abilities, like education (Krosnick and Alwin 1987; Narayan and Krosnick 1996; Malhotra 2008) and an age-related decrease in the working memory capacity (Knäuper et al. 1997; Knäuper 1999; Schwarz and Oyserman 2001). Consistent

with the satisficing model, motivation is also positioned as the foundation of many deviations in the response process. This makes a strong base for the expectation of mode-related differences in the occurrence of response sets, as motivation and task difficulty can be significantly influenced by a specific mode. Yet, many of these factors explain only a part of the problem.

3.4.1 Response order effects

Respondents are generally expected to process all response options in the presented order and with equal depth. In practice they sometimes devote more cognitive attention to response options presented at the beginning or at the end of the list. This leads to the emergence of primacy or recency response order effects, respectively (Krosnick and Alwin 1987). A type of effect that very much depends on the channel of presentation, inherent to a specific mode.

When response options are presented visually, the respondent may begin processing initial options with sufficient effort, but potentially in a confirmatory direction by considering the favourable reasons for its selection. If the respondent's cognitive burden surpasses their ability or motivation, they may start processing later options rather superficially. Superficial attention to later options was confirmed in an eye-tracking study by Galesic et al. (2008). This increases the likelihood of the selection of one of the initial options. According to Schwarz and Oyserman (2001), deviations from the optimal response processes occur because of limited processing time due to fatigue, interference of previously retrieved information with later retrieval processes, or the subjective feeling of a sufficient number of already endorsed options in multiple-answer questions. Primacy effects are thus more likely to occur in visual modes. According to Krosnick (1991), they may also appear when response options are presented orally if the respondent does not evaluate alternatives on the fly, but starts recalling the first one only after the whole list was read.

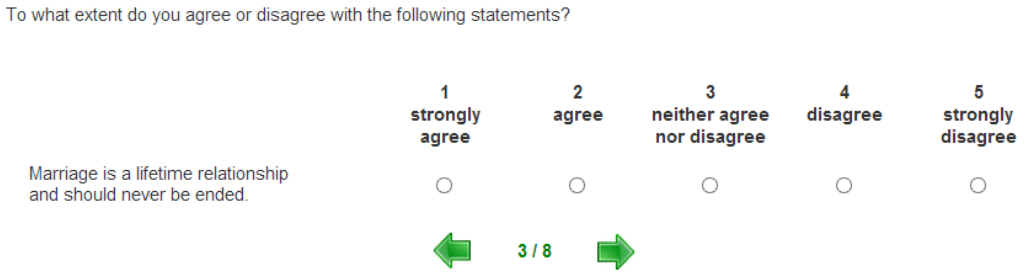
The mechanism of recency effects is somewhat different (Krosnick and Alwin 1987). They occur when processing of a previous response option is terminated by the presentation of a new one. This is more typical in auditory modes because new response alternatives are often presented relatively quickly. Time available to the respondent to consider each response option is thus much more limited than in visual modes. Since the processing of the last option is not terminated in this way, more processing time is available, and the option becomes more likely to be selected. In visual modes, response options are visible all the time and recency effects are thus less likely to occur (Ayidiya and McClendon 1990).

Several studies show an incidence of primacy effects in visual modes, such as mail surveys and face-to-face interviews (e.g. Krosnick and Alwin 1987; Ayidiya and McClendon 1990; Rasinski et al. 1994; Krosnick 1999; Duffy 2003). Consistent with their visual nature, the effects have also been observed in web surveys (Couper et al. 2004a; Smyth et al. 2006b; Galesic et al. 2008; Malhotra 2008; Kunz and Fuchs 2013).

Because of the flexibility of questionnaire design possibilities and technical aspects of computerization, response order effects in web surveys also substantially depend on implementation-specific and contextual characteristics of the survey. In section 3.1.2 (page 89) we reported on studies showing the more likely occurrence of response order effects with certain question formats. This is especially common with the formats that restrict the range of immediately accessible information, like drop-down menus (Couper et al. 2004a; Galesic et al. 2008). A greater selection of initial response options, although possibly more often related to overlooked information than to satisficing, can also occur due to specific technical characteristics of the respondent's device used to access the questionnaire. This may be particularly the case with low-resolution screens as presented in Figure 3.4 (page 124).

Although these empirical findings are largely consistent with the theoretical predictions of response order effects, some challenges reveal a lacking understanding of the underlying causes. There is no consensus on many factors moderating response order effects,

Figure 3.5: An example of horizontal scrolling matrix



and different studies often produced inconsistent findings (Dillman et al. 2009). An analysis of a large number of early Gallup experiments by Bishop (1997; also in Bishop and Smith 2001) revealed a higher occurrence of recency than primacy effects with the use of showcards (visual input). Duffy (2003) also shows that many respondents do not process visually presented questions with a top-down reading of response options, but rather read bottom-up or keep jumping up and down. Finally, when some response options are highly implausible, either primacy or recency effects could occur, and both types may even be cancelling each other out at the aggregate level (Sudman et al. 1996). On the basis of empirical studies we can thus expect the visual question presentation channel of web surveys to foster primacy effects, but several other factors can supress their emergence or even produce the opposite effect.

3.4.2 Non-differentiation

Non-differentiation (*straight-lining*) is probably the most typical strong satisficing strategy occurring in scale questions. Straight-lining respondents simply select the same scale value to all or almost all items, regardless of their content. This allows them to abandon virtually all stages of the response process (Krosnick 1991).

Web surveys are often regarded as particularly inviting for non-differentiation. The visual presentation of scale items offer an idea for forming the shortcutting strategy to strong satisficers, who make every effort to reduce their cognitive commitment to a minimum that still allows a seemingly reasonable answer.

Empirical comparisons, however, do not offer consistent support of increased non-differentiation in web surveys. Compared to telephone interviewing, Fricker et al. (2005) found more non-differentiation on the web, while Chang and Krosnick (2009) obtained the opposite results. Heerwegh and Loosveldt (2008) report slightly higher non-differentiation on the web compared to face-to-face interviewing with showcards, although both modes rely on a visual input channel for presentation of response categories. The primary source of non-differentiation in web surveys may thus not be the visual channel of presentation, but rather self-administration and the corresponding lack of interviewer-provided extrinsic motivation.

Better evidence of the role of the question presentation channel in fostering non-differentiation in web surveys can be obtained from within-mode comparisons. Some studies observed significant variations in its occurrence across different visual presentations of web questionnaires. Tourangeau et al. (2004) observed increased straight-lining if all scale items were presented on one screen. The effect decreased when items were distributed across several questionnaire pages. To address this issue, Klausch et al. (2012) proposed the format of horizontal scrolling matrix, where one item at a time is presented to the respondent, similar to auditory modes (Figure 3.5). This resulted in a significantly lower level of non-differentiation compared to ordinary scale questions with several concurrently visible items.

Non-differentiation in web surveys is thus most likely affected by both, self-administration and the visual input channel. Further research on the comparison of horizontal matrix scale questions with auditory modes will most likely help establish more solid explanations of mode-related influences on non-differentiation.

3.4.3 Acquiescence

The tendency to agree with assertions lies at the intersection of satisficing and impression management. As a shortcutting strategy it is the result of question processing in a

confirmatory direction, while impression management is manifested by editing of answers in an agreeable direction due to fear of disapproving reactions from others (Krosnick 1991)[25]. Although some authors (like Knowles and Condon 1999) lean more toward the former explanation and others towards the latter (e.g. Smyth et al. 2009a), both offer support to a hypothetical attribution of mode-specific acquiescence to the same mode characteristics that generally influence satisficing and social desirability. In line with our previous treatment of these topics, we would expect acquiescence to be reduced in the web mode, at least in the part arising due to impression management issues.

Unfortunately, current research again offers no consistent empirical evidence. Comparisons between web surveys and other modes yielded overwhelmingly non-significant results (Knapp and Kirk 2003; Fricker et al. 2005; Smyth et al. 2006b). Smyth et al. (2009a) report generally non-significant and non-conclusive results also for comparisons between other modes (e.g. de Leeuw 1992; Jäckle et al. 2006), concluding that more research is needed to understand the underlying causes of acquiescence response bias in different modes.

3.4.4 Answers to scale questions

A large body of empirical studies observed differences in answers to scale questions between modes. Focusing on comparative studies of the web mode, the most consistent results were found against telephone interviewing. Respondents to telephone surveys are more likely to choose extreme scale points (de Leeuw et al. 2010b), especially the ones on the positive end of a scale (Taylor 1999; Roster et al. 2004; Christian et al. 2007a;

[25] Krosnick (1991) treat the latter cause of acquiescence as a form of strong satisficing, claiming that respondents in this case completely dismiss the retrieval or judgment stages of the response process. However, authors of some later studies (e.g. Knowles and Condon 1999; Smyth et al. 2009a) regard impression management as editing behaviour occurring during the response formulation stage. See also section 2.1, page 67.

Dillman et al. 2009). In contrast, web respondents more often choose the middle categories (Roster et al. 2004; de Leeuw et al. 2010b), or express negative positions (Roster et al. 2004).

Results are somewhat less clear for comparisons between web surveys and other visual modes. Some studies found more favourable responses in web than in mail surveys (N. Bates 2001; Carini et al. 2003; Smither et al. 2004). As regards the extremity, findings are mixed. McDonald and Adam (2003) report on somewhat more extreme responses in the mail mode, while Grandcolas et al. (2003) observed the opposite. Dillman et al. (2009) did not find significant differences in extremeness of scale responses between web and mail modes. Finally, in the comparison of web and face-to-face survey with showcards by Heerwegh and Loosveldt (2008), respondents more often used the middle category in the web mode, while there was no significant difference in the extremeness of responses.

To begin exploring how much these variations can be attributable to mode effects in web surveys and how much to the effects of other modes, it is first useful to consider previous research comparing mail surveys with face-to-face and telephone modes. Dillman (1991) ascribes less extreme answers in mail surveys to a more complete utilization of the entire scale range. He attributes this to three key factors related to characteristics of the mail mode, but admits that they account for only a part of the observed differences between modes:

- Visual presentation provides context more explicitly and enables respondents to take into account both preceding and later items.
- Less socially desirable responding reduces the selection of more desirable extreme points.
- The locus of control at the respondent's side allows higher control over pace of responding and contributes to less top-of-the-head answers.

Two other potential influences of visual presentation can be exposed, although not directly mentioned by Dillman (1991). According to Jenkins and Dillman (1997), the horizontal layout of response options establishes an impression of relations between

scale values, creating a more explicit representation of a continuum[26]. In addition, the visual presentation may help reduce the impact of recency effects on scale point selection (de Leeuw 1992).

More recent research justifies to certain degree Dillman's reservations on the explanatory power of these factors. The findings of various studies express some doubt on the role of visual presentation in the utilization of the whole range of scale values. The importance of considering previous and later items is questioned by the experiment using a horizontal scrolling matrix by Klausch et al. (2012). The authors discovered even *less extreme* answers to the web mode when they were presented one-by-one instead of in a typical grid format. Although less extreme responses can also be the result of a satisficing tendency to select middle values, they found no such evidence.

Tourangeau et al. (2013) note the lack of evidence that more explicit scale representation using a horizontal layout would produce substantial differences in answers to web questionnaires. Exceptions arise with some specific questionnaire-provided contextual information discussed in section 3.1.3 (pages 95 and 97). For example, scale-point distances and differentiating colours can significantly influence the obtained estimates (Tourangeau et al. 2004; Tourangeau et al. 2007). Visual presentation can thus bias scale estimates when information transferred through the visual input channel contribute to distorted impressions of a continuity and balance of the scale.

In addition, response order effects do not seem to be a major determinant of between-mode differences. Tarnai and Dillman (1992) provided telephone respondents with printed questionnaires to give them a visual representation of the scale, but this did not reduce the extremity of responses. Christian et al. (2007a) also show the persistence of more extreme responses regardless of the scale direction and the number of scale points.

[26] Their argument is well-grounded from the perspective of conceptualization of response scales. Scales are based on an assumption that individual's attitudes can be expressed on a continuum with endpoints representing extreme unipolar or bipolar opposites. The idea is derived from the concept of psychological continuum for stimuli comparison introduced by Thurstone (1927).

More favourable empirical support is offered to the social desirability hypothesis. A recent meta-analysis of response distributions in scale questions across different survey modes by Ye et al. (2011) confirmed generally more extreme *positive* responses in telephone interviews than mail and web modes, but telephone surveys did not differ significantly to face-to-face interviews. The authors thus regard the interviewer involvement as the most probable major cause of between-mode differences in answers to scale questions. However, they caution against attributing more positive responses to social desirability, which is a misreporting of (un)desirable behaviours or characteristics. It is necessary to take into account a more general view on the respondent's reluctance to express negative evaluations in the presence of a stranger (i.e. an interviewer).

These effects may be confounded with a higher control over the pace of responding in self-administered modes. More superficial processing due to time pressure in telephone interviewing can result in the less careful consideration of the whole scale (Christian et al. 2007a). The results of the meta-analysis are not consistent with this hypothesis as they show even stronger shifts to extreme values in the face-to-face mode, where time pressure is reduced by a direct personal interaction. Of course, this does not preclude the possibility of simultaneous positive influences of both privacy perceptions and increased control over the response pace in web and mail surveys.

Putting all this together, recent research indicates a lower impact of mode effects on scale questions in web surveys than in interviewer-administered modes. Unfortunately, little can be concluded about the causes of differences between web and mail surveys, apart from situations when computerization negatively influences the perception of privacy (section 3.2.3, page 118). The exact causes of differences between different modes are therefore difficult to pinpoint, but they are more likely to be a conglomerate of various mode-related factors than attributable to one single factor (Dillman et al. 2008).

3.4.5 Non-substantive responses

With attitudinal and many factual questions it is possible that the respondent does not know the answer or does not have enough information to form an opinion. Inclusion of applicable non-substantive response options might thus seem a meaningful approach. However, according to Krosnick and Fabrigar (1997; also Krosnick et al. 2002), studies do not provide consistent evidence whether inclusion of "don't know" responses increase data quality or not. In addition, there is a general pattern of more non-substantive responses when respondents are explicitly offered the option. This lead to the general practice of avoiding explicit presentation of non-substantive response categories, unless they are essential for research (de Leeuw et al. 2003). However, interviewers usually record such responses if reported by the respondent anyway (Dillman and Christian 2005). This logic of non-explicit response options is much more difficult to achieve in visual self-administered modes.

An increased selection of explicitly offered non-substantive options dismisses the objective inability to provide an answer as the only respondent's motive for saying "don't know". Krosnick and Fabrigar (1997) list several other possible reasons: ambivalent attitudes in absence of scale midpoint, a neutral standpoint in absence of a neutral response option, inadequate understanding of the question, ambiguity of scale point meaning, and satisficing. Mode effects on non-substantive responses can thus be attributable to a range of mode-related influences on the respondent's comprehension of the question and the increased likelihood of satisficing.

Two studies comparing web and mail surveys produced inconsistent findings: Shin et al. (2012) found no significant differences after controlling for socio-demographic variables, while Bech and Kristensen (2009) did observe a higher proportion of non-substantive responses to socio-demographic questions. The authors, however, make no explanations for the observed effects.

Empirical verification of the mode-related influences between web and interviewer-administered modes is even more limited due to mode-specific practices of handling non-

substantive responses described above. For example, Kreuter et al. (2008) suggest that for sensitive items “don’t know” answers are more common in telephone interviewing than on the web, but their implementation of the web questionnaire does not allow for the separation of item nonresponses from non-substantive responses. Heerwegh and Loosveldt (2008) found more non-substantive responses in web than in face-to-face mode; however, in the former, non-substantive categories were provided explicitly, while in the latter, respondents were merely informed about the possibility of saying “don’t know”. A more comparable experimental study was conducted by de Leeuw et al. (2010a). The authors explicitly provided “don’t know” response options in both telephone and web survey, and obtained a higher proportion of such responses on the web.

The results tentatively indicate the potential contribution of the visual presentation channel to the more frequent use of non-substantive responses. The visibility principle may backfire here, since visible non-substantive options offer an inviting opportunity for low-motivated respondents to find an easy way out. Of course, we cannot exclude the role of lacking extrinsic motivation in self-administered modes, directly related to the probability of satisficing.

3.4.6 Item nonresponse

Item nonresponse occurs when the respondent reports no answer. It can occur early in the response process or just before the completion of the response stage, but can also be the result of a completely skipped question without any attempt to initiate the question processing. Wolfe et al. (2008) summarize various studies dealing with the causes of item nonresponse. They list several question-related factors that increase item nonresponse in different modes, including sensitivity, conditionally applicable questions, a higher number of response options, and the explicit provision of non-substantive response categories. It can be also influenced by an inappropriate questionnaire design or comprehension difficulties (de Leeuw et al. 2003).

Although item nonresponse does not match the definition of satisficing[27], it stems directly from the response process and appears to be strongly affected by the very same key variables as satisficing: task difficulty, motivation, and cognitive ability (Krosnick 1991). Comparable to satisficing, studies also relatively consistently report a tendency of higher item nonresponse among lower-educated and older respondents (Wolfe et al. 2008).

A lack of interviewer probing and interviewer-provided extrinsic motivation are considered the most important factors of generally higher item nonresponse in self-administered modes (de Leeuw 1992). Web surveys were shown to produce higher item nonresponse than face-to-face (Heerwegh and Loosveldt 2008) and telephone interviewing (Roster et al. 2004; Smyth et al. 2008), although comparisons with the latter are largely inconsistent in the magnitude of differences. However, when items are sensitive, respondents more often refuse to provide an answer in the interviewer-administered modes (Bradburn et al. 1978). Consistent effects were found in the comparison of web and telephone interviewing by Kreuter et al. (2008).

A majority of studies reported lower item nonresponse in web surveys compared to mail surveys (Stanton 1998; Klassen and Jacobs 2001; Boyer et al. 2002; Kwak and Radler 2002; Truell et al. 2002; Lorenc 2010; Shin et al. 2012). Better performance of the web mode is generally attributed to the features of computerization: automated routing (Pealer et al. 2001; Kwak and Radler 2002; Wolfe et al. 2008; Bech and Kristensen 2009) and real-time validations of answers (Klassen and Jacobs 2001; Mangunkusumo et al. 2006). This is in line with our discussion on the use of dynamic web questionnaire features to compensate for the lack of interviewer's help with routing through the questionnaire (section 3.2.1).

Some authors, however, report a higher proportion of skipped items on the web than in the mail mode (N. Bates 2001; Lozar Manfreda and Vehovar 2002a; Denniston et al.

[27] Namely, Krosnick and Alwin (1987) understand satisficing as a response strategy used by the respondent uses to produce a *seemingly* reasonable answer, which is not the case if no answer is provided at all.

2010). Although rather speculatively, this can be to some degree explained by the implementation-specific influences of computerized questionnaires on the transmission of information through the visual input channel. In section 3.1 we stressed the importance of the visibility concept and clarity of presentation for the reduction of the respondent's burden and the appropriate processing of questions. Although this is equally important with paper questionnaires, computerization brings additional specifics. For example, item nonresponse in web surveys is sometimes increased when there are a large number of questions on a single page (Lozar Manfreda et al. 2002; Peytchev et al. 2006; Toepoel et al. 2009b) or for items appearing at the lower edge of the screen (Lorenc 2010).

Two final remarks need to be made regarding the item nonresponse comparisons between modes. First, the rates of unanswered questions and corresponding differences between modes vary widely across studies. For example, Lorenc (2010) reports the mean item nonresponse rate of 1.6% for mail mode and 0.9% for the web mode, while a survey by Lozar Manfreda and Vehovar (2002a) produced the mean rates of 8% and 17%, respectively. Similarly, Roster et al. (2004) found profoundly higher item nonresponse for all questions on the web compared to the telephone, while in the study by Smyth et al. (2008) significant differences were observed on a vast minority of items.

Secondly, there is a strong indication that the question type is likely to importantly interact with the mode of administration and may be a more influential factor than the mode itself. In the mode comparison study by Borkan (2010), for example, a mean item nonresponse for scale items was 1.5% for mail and 1.2% for web mode, while for open-ended questions the proportions increased to 12.2% and 12.9%, respectively. Denscombe (2009) and Wolfe et al. (2008) point out disproportionally higher item nonresponse in the mail mode for open-ended items compared to web. Similarly, Roster et al. (2004) found the largest difference in item omission between telephone and web mode for demographic items (4.4% compared to 14.5%) and the lowest for 10-point scale questions (12.6% and 16.1%, respectively).

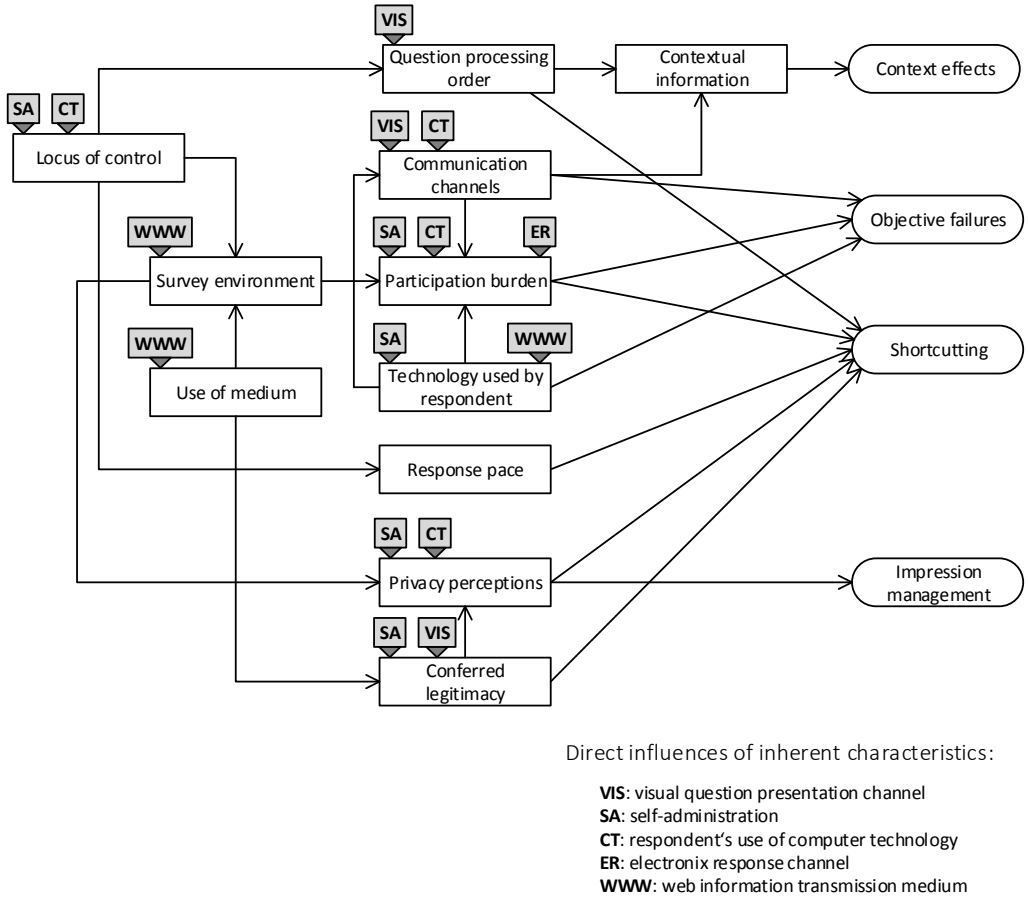
3.5 The complex structure of mode effects in web surveys

The most important conclusion derived from this extensive overview of distinctive aspects of web surveys is that inherent characteristics of the mode do not simply determine the occurrence of mode effects. They always emerge as a consequence of a particular combination of mode characteristics and external factors related to specific survey design and characteristics of respondents.

To create a big picture of sources of mode effects in web surveys, it is first necessary to establish relations between the inherent, implementation-specific, and contextual characteristics of the web mode. Each mode is best understood as the foundation that loosely defines further possibilities of building the survey project. The first layer around this basis are implementation-specific and contextual mode characteristics. As defined in section 1.1.2 (page 37) they are constrained and influenced by the mode itself, but to a large degree also vary with specific survey design approaches, respondent's characteristics, behaviours and attitudes, as well as broader social perspectives. We have summarized several partially overlapping and complementing typologies by different authors in Table 1.2 (page 36).

Tourangeau et al. (earlier also Tourangeau and Smith 1996; 2000) proposed a model which positions three psychological variables as mediating factors of mode effects, directly influenced by inherent mode characteristics: impersonality (in section 3.2.3 we referred to this as *perception of privacy*), legitimacy, and cognitive burden. They also acknowledged the mediating role of surveying pace, the order of question processing, and mental models, although they do not include them in the model explicitly. The influence of inherent mode characteristics through these mediating variables reflects in the error of the survey estimate due to a decreased level of reporting, lower accuracy and reliability, or an increased rate of missing data. Some extensions to this conceptual model present a solid basis for a summary of the complex structure of mode effects in

Figure 3.6: Relations between inherent, implementation-specific, and contextual characteristics of the web mode and their influences on measurement error



web surveys. In this chapter we turn back to some of the key conclusions from the literature review above and summarize them as part of the extended model, presented in Figure 3.6.

The large number of presented factors and their interrelations make the immediate comprehension of the model somewhat difficult. Before reviewing the details of the proposed model, we therefore present an illustrative example of impression management factors, which we addressed in sections 2.2.2, 3.2.3, and 3.3 above. We have generally concluded that the problem of impression management is reduced in web surveys compared to interviewer-administered modes, particularly due to the higher privacy perceptions fostered by self-administration. However, according to various studies discussed above, privacy perceptions in web surveys can be influenced by several other mode-related factors. The model in Figure 3.6 shows the potential influences of

such factors. Computer technology can reduce the perceived privacy by introducing the effects of media presence if highly interactive questionnaire features are used. Some respondents may also express concerns over anonymity online. The ability of the web mode to confer the legitimacy of the survey request is significantly reduced by the lack of interviewers (self-administration), but can be improved to a certain degree using appropriately visually presented questionnaire. In addition, web users often face illegitimate requests and untrustworthy information on the web, which can additionally question the legitimacy. Self-administration and computerization of the questionnaire both determine the respondent's locus of control over the surveying process, and in effect allow the respondent to choose a preferred environment in which they will complete the questionnaire. Some of these environments (e.g. public places) may not assure the respondent sufficient privacy of provided answers.

Following the logic presented in this example, we here disentangle the individual components of the model and summarize their contribution following the findings of studies, elaborated previously in this chapter.

3.5.1 Relations between characteristics of the web mode

One of the most highlighting advantages of web surveys is their flexibility, and this is reflected in numerous overlapping influences of inherent mode characteristics on implementation-specific and contextual variables. These variables are presented as rectangular shapes in Figure 3.6 with the direct influences of the inherent characteristics listed above them. As it is evident from the figure, mediating variables include those proposed by Tourangeau et al. (2000) and several others that we identified as important in this chapter.

Communication channels are defined by visual presentation as consisting of words and numbers, nonverbal symbols, and graphic paralanguage. Properties of communication channels are determined by the questionnaire design and other included visual elements, which are additionally specified by implemented computerized questionnaire

features. Furthermore, the nature of visual presentation can also be affected by properties of the device used by the respondent, for example a small screen size.

Locus of control is put completely on the respondent's side by self-administration, but can be, to some degree, restricted by interactive and dynamic features of web questionnaires, like automated skips and validations. Depending on the questionnaire programming, the respondent thus has more or less control over the questionnaire administration. This directly influences the flexibility of **response pace** as respondents can choose the amount of time they devote to each question, without external social pressures.

Question processing order is directly affected by visual presentation, which allows several questions to be visible simultaneously on the same questionnaire page. When more flexible routing through the questionnaire is allowed by the locus of control, respondents are able to simultaneously or in varying order consider information from preceding and subsequent questions. The order of question processing and used visual elements transmitted through communication channels of visual presentation can both convey various **contextual information** in the form of additional interpretative cues to the respondent. The nature of visual communication makes these contextual information much more explicit compared to auditory modes.

Web, as an information transmission medium, and computerization of the questionnaire impose specific requirements on the **technology used by respondent** to answer the questions. Still, a wide variety of devices with the capability to browse the web can be used, including personal computers, mobile phones, and tablet computers.

Use of medium refers to individually and socially perceived common practices of the medium utilization (de Leeuw 1992). Since web users often experience untrustworthy information and requests on the web, use of the web medium can be regarded as a mediating factor of lower ability of the mode to **confer legitimacy** of the survey. The problem of legitimacy is further aggravated by impersonality due to self-administration. However, professional design conveyed through the visual input channel can be used to overcome this problem to a certain degree.

Practices of web use also relate to the **survey environment**, since the web itself can be regarded as a virtual environment from which the respondent can receive a large amount of information also during participation in the survey. Other attributes of this environment are influenced by the physical setting, flexibly chosen by the respondent due to the locus of control on their side.

Privacy perceptions in web surveys are significantly increased by the sense of impersonality due to self-administration. However, they can be decreased by computerization when human-like interactive questionnaire features are used, anonymity is not assured, or the respondent has a negative attitude towards computer privacy. Perceptions of privacy can also be affected by the problem of legitimacy of survey requests on the web as well as the specific of environment in which surveying takes place.

Finally, the mode-related **participation burden** is probably the most diversely influenced mediating variable of the web mode. This mediating variable contains a somewhat broader range of aspects than the concept of cognitive burden proposed by Tourangeau and colleagues (Tourangeau and Smith 1996; Tourangeau et al. 2000). In addition to **cognitive demands** arising from mode characteristics (like required reading ability), the mode-related participation burden in web surveys can also be increased by a lack of interviewer-provided help and motivation due to self-administration, questionnaire design or technical characteristics of the device used for surveying, use of an electronic response channel and computer technology, computerized interactive features of the questionnaire (like answer validations), distractions from the environment chosen by the respondent to participate in the survey, and so on. Whether or not these influences actually increase task difficulty often depend on the respondent's ability, knowledge, and previous experiences with various aspects of the web survey process. For example, computer technology may increase task difficulty only for less experienced computer users, and written text conveyed through the communication channels of the visual input may be challenging only for respondents with lower reading abilities.

3.5.2 Deviations leading to errors

The relations between characteristics of the web mode establish the basic structure for the potential emergence of mode effects. In a vast majority of cases they are actually triggered after additional variables related to a specific survey design or respondent start interacting with this structure. A recent study by Bennink et al. (2013) demonstrated that mode differences between web and face-to-face surveys emerge only with some combinations of visual presentation, question topics, availability of non-substantive response category, order of the answer categories, mandatory answers in the web mode, and so on.

In contrast to Tourangeau et al. (2000), we propose a somewhat different set of typology of errors due to mode effects. Instead of including more general deviations in reliability and validity, we expose four key specific types of deviations identified throughout this chapter (presented as rounded shapes in Figure 3.6): context effects, objective failures to provide an accurate answer, shortcutting (satisficing and deliberate item nonresponse), and impression management. This alternative typology allows more detailed description of the nature of errors caused by mode effects, but can still be conceptually mapped to the proposal by Tourangeau and colleagues. Impression management directly corresponds to their notion of the level of reporting. Objective failures and shortcutting can both result in the increased rate of missing data. Finally, each type of deviation exposed by our model can decrease the reliability and accuracy of the obtained survey estimates.

All these types of deviations can also be caused by sources unrelated to mode, but we focus only on mode-specific causes. Furthermore, different types of errors should not be regarded as completely independent from each other. For example, context effects can be fostered by satisficing if the respondent starts relying on various contextual cues in order to shortcut through the response process. These relations are, however, not included in Figure 3.6 in order to avoid introducing additional complexity.

Context effects in web surveys are a direct result of the incorporation of additional cues transmitted through visual communication channels into the response process. The exact mechanisms are largely unknown, but crucially depend on the type of information. Examples exposed in section 3.1.3 include the content of surrounding questions, visual format of scale questions, meanings derived from colours, and information provided by images and other graphical elements. Even when such information is available, the emergence of context effects may still be inhibited by other interpretative cues, like textual labels next to coloured scales.

Objective failures to provide an accurate answer occur due to the respondent's objective inability to completely and thoroughly process the question (Knapp and Kirk 2003). This can result in either misreporting or item nonresponse. In the former case, the respondent is incapable to perform one or several cognitive stages adequately, but still derives an (inaccurate) answer. In the latter case, the response process is terminated prior to the completion of the reporting stage.

Some forms of objective failures depend on the respondent's ability and experiences. The transmission of information through visual communication channels may cause comprehension failures among respondents with low reading ability, especially with complex question wordings. Respondents may also be incapable of answering questions correctly due to the computerization of the questionnaire if they do not possess sufficient skills in using computer technology. Another set of objective failures occur because some information is never transmitted to the respondent. This can be the result of a failure of the questionnaire design to meet the visibility principle or due to technical incompatibilities of the web questionnaire with specifications of the respondent's device. In such cases information is not even considered for inclusion into the response process.

The occurrence of **impression management** is primarily related to specific topics of questions that are regarded intrusive, threatening or socially (un)desirable by the respondent. It commonly occurs in the form of social desirability bias, but we use the term

"impression management" to emphasize the respondent's purposive distortion of answers in contrast to the unrealistic self-deception (Paulhus 2002, see also section 2.2.2). According to our elaboration in section 3.2.3, privacy perceptions are one of the key variables affecting the likelihood of impression management. Self-administration thus importantly reduces the impression management tendencies in web surveys. However, as discussed above, the problem may occur due to negative influences of survey legitimacy, specific environments, negative attitudes toward computer technology, effects of media presence caused by highly interactive questionnaire features, and other factors.

The most complex, but probably also the most frequent, deviations caused by mode effects in web surveys are related to **shortcutting**. We use this as a general term to embrace deviations arising from the respondent's strategies to reduce the burden of survey participation. This includes satisficing as well as item nonresponse due to the respondent's intentional termination of the reporting stage. Among mode characteristics the error is most explicitly fostered by an increased participation burden, influenced by the factors listed above. However, we have recognized several other potential sources related to mode characteristics and specific respondent's behaviour. One example is the search for the fastest route through the questionnaire enabled by the custom question processing order enabled by the respondent's locus of control. Because the latter allows the respondent to choose the preferred response pace, they may opt for speeding through the questionnaire without thoroughly processing questions. Also, the reduced perceptions of privacy due to computerization and lower conferred legitimacy may lead respondents to resort to some form of shortcutting.

Consistent with Krosnick (1991), a respondent's ability and motivation play an important role in the emergence of mode-specific shortcutting through interactions with mode characteristics. All these three variables – ability, motivation, and task difficulty – reflect in participation burden and are interlinked. Increased task difficulty may lead to lower motivation and its perception depend on the respondent's abilities. The role of these variables is further complicated by influences unrelated to mode, for example topics of

questions that require a high amount of retrieval and judgment effort to answer. However, a lowered motivation to participate in the survey due to burden increased by such influences may ultimately reflect also in a stronger interaction with mode characteristics.

3.5.3 Susceptibility of web surveys to mode effects

The volatile nature of mode effects and consequential inconsistent findings between studies becomes unsurprising when this complex structure of causes is taken into account. However, this makes it very difficult to elaborate how critical the problem is in web surveys. The severity of the problem further depends on the type of estimated parameter. The form-resistant hypothesis claims that even significant effects on estimates of individual variables often do not change the correlations between them (Krosnick and Alwin 1987; Jäckle et al. 2006). Indeed, measurement characteristics of the web mode were often showed to be at least equal if not better than in other modes. Miller et al. (2002) observed no significant differences in reliability for a vast majority of items in face-to-face and web survey. Similar conclusions were obtained by comparisons with the mail mode (Buchanan and Smith 1999; Hertel et al. 2002; Mangunkusumo et al. 2006; Börkan 2010), while telephone comparisons generally revealed higher reliability of the web mode (Roster et al. 2004; Braunsberger et al. 2007; Chang and Krosnick 2009).

Based on our review, mode effects in web surveys generally seem to be the most strongly influenced by the downsides of self-administration. While computerization can substantially relieve the burden of routing through the questionnaire, the tasks of appropriate attendance to questions, interpretation of questionnaire-provided information, and overall performance of the response process ultimately depend on the respondent. Self-administration also limits the possibilities of providing additional extrinsic motivation to respondents. This can significantly increase the likelihood of satisficing or item nonresponse in the interaction with less motivated participants. The issue places self-administration in an unfavourable position of the moderating factor of

many mode effects stemming from other inherent mode characteristics. Virtually any mode effect that results in increased satisficing or item nonresponse can be suspected to be additionally pronounced by the problems of self-administration.

An additional set of common problems arise from abuse of the visual input channel and computerization capabilities. Overexcitement with the interactivity and visual elements can introduce measurement errors absent in other modes, or even resemble effects of interviewer-administered surveys. The careful and somewhat conservative use of these features, proposed for example by Couper (2008), is essential to reduce the potentially damaging impact on data quality.

Paradoxically, self-administration, computerization, and visual question presentation offer some of the most pronounced benefits of web surveys, including more accurate reporting on sensitive topics, control over data quality, and extensive possibilities of data collection. Understanding potential sources of mode effects in web surveys is therefore crucial to take full advantage of the mode without introducing mode-specific errors.

3.6 Methodological approaches to studying mode effects

The complexity of mode effects makes comprehensive empirical research of the problem difficult. Empirical attribution of measurement errors to specific mode characteristics requires some form of experimental design. To isolate the role of each and every inherent mode characteristic contributing to differences between modes, experiments varying in only one characteristic at a time would be needed. In many cases this is impossible to accomplish in practice due to the close relation between certain inherent characteristics and incompatibility between others. There are also combinations that can be implemented easily, but would be normally be regarded as unusual or artificial, like a web survey with an auditory instead of visual presentation of questions

(Couper 2008). Experimental research into mode effects is thus focused predominantly on commonly used survey modes.

Throughout our discussion in this chapter we have stressed the benefits of both within-mode and between-mode experimental comparisons. A combination of both offers a powerful tool to better understand mode effects and especially their mapping to individual error sources. As we stressed in the first chapter (sections 1.2.3 and 1.3), different sources may influence survey errors in opposing directions and therefore cancel out the overall effect.

3.6.1 Experimental studies

Mode effects studies are, however, more commonly associated with between-mode experimental comparisons. When interpreting the results of these studies, it is important to bear in mind the distinction of mode effects from between-mode differences (Figure 1.4 on page 60). Observed measurement difference between two modes is a strong indication of mode effects if experimental design successfully controls other potential factors as discussed below. However, in the absence of a known true value of the estimate, further (theoretical or empirical) investigation is needed to identify how each of the modes is affected by mode effects. Of course, mode comparisons can reveal only mode effects that result in differences between the compared modes, which may obscure some effects of the individual mode.

Within-mode experiments are especially suitable for studying the contribution of implementation-specific and contextual characteristics to mode effects. Of course, appropriate theoretical relations with inherent mode characteristics need to be established first. The most highlighting example addressed in the above literature review are manipulations of the visual presentation channel to observe contextual effects (section 3.1.3). According to our theoretical conceptualization, these effects are attributable to the mode, because they are essentially enabled by the visual presentation channel and cannot occur in auditory modes.

Various experimental designs are used for mode comparisons, but they roughly fall into three groups: split-sample (randomized field experiments), resurveying, and laboratory experiments. Each of them has important advantages and disadvantages. Several mode comparison studies, however, do not use any explicit experimental design. A non-exhaustive list of examples from our literature review include assignments into different groups based on the respondent's selection of survey mode (Tomsic et al. 2000; Burr et al. 2001), availability of e-mail addresses (Griffis et al. 2003; Deutskens et al. 2006), sequential mixed-mode designs for nonresponse reduction (Carini et al. 2003), or the use of entirely different samples (Roster et al. 2004; Braunsberger et al. 2007; Chang and Krosnick 2009). In general, results of these studies should be viewed with a great deal of caution as potentially uncontrolled differences in coverage, sampling, and nonresponse errors between groups can substantially affect the results (Lozar Manfreda and Vehovar 2002b).

Split-sample comparisons

In split-sample experimental designs, a sample of the target population is randomly assigned to experimental groups. A similar approach is usually used for within-mode experiments (Groves 2004). The main problem of using a split-sample design for mode comparison studies is the possibility of confounded measurement and nonresponse errors between modes (Groves 2004; Vannieuwenhuyze and Loosveldt 2012). Even when the starting assignment to groups is well-controlled, variations in unit nonresponse patterns are likely to distort the final composition of groups. This violates the assumption of fixed essential survey conditions apart from the mode itself. The problem may be especially pronounced with web surveys due to the lower availability of Internet access among some groups of general population.

The issue is usually addressed by using post-survey adjustment techniques, like propensity score weighting. This assumes a conditionally independent relation between treatment differences and the target variable given the available auxiliary information (Lorenc 2010). Guzy and Leitgöb (2012) demonstrate the importance of adjustments by

reporting significantly reduced variations between modes after including the control variables into their model.

Another disadvantage of split-sample design for studying measurement errors is its general inability to separate systematic and variable measurement errors (Groves 2004). This problem can be addressed by resurveying as an alternative approach.

Resurveying

The resurveying (test-retest or repeated measurement) design to studying mode effects is based on multiple applications of the survey to the same sample of respondents, but with a different mode of administration in each wave. By analysing only respondents to both waves, the influence of differential nonresponse error between experimental groups can be eliminated. In addition, it is possible to compare answers from each wave, and estimate and compare test-retest reliabilities (Miller et al. 2002; Alwin 2007). Because the design is substantially more complex and demanding for the researcher and respondents, there are only a few resurveying studies of mode effects in web surveys available (Miller et al. 2002; Lozar Manfreda and Vehovar 2002b; Mangunkusumo et al. 2006; Rivara et al. 2011). Chang and Krosnick (2009) performed test-retest mode comparison with repeated measurements using the same mode, but with different sample sources for each mode.

There are some important potential disadvantages of resurveying design. The respondent's participation in the second wave of surveying may be affected by the experience from the first wave. The responses to the second wave may thus be subjected to consistency effects. On the other hand, unexpected changes in the true value of target variables may occur between both waves, leading to skewed reliability estimates. However, the most critical issue of resurveying design for investigation of mode effects lies in the self-selection. We can expect that only highly motivated respondents will take part in both data collection waves to be eligible for inclusion in the test–retest analyses.

Since motivation is one of the key factors of response process deviations in self-administered surveys, the generalization of the results may lead to a significant underestimation of the problem.

Laboratory studies

Studying the differences between modes in a laboratory setting enables substantially higher control over the experimental situation, either split-sample or resurveying. Chang and Krosnick (2010) used a laboratory setting to compare responses to computerized self-administered questionnaire and interviewer-administered intercom survey. By randomly assigning the participants to experimental group upon arrival to the lab, their study enabled a complete elimination of other error sources. Jäckle et al. (2006) report on a similar experiment conducted by Gallup, where participants were randomly assigned to one of four modes and then resurveyed with another mode. Laboratory studies are to some degree also resembled by group administration of self-administered questionnaires with a random allocation to the mode groups (Denscombe 2006; Denscombe 2008; Denniston et al. 2010). Laboratory research also enables some special non-experimental observations relevant for mode effects, such as eye-tracking to identify response order effects (Galesic et al. 2008).

The clearest problem with laboratory studies is the artificiality of the surveying situation that may cause changes in the respondent's behaviour. This is even more the case for studying web surveys and other self-administered modes, which can be administered in and be affected by various environmental factors. The laboratory setting may also decrease the sense of privacy if responses are directly visible to other people in the room (Denniston et al. 2010). However, as noted by Groves (2004), laboratory studies should not be used for estimating measurement errors directly, but rather for manipulating and observing specific causes of these errors.

3.6.2 Indicators of data quality and measurement errors

A general issue of mode effects measurement, shared with virtually all research on survey errors, is related to the estimation of the bias. The true value of the variable is very rarely known or does even not objectively exist (Groves 2004), therefore a direct observation of the bias as a difference between the true value and the obtained estimate is not usually possible. Exceptions include factual data obtainable from official registers, provided these are accurate. Among studies included in our literature review, Klassen and Jacobs (2001) used company-related registry data to check the accuracy of reports, while Kreuter et al. (2008) evaluated social desirability of student answers by comparing them to the university records.

Identification of mode-specific measurement errors therefore usually requires a reliance on indirect indicators and within-mode experimental comparisons. For example, response order and question order effects are commonly identified by comparing two or more experimental groups with varying order of items or response options (e.g. Krosnick and Alwin 1987; Schwarz et al. 1991; Couper et al. 2004a; Malhotra 2008). This, of course, only indicates the presence of effects, but cannot be used to estimate the magnitude of the error. Similar often holds true for studies observing differences in scale answers and other mode-specific response sets.

In many cases variations in results between compared groups are interpreted according to theoretical expectations. This is typically done in comparative studies of impression management, where answers in a less desirable directions are considered more accurate (Tourangeau and Yan 2007; Kreuter et al. 2008). Again, this does not allow making conclusion about the magnitude of the bias, but enables the identification of less biased groups, provided the above theoretical assumption holds true[28]. Sometimes, however, more innovative approaches are used. As an interesting, yet somewhat controversial

[28] The assumption can be regarded as reasonable, but we should not forget the caution by Schwarz and Oyserman (2001) that what is desirable may vary across different groups of respondents.

alternative approach to experimental observation of social desirability, it is worth mentioning the *bogus pipeline*, a dummy device (a fake polygraph) presented to the respondent as being capable of detecting false responding (Jones and Sigall 1971). While shown to be effective under various circumstances (Roese and Jamieson 1993; Aguinis and Henle 2001), its impracticality for field work, its possible impact on participation and raised ethical objections contribute to its limited applicability for general survey research.

Another set of indicators of mode effects and between-mode differences is based on the observation of correlations, latent structures, reliability, and validity. Examples include measurement equivalence (Revilla 2013), reliability (e.g. Buchanan and Smith 1999; Hertel et al. 2002; Miller et al. 2002; Mangunkusumo et al. 2006; Chang and Krosnick 2009), latent structures (Buchanan and Smith 1999; Roster et al. 2004; Deutskens et al. 2004) and concurrent validities (Hertel et al. 2002; Chang and Krosnick 2009).

3.6.3 Limited scope of studies

We conclude this chapter by exposing some additional problems of empirical research, that arise more from the research practice than the complexity of the problem itself.

Approximately half of the published comparisons between web and other survey modes used in our literature review is conducted on student populations, and the proportion is even higher if only split-sample and resurveying experimental studies are considered. This is most likely due to the easier accessibility of student populations and high Internet penetration among them. On the other hand, very few experimental studies are done on general populations of Internet users and especially among older respondents, although these are some highlighting exceptions (e.g. Lozar Manfreda and Vehovar 2002b; Bech and Kristensen 2009; de Leeuw et al. 2010a). One of the consequences of studying substantially homogeneous populations is lacking the understanding of differential

mode effects among specific groups, like those with lower cognitive abilities, less educated, and the elderly. The importance of this issue is supported by studies on satisficing that confirm an increased susceptibility to deviations in the response process among such groups (Knäuper et al. 1997; Knäuper 1999; Malhotra 2008).

Another limitation of existing research is related to the focus on a restricted set of consequences of mode effects. A vast majority of studies take into account only some errors due to mode effects, like social desirability, differences in satisficing, or general comparisons of estimate. Very few of them consider a more comprehensive set of indicators at once to illuminate a wider range of mode effects within the study (e.g. de Leeuw 1992; Jäckle et al. 2006). A related issue is the inadequate attention to relations between contributing factors of mode effects. As we have concluded above and is demonstrated by Bennink et al. (2013), mode effects are stimulated or inhibited by a large number of different implementation-specific and contextual variables. Without in-depth empirical treatment of a larger set of these variables, the causes of mode effects are often unclear.

Chapter 4
Analysis of mode effects in the GGS Pilot survey

Up to this point we have provided a detailed general evaluation of different sources of mode effects in web surveys, which we consider the main contribution of this dissertation. The outlined complexity of factors contributing to their occurrence highlights how challenging the problem is for empirical investigation.

In the empirical part of the dissertation we turn our focus to a small subset of the problem: the manifestation of mode effects in answers to scale (grid) questions. The study is intended to be more exploratory than explanatory, with the key aim of illustrating the volatile nature of the problem by observing the consistency of selected forms of mode effects on a larger set of variables. At the same time it attempts to sharpen the understanding of often inconclusive differences in responses to scale questions between web and interviewer-administered surveys as discussed above (section 3.4.4). It also compares findings obtained using different methods to underline the importance of observing different data quality indicators in order to get a more complete picture of the problem.

We focus on four response sets that are often, but not always consistently, found to be influenced by the data collection mode: impression management, extreme and midpoint answers, and non-differentiation. The analysis is based on the observation of differences in answers by respondents randomly allocated to web, telephone (CATI), or face-to-face (CAPI) implementation of the same questionnaire. Rather than focusing only on a few target variables, we perform the analyses across a large number of scale items, which enables us to offer a more general evaluation of identified response sets.

We begin this chapter by establishing general hypotheses to guide the analysis. We then describe the main methodological background of the study, indicators of response sets,

and the most important details of analytical approaches. Finally, we present the results, comment the findings, and discuss necessary improvements for future research.

4.1 General hypotheses

In the empirical study we attempt to verify four main hypotheses. The first two hypotheses (H_1 and H_2) state the absence of negative influences of the web mode, while the second two (H_3 and H_4) state the presence of such influences. Because the theoretical background and results of previous empirical studies on which these hypotheses are based were already extensively discussed above, we only briefly justify each hypothesis here and add references to the corresponding sections.

H_1: Web respondents express lower impression management tendencies.

The likelihood of impression management in survey situation is importantly determined by the respondent's perception of privacy. The self-administered nature of web surveys generally offers respondents a higher level of privacy than interviewer-administered CATI and CAPI modes. We can therefore expect web respondents to be less reluctant in providing answers that may be regarded as less appropriate in a specific social context. See sections 2.2.2 and 3.2.3.

H_2: Web respondents are less likely to answer with lower or upper extreme scale values than CATI and CAPI respondents.

Although lacking a complete theoretical explanation, several empirical studies found a higher frequency of extreme answers in telephone and (to lesser degree) face-to-face surveys than on the web. Some of the factors discussed in section 3.4.4 include the respondent's ability to pay more attention to all answers due to self-administration, the related visual presentation of questions, and confounding with lower impression management. Because of the latter factor, we verified an additional hypothesis:

H_{2A}: Difference in extreme responses between web and interviewer-administered modes is most pronounced for questions susceptible to impression management.

H_3: Web respondents tend to select middle scale values more often than CATI and CAPI respondents.

While the tendency to select mid-answers is closely related to extreme responses as a specific response set in scale questions, it is also often assumed to be the result of satisficing. While an inadequate question processing can also result in extreme answers, the usual neutral meaning of the mid-point scale value may offer the respondent a safer option for providing apparently valid answer even when they resort to strong satisficing. Susceptibility of web surveys to this issue is mostly attributed to the double-edged sword of self-administration. Because respondents need to assume (almost) complete locus of control over the survey flow, and possibilities to offer them additional extrinsic motivation are limited, there is a higher risk of resorting to sub-optimal response strategies. This may be especially true for long questionnaires, like the one we used in this study. See sections 3.2.2 and 3.4.4.

H_4: Web respondents are more likely to resort to non-differentiation.

Assignment of the same response value to all or almost all items within a scale is one of strong satisficing strategies (sections 2.2.1 and 3.4.2). As with the previous hypothesis, we expect this to occur more frequently in web mode than in CATI and CAPI mainly due to the lower motivation of respondents to devote sufficient effort to the response process.

To aid interpretation of the results, we also analyse general patterns of differences between modes and comment on other potentially interesting and important observations.

4.2 Methodology

To verify the stated hypotheses, we used data collected within an experimental evaluation of the *Generations and Gender Survey*. Here we briefly present the survey and discuss important background methodological information relevant to our study.

4.2.1 Study description

The *Generations and Gender Programme* (GGP) is an international research infrastructure for academic research and population-related policy development. It is managed by 11 European institutes and implemented in cooperation with the *United Nations Economic Commission for Europe* (UNECE). The central activity of the programme is a large-scale comparative longitudinal *Generations and Gender Survey* (GGS), conducted in 19 countries worldwide. The survey is applied to a longitudinal panel of approximately 10,000 respondents per country in three waves with a three-year interval between them (NIDI 2011).

In 2011, a feasibility pilot study of mixed-mode GGS implementation was conducted in Slovenia. The rationale was to explore the possibilities of optimizing data collection costs by introducing the web data collection mode for some parts of population. The experiment was conducted in two phases. The first phase was aimed at estimating the magnitude of mode effects between face-to-face (CAPI), telephone (CATI), and web mode. In the second phase, various alternative mixed-mode designs were compared. Our analyses are based only on the first wave.

Questionnaire and questions

The experimental survey was based on a draft questionnaire for GGS 2015 (Aassve et al. 2011), consisting of eleven thematic modules and covering a variety of demographic and related topics. In total, the questionnaire contains approximately 340 different questions. Some of them are repeated several times (e.g. questions about each household

member), resulting in a relatively long questionnaire with an average duration of approximately 45 minutes in the face-to-face mode.

For the study presented here, we selected only single-item and multiple-item scale questions. We also included “yes”/”no” (forced-choice) questions presented in the grid format. However, we dismissed questions applicable only to some respondents (conditionally presented questions), and those included in various within-questionnaire experimental manipulations. This enabled us to make a comparison of results across several questions on the same set of respondents.

We ended up with 20 questions, containing a total of 89 items:

1. Health evaluation (1 item);
2. Personality (15 items);
3. Sense of control (5 items);
4. Happiness (1 item);
5. Loneliness (6 items);
6. Depression (7 items);
7. Income adequacy (1 item);
8. Affordable goods and services (6 items);
9. Payment inability in 12 months (4 items);
10. Religiosity (1 item);
11. Importance of religious ceremonies (3 items);
12. Planning for future (1 item);
13. Marriage and children (9 items);
14. Family risk responsibilities (5 items);
15. Childcare responsibilities (3 items);
16. Elderly-care responsibilities (4 items);
17. Gender roles (8 items);
18. Overall survey experience (1 item);
19. Survey feedback (5 items);
20. General opinion about surveys (3 items).

The corresponding scales consisted of a varying number of answer categories (values), ranging from two (“yes”/”no”) to eleven. Additional details of the selected questions, including full question wording, are presented in Appendix A.

In order to minimize the differences between the three modes, only the most essential mode-specific adjustments were made to the questionnaire. Following this principle, grid questions were presented item-by-item in both, interviewer-administered and web

versions of the questionnaire. A horizontal scrolling matrix, similar to the one shown in Figure 3.5 (page 130), was used on the web.

Sampling, allocation to experimental groups, and participation

The sample of respondents was obtained from a commercial online access panel maintained by the Slovenian marketing research company *Valicon*. Although the sampling is non-probability and cannot be used for valid inferences to the general population, it offers an important advantage of minimizing confounding non-coverage effects between modes and reaching demographically diverse population. The most prominent potential problem is a high familiarity of these respondents with web survey participation. Although Toepoel et al. (2008) found only minor differences between new and trained panel respondents regarding the incidence of response effects, we cannot rule out the possibility that different effects may be found among participants less accustomed to web surveys and the use of computer.

The company obtained all the necessary contact information (e-mail, address, and telephone number) from 743 panel members, who were randomly assigned to one of three modes. Because the threshold required number of respondents was set to approximately 200 for each mode, the additional recruitment of 72 persons for CATI and 32 for CAPI was performed during the course of data collection. The overall final response rates were 87% for the web mode, 61% for CATI, and 74% for CAPI. In total, data from 623 respondents were used for the analyses. There are very minor variations in generally low item nonresponse across the analysed items. The number of cases for each item therefore vary between 611 and 618 (Appendix C).

Although the demographic structure of respondents somewhat differed between modes, none of the differences were found to be significant (Table 4.1). However, we included demographic variables in all statistical models in order to control the potential influence of differential nonresponse.

Table 4.1: Differences in basic sample composition between modes

Characteristic	Web	CATI	CAPI	Test of differences*
Gender (%)				$\chi^2(2) = 0.90^{ns}$
Male	47.9 %	43.5 %	47.1 %	
Female	52.1 %	56.5 %	52.9 %	
Total	100.0 %	100.0 %	100.0 %	
Age				
mean	39.28	41.17	41.52	$F = 1.74^{ns}$
standard deviation	12.99	13.40	13.35	$W = 0.22^{ns}$
median	37	40	41	
Education				$\chi^2(12) = 14.70^{ns}$
Basic or less	2.8 %	2.5 %	1.5 %	
Vocational secondary	11.6 %	13.5 %	8.4 %	
Technical secondary	19.5 %	21.5 %	26.6 %	
General secondary	21.9 %	19.0 %	27.1 %	
Short cycle higher	12.6 %	15.0 %	9.4 %	
Professional higher	8.4 %	9.0 %	10.8 %	
University degree	23.3 %	19.5 %	16.3 %	
Total	100.0 %	100.0 %	100.0 %	

*Note: Tested using chi-square test (χ^2), one-way ANOVA F-test, or Levene's test (W).
ns not significant, $p \geq 0.1$

4.2.2 Indicators of response sets

In order to test the stated hypotheses, specific indicators of response sets were needed. For extreme and midpoint answers the indicators are simple binary variables, with a value of 1 if the target answer (lower extreme, upper extreme, or mid-value) is selected and 0 otherwise. However, the detection of impression management and non-differentiation requires somewhat more elaborate approaches.

Impression management

An analysis of impression management is ideally supported by a prior identification of potentially susceptible items, and known expected scale direction to which answers shift under the influence of impression management (e.g. towards a higher agreement with

a specific statement). For some topics, like drug use or school performance, this can be done with a relatively high confidence, but in many other cases the situation is much less clear. After the direction is identified, the common assumption is that the mode with a higher tendency toward these answers is more affected (Bradburn et al. 1978).

To identify susceptibility to impression management, each item in the questionnaire was rated by three survey methodology experts using the coding scheme for identification of sensitive questions (included in Appendix B). The scheme was developed by the GGP project team to distinguish between three types of question sensitivity as proposed by Tourangeau et al. (2000): intrusiveness, fear of disclosure, and social desirability (see section 2.2.2). Because we were interested in purposive impression management, we focused on the last indicator only, measured as a potential for over-claiming in terms of presenting oneself in a more favourable light (Nederhof 1985; Paulhus 2002).

The potential for over-claiming was rated on a three-point scale, with the value 1 indicating no potential and the value 3 a high potential. We averaged the ratings across all three experts and marked the item as potentially susceptible to impression management if the mean rating was 2 or higher. The majority (76%) of 89 items included in our analysis were found to be potentially susceptible. An overall agreement between experts, measured using Krippendorff's alpha, was moderate ($\alpha_K = 0.51$) and indicate the presence of high variation among the raters for some of the items. However, we accepted this level of agreement as acceptable and did not consider the variation between experts further in the analyses.

Because a large part of analysed items ask respondents about their opinion on different topics, it is often difficult to identify scale points with a higher likelihood of selection under the influence of impression management. Depending on which social values and norms are invoked by a respondent during the response process, answers may be often shifted to either side of the scale (Schwarz and Oyserman 2001; Näher and Krumpal 2012). Without taking the risk of considerable over-guessing, we ascribed the assumed direction to 49 items. However, among them are no items with value-related topics for which determination of the most likely direction is especially problematic.

Non-differentiation

Under the satisficing model (Krosnick and Alwin 1987; Krosnick 1991), non-differentiation is often understood in a literal sense as the lack of (any) differentiation between scale values. Such complete non-differentiation can be measured in a straightforward way by discovering cases with a zero standard deviation of answers across all scale items. On the other hand, some authors expose benefits of using more sensitive methods that take into account different levels of differentiation (e.g. McCarty and Shrum 2000; Jäckle et al. 2006; Heerwegh and Loosveldt 2008; Chang and Krosnick 2009). This seems reasonable as some respondents may differentiate between scale values for only few scale items.

To measure the level of differentiation within each scale, we used the probability of differentiation index as proposed by Linville et al. (1989; also used by McCarty and Shrum 2000; Heerwegh and Loosveldt 2008):

$$P_d = 1 - \sum_{i=1}^{k} p_i^2 \qquad \textit{Eq. 4.1}$$

where k is the number of answer categories, and p_i the proportion of scale items with selected answer category i. The index is defined in the range $[0, P_{d_{max}}]$, where 0 represents the complete non-differentiation (the respondent assigns the same scale value to each item), and $P_{d_{max}}$ the maximum differentiation. Because the value of $P_{d_{max}}$ depends on the number of items and scale values, it is not readily clear when maximum possible differentiation is achieved. For the sake of clarity, we normalized the index values to the range [0, 1] by taking into account the theoretical maximum number of different answers within the scale[29]. However, direct comparability of the index between different scales remains questionable as the number of theoretically possible values within this normalized range still depends on scale characteristics.

[29] Already intuitive reasoning brings us to the conclusion that assigning different answers to all items is not possible when the number of items exceeds the number of scale values. For example, if the scale consists of six items and five scale values, at least two items will essentially have the same value.

Level of differentiation should be interpreted cautiously as higher differentiation index does not necessarily indicate better data quality. On the contrary, even equal answers to all scale items may represent a true value for some respondents.

4.2.3 Approach to analysis

Similar to a majority of other empirical studies, the grounds for our exploration of mode effects is the analysis of differences between web and both interviewer-administered modes (CATI and CAPI). Because the true value of any estimated parameter (θ) is unknown, the decision about the effects of which mode are causing the between-mode difference in estimates, $\Delta\left(\theta'_{M_1}, \theta'_{M_2}\right)$, needs to be predominantly theoretically-driven. For example, if a variable has a significantly higher mean in the known direction of impression management in mode *A*, we would most likely conclude that a higher impression management susceptibility of the mode *A* partially accounts for the discovered difference.

While we cannot avoid relying on such theoretical expectations for an interpretation of results, we devote special attention to summarizing the effects across all items. The main advantage of this approach is the ability to observe the consistency of the effects, but at the cost of a somewhat less thorough consideration of individual items and their content. A similar methodology was used before by Jäckle et al. (2006) in their comparison of telephone and face-to-face modes. Here we present some statistical techniques and considerations that may be important to understand and interpret the obtained results.

Models for observing between-mode differences

To observe between-mode differences, we used ordinary least squares (OLS) regressions, logistic regressions, and partial proportional odds models. All effects were controlled for basic socio-demographic structures (gender, age, and higher education) to reduce the potential confounding influence of differential unit nonresponse. The

analyses were performed using Stata 11 (StataCorp 2012) with some additional plug-in packages.

While ordinal variables of scale questions are often seen suitable for ordinary least squares (OLS) regressions, it is generally more appropriate to use models with an explicit recognition of the ordinal nature of these variables (Winship and Mare 1984). Jäckle et al. (2006) showed that ordinal-level analysis helps to reveal more detailed patterns of mode effects. We therefore use partial proportional odds modelling to observe differences in obtained answers between modes in addition to OLS regressions. Because this approach is somewhat less known and less frequently used, we clarify some of its main principles.

The proportional odds modelling (McCullagh 1980), also known as ordered logit, is based on a cumulative sequential partitioning of ordinal values into dichotomous categories. For example, for four ordinal values (*A*, *B*, *C*, *D*) of the dependent variable, dichotomizations are done in the sequence *A–BCD*, *AB–CD*, and *ABC–D*. To each of these dichotomizations the logistic regression model is then applied.

The central assumption, however, is that logistic regression coefficients of independent variables are the same for all dichotomizations (Peterson and Harrell 1990; Long 1997; O'Connell 2006). Studies of mode effects on scale questions highlight a potential for violation of this parallel lines assumption: telephone respondents are more likely choose extreme positive scale categories (Taylor 1999; Roster et al. 2004; Christian et al. 2007a; Dillman et al. 2009), while web respondents more often gravitate toward middle or negative categories (Roster et al. 2004; de Leeuw et al. 2010b). Mode effects may thus differentially influence some rather than all response categories (Jäckle et al. 2006).

One of the generalizations of the ordered logit model that relaxes the parallel lines assumption is partial proportional odds modelling (Peterson and Harrell 1990). Williams (2006) presents the generalized ordered logit model for an ordinal dependent variable Y with M categories and one independent variable X as:

$$P(Y_i > j) = \frac{\exp(\alpha_j + X_i\beta_j)}{1 + \exp(\alpha_j + X_i\beta_j)}, \qquad j = 1,2, \dots, M-1 \quad \textit{Eq. 4.2}$$

Of course, more than one independent variable can be included in the model. In case of K independent variables ($X_1, X_2, \dots, X_K$), we can rewrite the above equation as:

$$P(Y_i > j) = \frac{\exp(\alpha_j + X_{1i}\beta_{1j} + X_{2i}\beta_{2j} + \cdots + X_{Ki}\beta_{Kj})}{1 + \exp(\alpha_j + X_{1i}\beta_{1j} + X_{2i}\beta_{2j} + \cdots + X_{Ki}\beta_{Kj})}, \quad j = 1,2, \dots, M-1 \quad \textit{Eq. 4.3}$$

Under the parallel lines assumption of the proportional odds model, values of β_{kj} are the same across categories j of a dependent variable and only values of the intercept α_j may vary. In contrast, the partial proportional odds model allows specifications in which values β_{kj} are the same across categories of j for some independent variables X_k, but vary for others (Peterson and Harrell 1990; Williams 2006).

For our application of partial proportional odds models, we use the *Gologit2* package for Stata, developed by Williams (2006) as an improved version of an earlier package by Fu (1998). It includes features for the automated detection of independent variables that violate the parallel lines assumption. For such variables values of β, coefficients are allowed to vary across categories of the dependent variable while for others they remain fixed. This automated fitting of a model in some regards resembles data mining techniques and also its author acknowledges a potential danger of falsely detected violations of the assumption due to chance. However, in our case there is not sufficient theoretical knowledge available to establish appropriate hypotheses about which independent variables can be expected to violate the hypothesis. We therefore rely on empirical tests of violations.

Significance levels for multiple testing

In our analyses we fit one or more of the above models for each of 89 selected scale items. One important problem with the simultaneous testing of a large number of null hypotheses is the increased probability of false rejection (type I) error occurrence. A

common and simple, yet very conservative, way of dealing with this problem is the Bonferroni adjustment, where each of k hypotheses is tested at α/k. Although this approach works well for testing a small number of hypotheses (Rice 2007), it becomes problematic if k is large. Substantially lowered values of adjusted significance levels when testing numerous hypotheses essentially reduce the power of a test to detect investigated effects. Perneger (1998) also criticizes the Bonferroni method as being of little use for researchers since it tests the typically irrelevant general null hypothesis that all individual null hypotheses are true.

One of more recent approaches to tackle the issues of Bonferroni and similar methods is the *false discovery rate* (FDR), proposed by Benjamini and Hochberg (1995). The authors define FDR as the expected proportion of falsely rejected null hypotheses. A basic assumption is that errors due to falsely rejected null hypotheses are less critical when a large number of the hypotheses is truly false (Benjamini and Yekutieli 2001). In contrast to the Bonferroni method, FDR therefore allows for some type I errors at the level of individual hypotheses. To what degree critical values are adjusted at the chosen threshold depends on the obtained p values from individual tests: a larger number of low p values lead to less conservative adjustment.

The decision on how conservative or liberal to be with the significance level adjustments in multiple testing can become rather philosophical, strongly depending on the view of the individual researcher as well as on the purpose of a study. For example, Bender and Lange (2001) question the need for multiple testing adjustments in predominantly exploratory studies like the one we present in the dissertation. At the same time they emphasize the importance of adequate reporting of the utilized approach. This view is also supported by Rothman (1990, 43) who argues that scientists “should not be so reluctant to explore leads that may turn out to be wrong that they penalize themselves by missing possibly important findings”. We therefore arbitrarily set a general threshold for

interpretation of results as significant at $\alpha = 0.01$, but also report significance at adjusted levels using the Benyamini-Yekutelli method (α_{yek}) and more conservative Bonferroni method (α_{bnf}) where applicable[30].

Effect sizes

For some comparisons we additionally reported effect sizes for the proportion of variance in the target estimates accounted for by mode as well as mean differences between modes. We use partial η^2 as a measure of an effect size for explained variance, and Glass's delta (Δ_G) for an effect size of mean difference (Ellis 2010; Richardson 2011). The latter is calculated as the difference of adjusted means divided by the standard deviation of the control group, which is in our case the web, as the reference mode for comparison with CATI and CAPI:

$$\Delta_G = \frac{\bar{X}_C^* - \bar{X}_W^*}{s_W} \qquad \textit{Eq. 4.4}$$

An interpretation of whether a specific effect size is large or small is mainly the matter of content as well as individual researcher's perceptions and needs. Following a very generic proposal by Cohen (1988) we use the threshold values of $|\Delta_G| \approx 0.2$ for a small, $|\Delta_G| \approx 0.5$ for a medium, and $|\Delta_G| \approx 0.8$ for a large effect. The corresponding thresholds for partial η^2 are approx. 0.01, 0.06, and 0.14, respectively. For some descriptive comparisons of effects between different groups of items we summarized the effect sizes using simple averages (Turner and Bernard 2006; Maynard et al. 2007).

[30] We calculated the adjusted α values using the *yekutieli* method of *Multproc* package for Stata by Newson and The ALSPAC Study Team (2003). According to the author, the method does not assume independence between p values and allows for arbitrary correlations between them, but is more conservative than some other FDR methods.

4.3 Analyses and results

We begin our analyses by observing general differences in mean estimates and distributions between modes, and briefly compare results obtain using the two different modelling approaches. The results serve us for verifying the stated hypotheses in the remainder of the chapter.

4.3.1 Review of differences in answers

Many users of survey data and also a majority of current mode comparisons primarily focus on the estimation of means. We therefore begin by analysing mode differences in means of target variable. For this analysis we considered all variables with four or more scale values (answer categories) in the form of an ordered scale. Because of the large total number of items, only those with a significant mean difference between either web and CATI or web and CAPI ($p < 0.01$) are presented in Table 4.2. The original regression coefficients and model statistics for all items are listed in Appendix C.1.

Table 4.2: Differences in adjusted means between modes and effect sizes for items with four or more scales values and a significant difference ($p < 0.01$) between web and either CATI or CAPI

	Web	CATI compared to web		CAPI compared to web	
Item	$\bar{X}_m$	Δ_{C-W}	Δ_G	Δ_{C-W}	Δ_G
(7.05) **Personality**: 1 does not apply /…/ 7 applies perfectly					
b) Does thorough job	5.490	+0.200	0.165	+0.470##	0.389
c) Talkative	5.232	+0.269	0.183	+0.451#	0.307
h) Outgoing, sociable	4.999	+0.498##	0.364	+0.668##	0.488
j) Values artistic, aesthetic experience	4.571	-0.034	-0.023	+0.579##	0.392
n) Relaxed	4.719	+0.467#	0.339	+0.320	0.232
(7.06) **Sense of control**: 1 strongly agree /…/ 5 strongly disagree					
a) Cannot solve own problems	3.568	+0.354#	0.318	+0.407##	0.366
b) Feel pushed around	3.635	+0.287	0.261	+0.435#	0.395

(*Table continued on next page*)

	Web	CATI compared to web		CAPI compared to web	
Item	$\bar{X}_m$	Δ_{C-W}	Δ_G	Δ_{C-W}	Δ_G
(7.09) **Depression**: 1 seldom or never /…/ 4 most or all of the time					
a) Could not shake off blues	1.407	-0.284##	-0.445	-0.015	-0.023
b) Felt depressed	1.332	-0.220##	-0.426	-0.082	-0.158
c) Thought life is a failure	1.351	-0.241##	-0.568	-0.167#	-0.394
d) Felt fearful	1.526	-0.364##	-0.674	-0.181#	-0.335
e) Felt lonely	1.486	-0.313##	-0.502	-0.165#	-0.265
f) Had crying spells	1.293	-0.240##	-0.501	-0.099	-0.206
g) Felt sad	1.642	-0.351##	-0.571	-0.126	-0.205
(10.02) **Income adequacy**: 1 with great difficulty /…/ 6 very easily					
Making ends meet	3.402	+0.227	0.184	+0.435##	0.353
(11.04) **Importance of religious ceremonies**: 1 strongly agree /…/ 5 strongly disagree					
b) Religious wedding	3.782	+0.403#	0.340	+0.160	0.135
(11.07) **Planning for future**: 1 I plan f. fut. as much as possible /…/ 10 I just take each day as it comes					
Planning for future	4.700	-0.296	-0.116	-0.903##	-0.354
(11.08) **Marriage and children**: 1 strongly agree /…/ 5 strongly disagree					
b) Living unmarried together all right	2.005	-0.040	-0.040	-0.284**	-0.288
d) Divorce having children all right	1.838	+0.273#	0.254	+0.036	0.034
h) Woman w/o stable relationship with man having a child	2.489	-0.217	-0.210	-0.385##	-0.373
(11.11) **Elderly-care responsibilities**: 1 strongly agree /…/ 5 strongly disagree					
b) Children should adjust work to parents' needs	3.367	+0.108	0.108	+0.277#	0.277
c) Children should financially help parents	2.402	+0.307#	0.333	+0.143	0.154
(11.12) **Gender roles**: 1 strongly agree /…/ 5 strongly disagree					
a) Women really want home and children	3.363	-0.181	-0.153	-0.311#	-0.262
c) Man's task earning, woman's family	4.054	+0.118	0.136	+0.261#	0.302
d) Not good if woman works, man cares for children	3.593	-0.003	-0.003	-0.320**	-0.239
e) Working woman same relation with child	2.212	-0.235	-0.233	-0.415##	-0.412
h) Family life suffers because men too concentrated on work.	2.798	+0.375#	0.295	+0.199	0.157

(*Table continued on next page*)

Item	Web	CATI compared to web		CAPI compared to web	
	$\bar{X}_m$	Δ_{C-W}	Δ_G	Δ_{C-W}	Δ_G
(12.02) **Survey feedback**: 1 definitely not /…/ 5 definitely yes					
a) Questions difficult	1.731	-0.099	-0.125	-0.276#	-0.347
c) Questions made think	3.697	-0.513##	-0.375	-0.626##	-0.458
e) Questionnaire too long	1.991	+0.847##	0.714	-0.073	-0.062
Mean $\|\Delta_G\|$ (significant items)[a)]	-	-	|0.439|	-	|0.350|
Mean $\|\Delta_G\|$ (all items) [b)]	-	-	|0.171|	-	|0.188|

Notes:

a) Calculated as the simple average of the absolute effect sizes (Glass's Δ) across items with the significant corresponding mode coefficient in the OLS regressions. See page 170 for further discussion of the calculation.

b) Calculated as the simple average of the absolute effect sizes across all items in OLS regressions. All model coefficients are listed in Appendix C.1

- Control variables: gender, age, and higher education.
- $\bar{X}_m$: the marginal mean from the OLS regression; Δ_{C-W}: difference between the marginal means of web and each of the compared mode (equals the regression coefficient b of the compared mode).
- $^{**}\ p < 0.01$, $^{\#}\ p < \alpha_{yek} = 0.0052$, $^{\#\#}\ p < \alpha_{bnf} = 0.0006$. Coefficients for the items with $p \geq 0.01$ on both mode comparisons are listed in Appendix C.1.

Among all 73 compared items with four or more scale values, the means significantly differ at $p < 0.01$ between web and CATI on 16 items (22%), and between web and CAPI on 20 items (27%). Six web items significantly differ from both compared modes.

Differences between web and interviewer-administered modes are in general comparably small if all analysed items are taken into account. The mean absolute Glass's Δ is 0.171 for comparisons of the web mode with CATI and 0.188 for comparisons with CAPI. However, some of the items in Table 4.2 exhibit more prominent effects. The largest effect ($\Delta_G = 0.714$) is found between web and telephone respondents on the question about the questionnaire length (Q12.02E). Unsurprisingly, the telephone respondents experienced the lengthy questionnaire as "too long" to a larger degree than web respondents. Other medium to large effect sizes ($|\Delta_G| > 0.5$) are identified in web-CATI comparison for a majority of depression scale items (Q7.09), with telephone respondents claiming on average lower frequency of all depression symptoms. Effects of a similar size for web-CAPI comparison occur less frequently. The most highlighting differences include higher self-portrayal of CAPI respondents as being outgoing and sociable (Q7.05H), and higher reporting of web respondents about "being made to think by questions" (Q12.02C).

Analysis with partial proportional odds modelling (GO-logit[31]) offers an alternative insight into potential mode effects by exploring differences at the level of individual scale values. It also allows for the inclusion of variables with a lower number of answer categories, for which mean estimation would make sense. We therefore attempted to fit the models for all 89 selected items, but the estimation process did not converge for nine of them. This presumably occurred due to too low or even zero cell frequencies for some answer categories (Williams 2006).

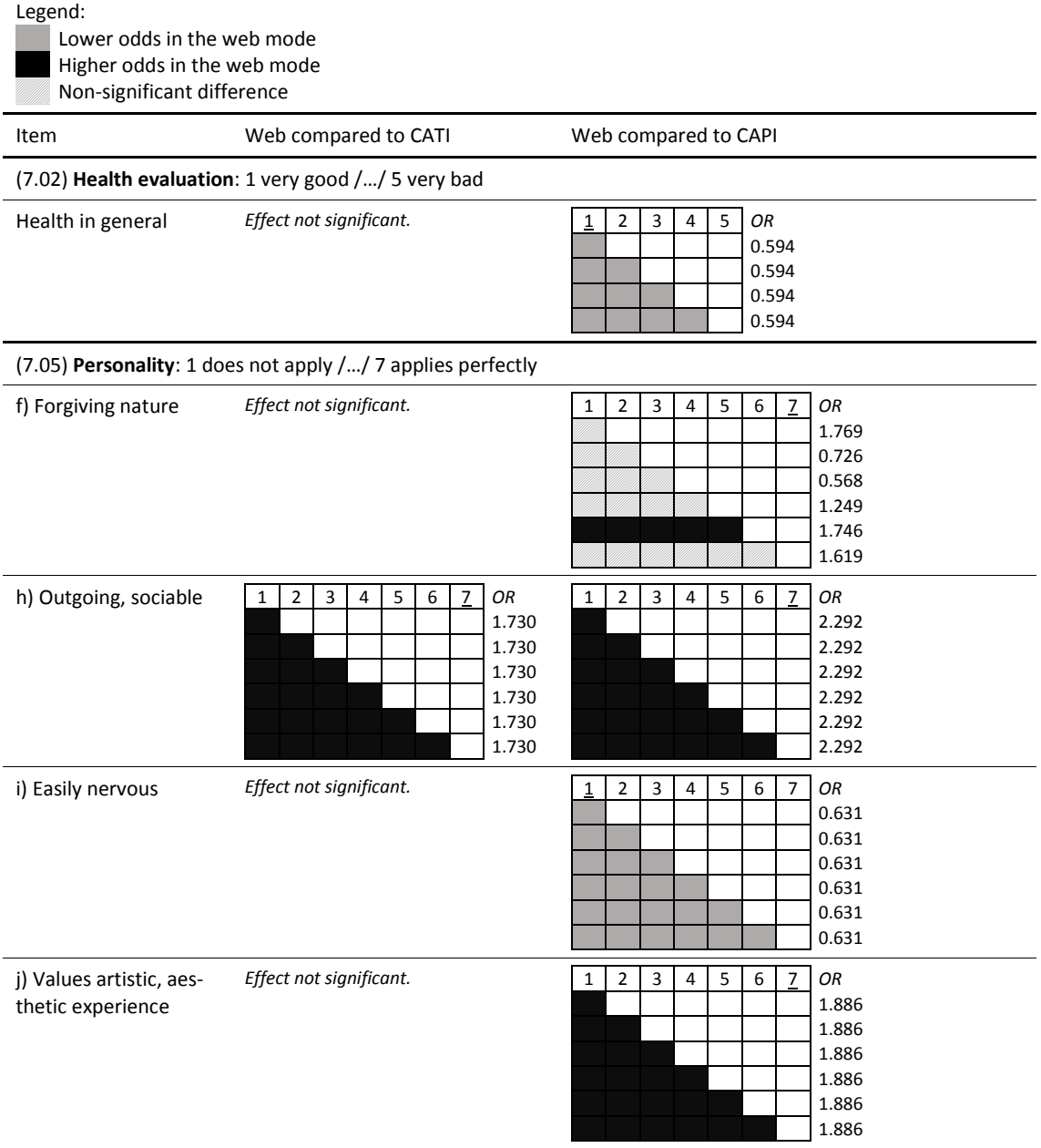
Figure 4.1 graphically presents differences between web and both interviewer-administered modes. Only items with a significant difference at $\alpha = 0.01$ are presented on the figure, while details for other items are provided in Appendix C.2. The black colour denotes that web respondents have higher odds of being positioned within the shaded range of variable values than the respondents of compared mode. Conversely, the grey colour represents lower odds of web respondents to be within the shaded value range against the respondents of compared mode. (Underlined response values on the figure mark the assumed response under extreme impression management, to which we refer later in this chapter).

The cumulative principle of proportional odds modelling in Figure 4.1 requires somewhat careful interpretation. The values of odds ratios are based on between-mode ratio of odds for cumulatively selecting one of the shaded answers against one of the unshaded ones. For example, for question Q7.02 the odds of selecting answer 1 against answers 2, 3, 4, and 5 are significantly lower in web than in CAPI mode, so are the odds of selecting answers 1 and 2 against 3, 4, and 5, etc. The resulting interpretation is that web respondents are less likely to select lower scale values generally across the whole scale range, or equivalently, web respondents are generally more likely to select higher scale values. For some items this parallel lines assumption does not hold true. Item Q7.05F, for example, shows that web respondents are more likely than CAPI respondents to select values 1–5 than the two upper extreme values, indicating a lower

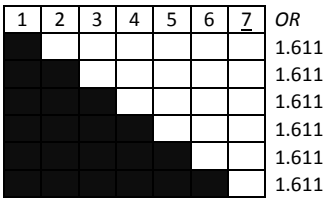
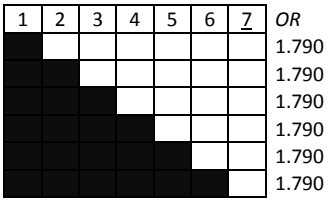
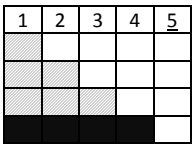
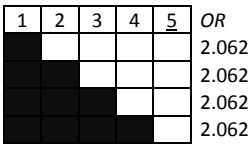
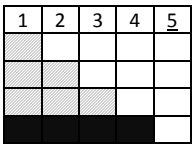
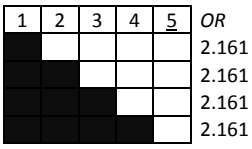
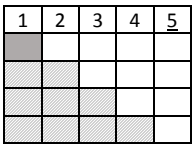
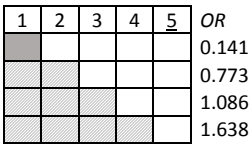
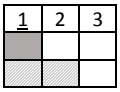
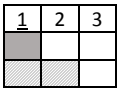
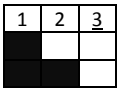
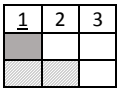
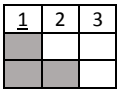
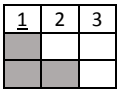
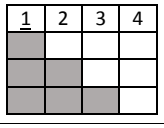
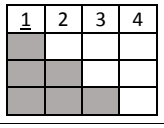
[31] We use GO-logit (generalized ordered logit) as an abbreviation for the partial proportional odds model, although the method is technically only one of possible generalizations of proportional odds models.

likelihood of perceiving themselves as having a very forgiving nature. However, differences are not significant for other answer categories.

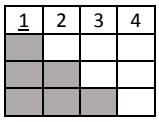
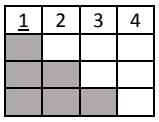
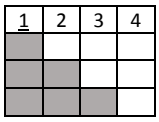
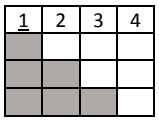
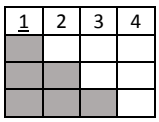
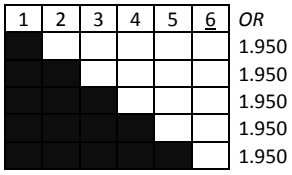
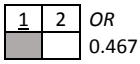
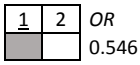
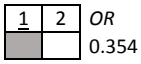
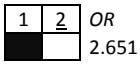
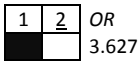
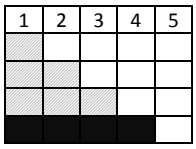
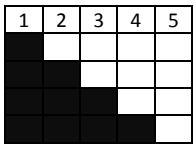
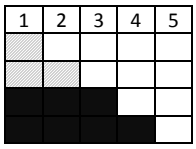
Figure 4.1: Comparison of differences in answer distributions for items with a significant effect of mode ($p < 0.01$)



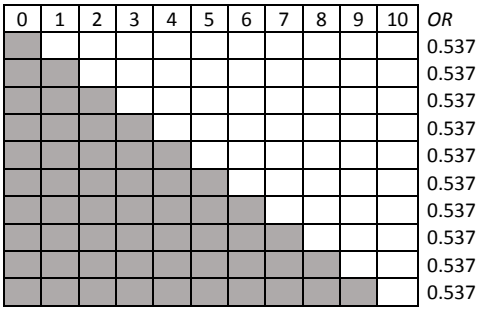
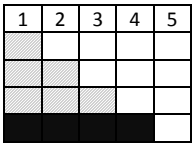
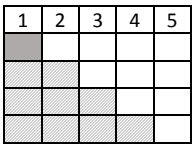
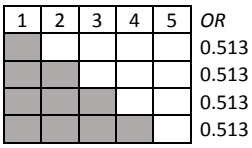
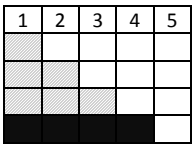
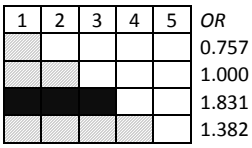
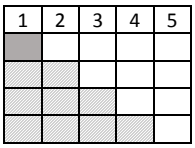
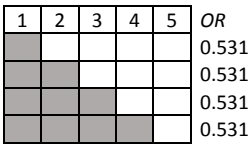
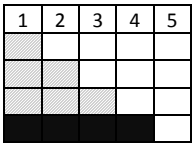
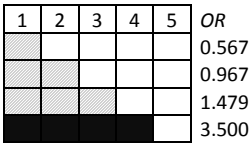
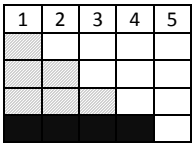
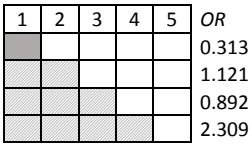
(*Figure continued on next page*)

Item	Web compared to CATI	Web compared to CAPI
k) Considerate and kind	*Effect not significant.*	
n) Relaxed		*Effect not significant.*
(7.06) **Sense of control**: 1 strongly agree /…/ 5 strongly disagree		
a) Cannot solve own problems		
b) Feel pushed around		
c) Little control over things		
(7.08) **Loneliness**: 1 yes / 2 more or less / 3 no		
a) Plenty of people to lean on		
b) General sense of emptiness	*Effect not significant.*	
d) Many people to count on		*Effect not significant.*
f) Enough close people		
(7.09) **Depression**: 1 seldom or never /…/ 4 most or all of the time		
a) Could not shake off blues		*Effect not significant.*
b) Felt depressed		*Effect not significant.*

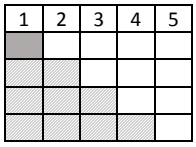
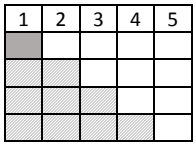
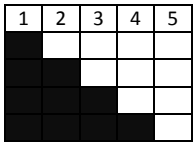
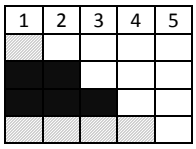
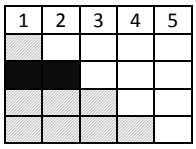
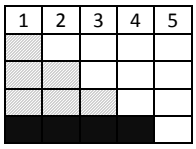
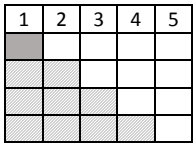
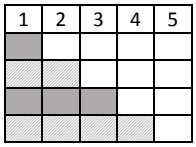
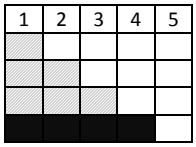
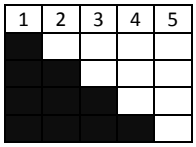
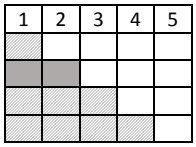
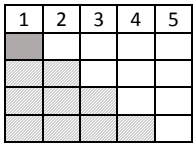
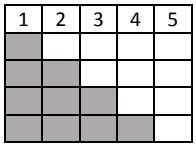
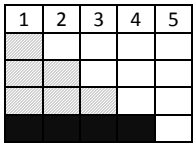
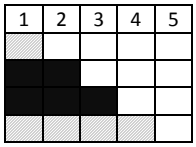
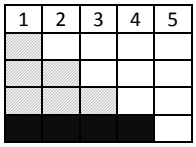
(*Figure continued on next page*)

Item	Web compared to CATI	Web compared to CAPI
d) Felt fearful		
e) Felt lonely		
g) Felt sad		*Effect not significant.*
(10.02) **Income adequacy**: 1 with great difficulty /…/ 6 very easily		
Making ends meet	*Effect not significant.*	
(10.03) **Affordable goods and services**: 1 yes / 2 no		
b) One-week holidays	*Effect not significant.*	
c) Furniture replacement	*Effect not significant.*	
f) Monthly dining out		
(10.04) **Payment inability in 12 months**: 1 yes / 2 no		
c) Utilities	*Effect not significant.*	
d) Loans	*Effect not significant.*	
(11.04) **Importance of religious ceremonies**: 1 strongly agree /…/ 5 strongly disagree		
a) Infant registered in religious ceremony		*Effect not significant.*
b) Religious wedding		*Effect not significant.*
c) Religious funeral		*Effect not significant.*

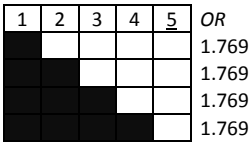
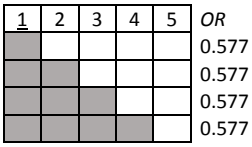
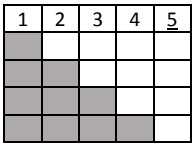
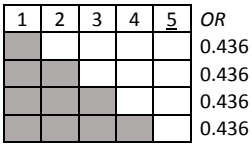
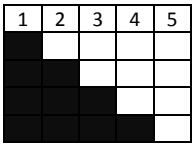
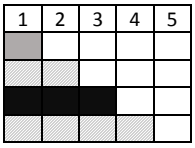
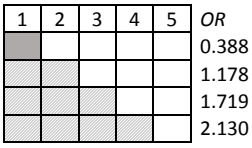
(*Figure continued on next page*)

Item	Web compared to CATI	Web compared to CAPI
(11.07) **Planning for future**: 1 I plan f. fut. as much as possible /…/ 10 I just take each day as it comes		
Planning for future	*No significant effect.*	
(11.08) **Marriage and children**: 1 strongly agree /…/ 5 strongly disagree		
a) Marriage outdated		*Effect not significant.*
b) Living unmarried together all right		
c) Marriage should not end		
g) Mother and father needed for happy child		*Effect not significant.*
h) Woman w/o stable relationship with man having a child	*Effect not significant.*	
(11.10) **Childcare responsibilities**: 1 strongly agree /…/ 5 strongly disagree		
a) Grandparents should help childcare		
c) Parents should adapt life to help adult children		

(*Figure continued on next page*)

Item	Web compared to CATI	Web compared to CAPI
(11.11) **Elderly-care responsibilities**: 1 strongly agree /…/ 5 strongly disagree		
a) Children should care for parents	 *OR* 0.517 1.179 1.512 5.576	 *OR* 0.531 1.186 1.246 2.970
b) Children should adjust work to parents' needs	*Effect not significant.*	 *OR* 1.698 1.698 1.698 1.698
c) Children should financially help parents	 *OR* 0.643 2.319 2.597 4.136	 *OR* 0.708 1.880 1.210 2.503
d) Children should live with parents for care	*Effect not significant.*	 *OR* 7.793 0.709 1.465 3.123
(11.12) **Gender roles**: 1 strongly agree /…/ 5 strongly disagree		
a) Women really want home and children	 *OR* 0.173 0.616 1.145 1.830	 *OR* 0.095 0.799 0.561 0.800
c) Man's task earning, woman's family.	 *OR* 0.370 0.821 1.109 1.772	 *OR* 1.802 1.802 1.802 1.802
d) Not good if woman works, man cares for children	*Effect not significant.*	 *OR* 0.507 0.444 0.627 1.129
e) Working woman same relation with child	 *OR* 0.443 0.783 0.921 0.759	
g) Family life suffers if mother works	 *OR* 0.385 0.701 1.084 1.930	*Effect not significant.*
h) Family life suffers because men too concentrated on work	 *OR* 0.585 1.933 2.162 2.699	 *OR* 0.378 1.137 1.767 3.382

(*Figure continued on next page*)

Item	Web compared to CATI	Web compared to CAPI
(12.01) **Overall survey experience**: 1 very unpleasant /…/ 5 very enjoyable		
Feeling about survey participation	*Effect not significant.*	
(12.02) **Survey feedback**: 1 definitely not /…/ 5 definitely yes		
a) Questions difficult	*Effect not significant.*	
c) Questions made think		
e) Questionnaire too long		*Effect not significant.*
(12.05) **General opinion about surveys**: 1 strongly agree /…/ 5 strongly disagree		
c) Surveys enable own opinion articulation		

Legend:
- Lower odds in the web mode
- Higher odds in the web mode
- Non-significant difference

Note:
- Only the results for models with at least one significant effect ($\alpha < 0.01$) are shown.
- Underlined scale values indicate the expected direction of impression management tendencies (i.e. socially desirable answer).

The GO-logit approach is particularly beneficial because it exposes answer categories with the strongest reflection of differences between web and each compared mode. The common patterns of differences indicated in the above figure are summarized in Table 4.3, where we take into account directions of scales and statements[32].

[32] It is important to note that the identification of pattern types depends on the selected α level. At more liberal levels, more odds ratios would be significant and may somewhat change the conclusions. However, a quick inspection of the original tables (Appendix C.2) reveal little substantive difference at $\alpha = 0.05$.

Table 4.3: Patterns of differences in answers between Web and CATI/CAPI on the basis of partial proportional odds models (GO-logit)

1. Generally higher responses in the web mode[*]

(Lower odds of selecting cumulative low values across the whole scale range in the web mode.)

Web–CATI comparison	Web–CAPI comparison
Web respondents are generally *more likely* than CATI respondents to report: • Not having enough close people (Q7.08F). • Being less able to shake off morning blues (Q7.09A), feeling depressed (Q7.09B), fearful (Q7.09D), lonely (Q7.09E), and sad (Q7.09G) more. • Inability to afford monthly dining out (Q10.03D). • Being made to think by the questions more (Q12.02C).	Web respondents are generally *more likely* than CAPI respondents to report: • Having worse health (Q7.02). • Being more easily nervous (Q7.05I). • Claim not to have enough close people (Q7.08F). • Feeling fearful (Q7.09D) and lonely (Q7.09E) more. • Inability to afford one-week holidays (Q10.03B), furniture replacement (Q10.03C), and monthly dining out (Q10.03D). • Taking each day more as it comes instead of Planning for future (Q11.07). • Less agreement and more disagreement toward living together unmarried (Q11.08B), woman having a child without stable relationship with a man (Q11.08H), and equality of relationship between working woman and her child (Q11.12E). • Finding questions more difficult (Q12.02A), and being made to think by the questions more (Q12.02C).

2. Generally lower responses in the web mode[*]

(Higher odds of selecting cumulative low values across the whole scale range in the web mode.)

Web–CATI comparison	Web–CAPI comparison
Web respondents are generally *more likely* than CATI respondents to report: • Being less outgoing and sociable (Q7.05H), and being less relaxed (Q7.05N). • More agreement and less disagreement with the importance of religious wedding (Q11.04B). • Not finding questionnaire too long (Q12.02E).	Web respondents are generally *more likely* than CAPI respondents to report: • Being less outgoing and sociable (Q7.05H), less value artistic and aesthetic experience (Q7.05J), and being less considerate and kind (Q7.05K). • Being more agreeable and less disagreeable that they cannot solve own problems (Q7.06A) and feel being pushed around (Q7.06B). • Experiencing sense of emptiness (Q7.08B). • Having more difficulties making ends meet with their income (Q10.02). • Being unable to pay for utilities (Q10.04C) and loans (Q10.04D) in last 12 months. • Being more agreeable and less disagreeable that children should adjust work to the needs of their parents (Q11.11B) and that man's task is earning while woman's task is family (Q11.12C). • Finding the overall survey experience more unpleasant and less enjoyable (Q12.01).

(*Table continued on next page*)

3. Less extreme responses at high values in the web mode

(Higher odds of selecting cumulatively lower responses than a high extreme in the web mode, non-significant odds ratios on other scale values.)

Web–CATI comparison	Web–CAPI comparison
Web respondents are *less likely* than CATI respondents to: • Extremely disagree that they cannot solve own problems (Q7.06A) and feel pushed around (Q7.06B). • Extremely disagree that it is important for an infant to be registered in a religious ceremony (Q11.04A), that marriage is outdated (Q11.08A) and should not end (Q11.08C), grandparents should help childcare (Q11.10A), parents should adapt life to help adult children (Q11.10C), man's task is earning while woman's task is family (Q11.12C), and that family life suffers if mother works (Q11.12G).	Web respondents are *less likely* than CAPI respondents to: • Extremely disagree that grandparents should help childcare (Q11.10A), children should live with parents for care (Q11.11D), and that family life suffers because men are too concentrated on work (Q11.12H).

4. Less extreme responses at low values in the web mode[*]

(Lower odds of selecting low extreme than cumulatively higher responses in the web mode, non-significant odds ratios on other scale values.)

Web–CATI comparison	Web–CAPI comparison
Web respondents are *less likely* than CATI respondents to: • Strongly agree that they have little control over things (Q7.06C). • Claim with certainty to have plenty of people to lean on (Q7.08A), and many people to count on (Q7.08D). • Strongly agree that living unmarried together is all right (Q11.08B), mother and father are needed for a happy child (Q11.08G), children should care for parents (Q11.11A), what women really want is home and children (Q11.12A), and that a working woman have the same relation with her child (Q11.12E).	Web respondents are *less likely* than CAPI respondents to: • Strongly agree that they have little control over things (Q7.06C). • Claim with certainty to have plenty of people to lean on (Q7.08A). • Strongly agree that parents should adapt life to help adult children (Q11.10C), children should care for parents (Q11.11A), and that surveys enable own opinion articulation (Q12.05C).

(*Table continued on next page*)

5. Other patterns

- Web respondents less likely disagree and strongly disagree that religious funeral is important (Q11.04C)
- Web respondents are more likely agreeable or neutral that children should financially help parents (Q11.11C).
- Web respondents are more likely agreeable or neutral that family life suffers because men are too concentrated on work (Q11.12H)
- Web respondents less likely strongly agree, but are also less likely disagreeable that surveys enable own opinion articulation (Q12.05C)
- Web respondents less likely answer 6 or 7 (applies perfectly) for having a forgiving nature (Q7.05F).
- Web respondents are more likely to disagree or strongly disagree that marriage should not end (Q11.08C).
- Web respondents are more likely to be agreeable that children should financially help parents (Q11.11C).
- Web respondents are less likely to strongly agree, but are also less likely to be disagreeable that what women really want is home and children (Q11.12A).
- Web respondents less likely agreeable that it is not good if woman works and man cares for children (Q11.12D).

* Note: "Yes"/"No" questions (10.03 and 10.04) with only two categories are included in the first or the second pattern. In case of two categories, either answer is by definition also extreme.

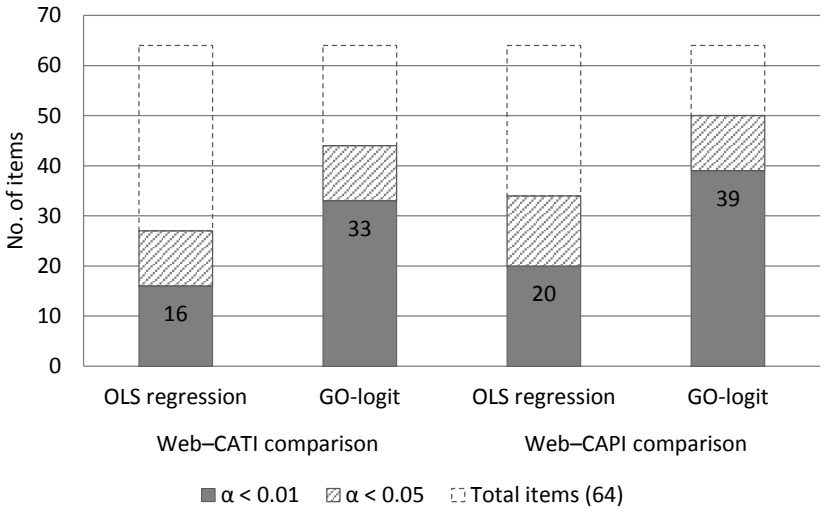
A quick inspection of results already tentatively supports our first two hypotheses on lower impression management and lower extremeness of responses in web mode. However, before further examining these topics, it is worthwhile making a brief comparison of results obtained using the two methods of analysis.

Because the GO-logit models did not converge for all items, and OLS regressions were used only for items with four or more answer categories, we compare the results for 64 items on which both models were successfully estimated. Overall, GO-logit models identified a substantially higher number of items with a significant difference between modes (Figure 4.2). The ordinal-level approach reveals 13 more items with a significant difference when comparing web and CATI, and 17 more items for a comparison between web and CAPI. On the other hand, all items significant in OLS regressions were also found significant by GO-logit models.

While statistical significance may not be a very good indicator of an effect's importance, the differences in results may indicate a higher sensitivity of analysis at the level of individual response categories. This is not unexpected as mode can potentially affect only a specific part of the variable's distribution without significantly altering the mean. Such effects go unnoticed in mean comparisons.

Indeed, a closer examination of the results from both models reveals that a vast majority of discrepancies in findings occur on the items characterized by a lower presence of extreme answers in web mode (3rd and 4th pattern of difference in Table 4.4). As shown in the table, these items also exhibit lower detected effect sizes in OLS regressions, measured as the proportion of variance in the target estimates explained by the mode (partial η^2).

Figure 4.2: The number of items with a significant effect of mode variable on means (OLS) and response distribution (GO-logit)



Note: The comparison is based on 64 items on which both models were estimated. This includes all items with four or more scale values, except nine items on which the GO-logit model did not converge.

Table 4.4: The number of items with significant mode effect and mean effect sizes by pattern of difference

Pattern of difference (web compared to other mode)	Significant ($\alpha = 0.01$) mode effects (OLS / GO-logit)		Mean η_P^2 of mode variable in OLS	
	Web–CATI	Web–CAPI	Web–CATI	Web–CAPI
1. Generally higher responses	6 / 6	8 / 10	0.044	0.018
2. Generally lower responses	4 / 4	7 / 9	0.033	0.019
3. Less extreme responses at high values	1 / 9	0 / 3	0.005	0.005
4. Less extreme responses at low values	0 / 6	0 / 4	0.003	0.002
5. Other	2 / 4	2 / 5	0.012	0.007

Note: The results are based on 64 items on which both models were estimated. This includes all items with four or more scale values, except nine items on which the GO-logit model did not converge.

These gaps in findings between the two analytical techniques are important beyond mere statistical curiosity. They demonstrate how (non)detection of mode differences can strongly depend on the type of estimated parameter. Suppose, for example, the interest in mean estimation of a (dis)agreement with the statement that it is "the task of a man to earn money and the task of a woman to look after home and family" (item Q11.12C). In this case we would find a small and non-significantly higher disagreement with the statement among CATI than web respondents (+0.12 on five-point scale). However, if our main interest was to observe the proportion of specific answers to the same question, we would find significantly more respondents (13 percentage points) who strongly disagree with the statement in CATI than in web mode.

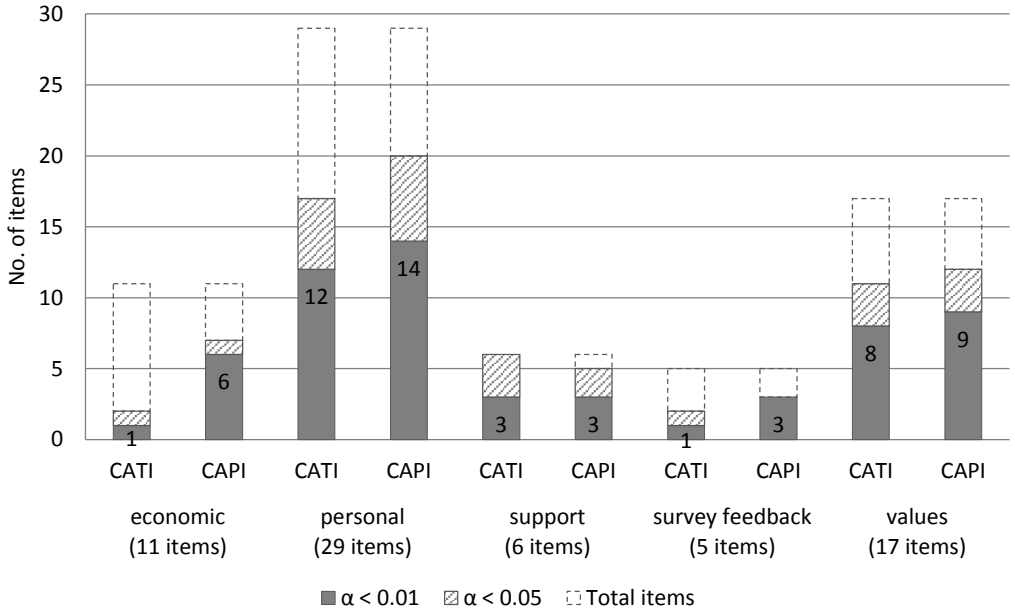
4.3.2 Impression management

Evaluators rated 68 items as potentially prone to impression management. Differences between modes were found to be significant at $\alpha = 0.01$ (either by OLS or GO-logit models) on 37% of these items in web–CATI comparisons and on 51% of them in web–CAPI comparisons. Figure 4.3 confirms that a large part of this discrepancy can be attributed to the varying behaviour of respondents in both interviewer-administered modes across different question topics. This variation is particularly large for questions regarding respondents economic issues. Compared to the web mode, CAPI respondents were significantly less likely to report their own economic issues on six items, while CATI respondents on only one item. This indicates that the physical presence of an interviewer substantially increased respondents' reluctance to admit personal financial issues.

Differences in impression management tendencies between modes can be most thoroughly explored for 49 items for which we were able to assume the most socially desirable answer. These answers are indicated by underlined scale values in Figure 4.1.

Items with significant differences between web and interviewer-administered modes from the GO-logit model strongly support the hypothesis of lower incidence of impression management on the web. An evaluation of Figure 4.1 confirms that in a vast

Figure 4.3: The number of items prone to impression management with a significant effect of mode variable by topic



Notes:

- The items rated as potentially prone to impression management were included.
- Significant effects found by either OLS or GO-logit model are counted.

majority of cases web respondents show significantly lower odds of selecting more socially desirable answers, either across all categories (pattern types 1 and 2) or primarily at the extreme values (types 3 and 4). In total, only two items with significant mode difference deviate from this (Table 4.5): compared to both interviewer-administered modes, web respondents have lower odds of strongly agreeing to have a little control over things (Q7.06C), and generally claim to have been made to think more by questions (Q12.02C).

Lower impression management tendencies of the web mode also reflect in mean differences, calculated using OLS regressions for items with four or more scale values. The only item with a significantly higher mean in a more socially desirable direction on the web is Q12.02C. Overall, the average effect sizes of mean difference (Table 4.6) are small across all items with assumed most desirable answers, and small to medium when considered only for items with a significant mean difference between modes.

Table 4.5: The Number of items with an assumed impression management direction and a significant difference between web and CATI/CAPI according to the identified pattern of difference in GO-logit

Pattern of difference (web compared to other mode)	Assumed impression management direction			
	Higher value		Lower value	
	Web–CATI	Web–CAPI	Web–CATI	Web–CAPI
1. Generally higher responses	**7**	**9**	1	1
2. Generally lower responses	0	0	**2**	**10**
3. Less extreme responses at high values	0	0	**2**	**0**
4. Less extreme responses at low values	**2**	**1**	1	1
5. Other	0	0	0	1
Total	9	10	6	13

Notes:
- The results are shown for items with a significant mode coefficient in the corresponding GO-logit model, for which the impression management direction was assumed.
- Bolded values represent the number of patterns indicating lower impression management tendencies in the web mode.

Table 4.6: The mean effect size of mode on mean estimates for items with assumed impression management direction

	Mean Δ_G (number of items)	
	Web–CATI	Web-CAPI
Items with significant effect ($\alpha < 0.01$)	0.394 (11)	0.297 (11)
Items with significant and non-significant effect	0.147 (33)	0.169 (33)

Notes:
- Calculated using the OLS models for items with four or more scale values and assumed impression management direction.
- Mean Δ_G was calculated by adjusting the sign in line with the assumed impression management direction. Larger positive values denote a higher tendency towards impression management.

The use of web mode from the perspective of impression management reduction is especially beneficial for some items. The most obvious advantage over telephone interviewing is observed for depression scale items (question 7.09), where absolute effect sizes (Δ_G) of mean difference range from 0.426 to 0.674. Interestingly, face-to-face

respondents are much less reluctant to admit depression symptoms, although still significantly more than web respondents (maximum $|\Delta_G| = 0.394$). An opposite example are finance-related questions about affordable goods (Q10.03) and payment inability (Q10.04), where the difference against web is higher in CAPI than in CATI mode.

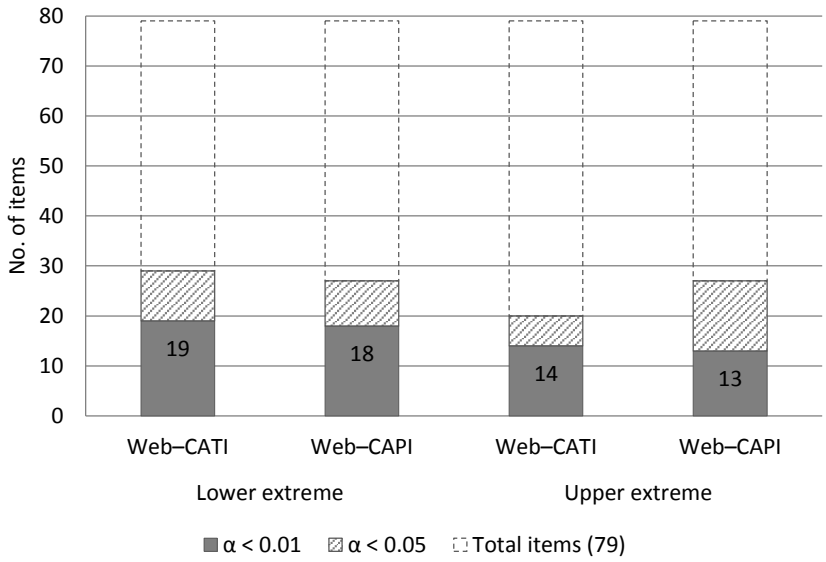
Finally, a further review of described differences in Table 4.3 indicated a relatively consistent performance of the web mode, also on opinion items for which we did not anticipate the most socially desirable scale values. The most highlighting is the more traditional (or less liberal) position of web respondents on a majority of opinion items about marriage (Q11.08) and gender roles (Q11.12).

While the results strongly confirm our hypothesis of lower impression management in web mode, this should not lead us to regard this mode as immune to the problem. As noted above, significant differences were observed for only about a third of items prone to impression management compared to CATI and a half of them compared to CAPI. One possible explanation is that these items are actually less susceptible to impression management. Another possibility is the increased tendency of web respondents to select socially desirable answers to these items, although there is no clear reason for such expectation. However, further investigation and more sensitive item evaluation methodology would be required to offer a usable non-speculative explanation.

4.3.3 Extreme and middle answers

We compared the level of extreme answer selection between web and both interviewer-administered modes for 79 items with at least three answer categories. The analysis is thus not performed for ten “yes”/”no” items of questions on affordable goods and services and payment inability (Q10.03 and Q10.04, respectively). An exploration of mid-answer selections additionally only applies to items with a midpoint value, further excluding seven items of depression scale (Q7.09) and the question on income adequacy (Q10.02).

Figure 4.4: The number of items with a significant effect of mode on lower or upper extreme answers



Note: The differences between modes were tested using logistic regression models.

For data analysis we used logistic regressions with three binary dependent variables indicating whether a respondent selected the lower extreme, upper extreme, or middle response category for a particular item. As with other models, we included gender, age, and higher education as control variables. The use of logistic regression on indicators of extreme responses differs from the GO-logit approach as it is not influenced by the distribution of non-extreme answers – it only tests whether or not odds of selecting the target answer are higher in either mode.

Because the size of tables with all model coefficients is relatively large and brings little added value apart from transparency of performed analyses, we only report summary results in the text. The relevant tables with detailed information are included in Appendix C.3 and Appendix C.4.

Extreme responding

Patterns of differences in answers between web and interviewer-administered modes, obtained using GO-logit models (Figure 4.1 and Table 4.3), give some interesting initial information on web mode performance in terms of extreme answers. The first revealing observation is the lack of pattern types in the direction of more extreme answers in web

Table 4.7: Median and mean odds ratios for extreme responses by impression management (IM) susceptibility of items

	Median (mean) OR for lower extreme answers		Median (mean) OR for upper extreme answers		Number of items
	Web–CATI	Web–CAPI	Web–CATI	Web–CAPI	
All items	1.680 (2.310)	1.409 (1.978)	1.323 (1.709)	1.331 (1.498)	79
Not prone to IM	1.565 (1.580)	1.362 (1.428)	2.062 (2.931)	1.540 (1.820)	21
Lower answer desirable	2.707 (3.019)	1.721 (1.700)	0.883 (0.885)	1.000 (1.020)	17
Upper answer desirable	1.107 (1.683)	1.000 (1.512)	1.328 (1.441)	1.620 (1.650)	22

Notes:
- The results are shown across all items with significant or nonsignificant effect of the mode variable.
- $OR > 1$: higher odds of selecting an extreme answer in the compared mode, $OR < 1$: higher odds of selecting an extreme answer in the web mode.
- Recommended consideration of median values due to high outlying odds ratios on some items.

mode. Secondly, of 33 items with identified significant difference in answer distribution between web and CATI, the most pronounced effects occur due to lower odds of extreme value selection in web mode for 17 items. In web-CAPI comparison, the lower tendencies of web respondents to select extreme answers are less pronounced and very much determine 8 out of 39 significant differences.

Logistic regressions on indicators of extreme responses for all 79 items offer further support to these observations. Significant odds ratios ($\alpha = 0.01$) of selecting lower extreme answers were found on 19 items (24%) in a comparison between web and CATI and on 18 items (23%) between web and CAPI items (Figure 4.4). Interviewer-administered modes exhibit a higher likelihood of extreme answers in all but one of these cases. The exception is the item about questionnaire length, where web respondents were more likely than CATI respondents to give an extreme answer that the questionnaire was “definitely not” too long (12.02E).

Results are not much different for upper extreme responses. CATI and CAPI respondents were found to have significantly higher odds of selecting upper extreme answers than web respondents on 14 items (18%) in web-CATI and on 13 items (17%) in web-CAPI

comparison. For none of these items are the odds of selecting an upper extreme answer significantly higher on the web.

Table 4.7 contributes some additional findings regarding differences in extreme answers. It summarizes average and median[33] odds ratios across items by taking into account significant and nonsignificant effects. For items not prone to impression management, an especially highlighting difference of less extreme responding in web mode is found for upper extreme answers against the CATI mode (median odds ratio 2.1). Unfortunately, it is not possible to pinpoint the exact reason for this, although it may indicate higher recency effects among telephone respondents.

The second important indication from Table 4.7 is the relation between impression management and extreme responding. Unsurprisingly, the observed effects are highest for extreme responses identified as more socially desirable, where a lower extremity of answers by web respondents syndicate with their lower tendency toward impression management. On the other hand, a lower extremity of web mode largely disappear or even tends to reverse for extreme answer categories opposite to the desirable end of a scale. It seems that if web respondents tend to select extreme answer categories more than telephone and face-to-face respondents, this most likely occurs due to reduced impression management. However, as stated above, none of these effects are significant.

Midpoint answers

The midpoint scale value is available to respondents in 71 analysed items. A significant effect ($\alpha < 0.01$) of mode variable was found in six (9%) comparisons with CATI and seven (10%) comparisons with CAPI[34]. The only case in which either of interviewer administered modes have significantly higher odds of selecting midpoint answer is the

[33] Exceptionally high odds ratios in a few items relatively strongly affect overall means of odds ratios. We therefore prefer to use median values for interpretation.

[34] If we disregarded the adjustment of statistical significance for multiple testing and choose $\alpha = 0.05$, we would obtain significant odds ratios for 23% of items in web–CATI comparisons and 27% in web-CAPI comparison.

Table 4.8: Median and mean odds ratios for mid-value answers by impression management (IM) susceptibility of items

	Median (mean) OR for middle answers		Number of items
	Web–CATI	Web–CAPI	
All items	0.786 (0.871)	0.797 (0.851)	71
Not prone to IM	0.884 (0.898)	0.941 (0.971)	21
Prone to IM	0.782 (0.859)	0.758 (0.800)	50
Prone to IM, 5 or more scale values	0.787 (0.914)	0.784 (0.872)	44

Notes:
- The results are shown across all items with significant or nonsignificant effect of mode variable.
- $OR > 1$: higher odds of selecting middle answer in the compared mode, $OR < 1$: higher odds of selecting extreme answer in web mode.
- Recommended consideration of median values due to high outlying odds ratios on some items.

neutral position of CAPI respondents towards the statement that "children should financially help parents" (Q11.11C). All other significant effects show higher middle answer selection among web respondents.

The mechanisms behind these effects are not very clear. Judging by the median odds ratios (Table 4.8), the higher tendency of web respondents to select midpoint answers further increases with questions prone to impression management. This remains the case also after removing six items with less than five scale values. Unfortunately, these results tell us little about whether a more pronounced midpoint answering is the result of specific deviations in the response process of web respondents, or merely a consequence of extremity that reduces the middle value selection in the compared modes.

Some tentative indication can be obtained by re-examining the results of GO-logit models of differences in the distribution of responses across the scale (Figure 4.1 and Table 4.3). The review of patterns of differences provides very little indication that variations in mid-point selection would profoundly affect distribution of responses. The pattern we would expect to observe when the difference between modes is prominently reflected

in the middle response value is demonstrated by the comparison of response distribution between web and CAPI on item Q11.08C (see the figure). However, this is the only item that shows such a pattern of difference.

4.3.4 Scale non-differentiation

We concluded the analysis of mode differences by comparing the proportion of non-differentiated answers and the amount of answer differentiation at the level of individual scales. We calculated the normalized differentiation index as described in section 4.2.2 (page 165) for each respondent across the items belonging to the common scale. This analysis, of course, does not apply to single-item scales. Complete non-differentiation occurs when the value of normalized differentiation index equals 0, meaning that a respondent answered with the same scale value to all items within the scale. A comparison of the proportion of respondents with complete non-differentiation between web and both interviewer administered mode was conducted using logistic regressions. For the comparison of mean normalized differentiation index, OLS regressions were used. Table 4.9 presents the results for both measures of non-differentiation.

The proportion of respondents with complete non-differentiation significantly differs at $\alpha = 0.01$ between web and CATI on two out of 14 scales, and between web and CAPI on four scales. The proportion of non-differentiating respondents is statistically significantly higher in both interviewer administered modes on the depression scale (question Q7.09). Face-to-face respondents also exhibit significantly higher non-differentiation on both scales related to their personal economic situation (Q10.03 and Q10.04). On the other hand, a statistically significant larger proportion of complete non-differentiating respondents is found in web mode on the importance of religious ceremonies scale (Q11.04).

An analysis of the differentiation index also reveals mixed situation. The mean value of the index is significantly higher in web than in CATI mode on the depression scale (Q7.09), but lower on the childcare responsibility scale and gender roles scale (Q11.10

Table 4.9: The percentage of complete non-differentiating respondents and differentiation index by mode

	% of respondents with complete non-differentiation			Mean normalized differentiation index[a)]		
Scale	Web	CATI	CAPI	Web	CATI	CAPI
(7.05) Personality	0.00	0.00	0.00	0.854	0.832	0.852
(7.06) Sense of control	16.22	16.44	20.72	0.516	0.537	0.494
(7.08) Loneliness	1.53	0.99	0.46	0.780	0.778	0.789
(7.09) Depression	24.98	60.81$^{\#\#}$	40.45$^{\#\#}$	0.420	0.198$^{\#\#}$	0.336$^{\#\#}$
(10.03) Affordable goods and services	52.72	58.84	65.06**	0.357	0.306	0.261**
(10.04) Payment inability	72.63	81.47	87.85$^{\#\#}$	0.221	0.152	0.097$^{\#\#}$
(11.04) Importance of religious ceremonies	61.16	48.57**	45.60$^{\#\#}$	0.279	0.363	0.398$^{\#\#}$
(11.08) Marriage and children	0.93	0.00	0.00	0.813	0.801	0.809
(11.09) Family risks responsibilities	11.37	9.47	11.32	0.610	0.646	0.602
(11.10) Childcare responsibilities	35.58	24.86	36.42	0.455	0.573$^{\#\#}$	0.471
(11.11) Elderly-care responsibilities	9.49	3.51	7.80	0.649	0.699	0.659
(11.12) Gender roles	2.49	0.00	0.93	0.714	0.771$^{\#\#}$	0.760**
(12.02) Survey feedback	0.00	0.00	1.47	0.758	0.776	0.761
(12.05) General opinion about surveys	2.34	3.01	2.46	0.796	0.780	0.771

Notes:

a) Differentiation index is normalized to [0, 1], but the number of theoretically possible values within this interval depends on the number of items in scale and the number of scale values. Direct comparison of values between items may not be appropriate.

- Logistic regressions used to test differences for proportion of non-differentiated answers and OLS regressions for mean differentiation index. Control variables: gender, age, and higher education.
- Reported are marginal means and proportions from logistic/OLS regressions.
- Significances reported for web–CATI and web–CAPI comparison:
  $^{**}\ p < 0.01$, $^{\#\#}\ p < \alpha_{bnf} \approx 0.004$

and Q11.12, respectively). In comparison with CAPI, web respondents show significantly more differentiation on the depression scale and both economic-related questions (Q10.03 and Q10.04), while lower on the importance of religious ceremonies scale (Q11.04), and gender roles scale (Q11.12).

While the obtained results are quite inconclusive, they also reveal some interesting consistencies with previous findings. Web respondents show significantly higher differentiation and lower proportion of complete non-differentiation on scales which were found to be most profoundly affected by impression management tendencies: the depression scale in both interviewer-administered modes, and scales on personal economic issues in CAPI. The higher non-differentiation is here most likely a direct

consequence of impression management as all socially desirable answer categories are located on the same side of the scale.

An evaluation of directions of all significant and non-significant differences reveals that web respondents tend to differentiate less on a majority of values-related scales (q. 11.04–11.11), with the exception of the marriage and children scale. However, differences are in most cases too small and do not indicate generally higher satisficing behaviour of web respondents.

4.4 Summary of findings and limitations

The results of the empirical study offer mixed support for our general hypotheses, but are mostly consistent with a large part of existing research on scale questions and encouraging from the viewpoint of data quality in web surveys.

We analyzed 20 scale questions with a total of 89 items from a pilot version of the *Generations and Gender Survey* questionnaire. Our main interest was to identify the consistency of differences in answers between the web survey and both interviewer-administered modes for four indicators of response sets in scale questions: impression management, extreme responding, midpoint answers, and non-differentiation.

Differences between web and both compared modes were found to be comparably small on a majority of analysed items. Mean estimates significantly differed on 22% of items compared to CATI, and on 27% of items compared to CAPI. Mean effect sizes, measured using Glass's Δ were 0.171 and 0.188, respectively. A substantially higher number of significant effects was found in the analysis of distribution of responses at the level of individual answer categories, conducted using partial proportional odds modelling. This approach revealed that on several items only particular responses are significantly affected by mode. This was most often the result of lower odds of extreme low or extreme high scale point selection among web respondents. These findings also

encourage researchers to avoid relying predominantly on the comparison of means for the detection of mode effects.

4.4.1 Verification of general hypotheses

Our analysis of differences in impression management strongly support the first hypothesis that *web respondents express lower impression management tendencies*. Significant differences on items rated as potentially prone to impression management were observed in 37% comparisons between web and CATI and 51% comparisons between web and CAPI. An analysis of 49 items, for which we were able to identify the most desirable response categories, revealed that when significant differences between compared modes occur, web respondents are less inclined toward impression management.

We also found strong support for the second hypothesis that *web respondents are less likely to answer with lower or upper extreme scale values than CATI and CAPI respondents* and that the *difference in extreme responses between web and interviewer-administered modes is most pronounced for questions susceptible to impression management.* This was indicated already by a general observation of differences in response distribution using partial proportional odds modelling and further confirmed by analysing proportions of selected lower and upper extreme responses. A significant difference in likelihood of extreme response selection was found on 24% of items in web–CATI and 23% of items in web–CAPI comparison. The effects were in the direction of lower likelihood of extreme answer selection for all but one of these items. The difference between web and compared modes in extreme answer selection was especially prominent for items prone to impression management, confirming the second part of the stated hypothesis.

However, only limited support was found for the hypothesis that *web respondents tend to select middle scale values more often than CATI and CAPI respondents.* The likelihood of midpoint answer selection significantly differed on 9% of items compared to CATI and 10% of items compared to CAPI. In all but one of these cases web respondents were

more likely to select a mid-point answer. However, it remains unclear whether these effects are caused by response deviations among web respondents or by the lower likelihood of mid-answer selection in the compared modes due to higher tendency to select extreme answers. Patterns of differences in response distributions show little support for the former explanation.

Inconclusive findings were obtained also with regard to the last hypothesis that *web respondents are more likely to resort to non-differentiation.* Differences between modes in non-differentiation performance were found to strongly depend on an individual scale. Respondents to interviewer-administered modes exhibited lower differentiation for scales containing items that were previously found to be most affected by impression management. On the other hand, web respondents differentiated less primarily on opinion and value-related questions, but differences between modes were predominantly small. Therefore the results do not allow us to confirm the hypothesis of higher non-differentiation due to satisficing in web mode. One of the possible explanations for this finding lies in the use of the horizontal scrolling matrix format of scale questions, which does not resemble the more traditional grid layout.

In sum, of four main hypotheses we were able to relatively firmly confirm two, both indicating an *absence* of specific mode effects on scale questions in web mode. The answers of web respondents were found to be less prone to impression management, less extreme, but without a confirmed tendency to resort to mid-point selection or non-differentiation. Although there is some indication of a higher presence of the latter form of satisficing on the web for some scales, the effect is generally small.

4.4.2 Consistency with previous studies

The results are generally in line with previous research; unfortunately, partially also from the aspect of providing inconsistent findings.

Findings on impression management in web surveys agree with a vast majority of other studies on social desirability and sensitive questions in web surveys compared to other

modes (Lozar Manfreda and Vehovar 2002b; Jäckle et al. 2006; Kreuter et al. 2008; Chang and Krosnick 2009; Tourangeau et al. 2013). Of course, as with other studies on this issue, the underlying assumption is that higher reporting of less desirable answers gives more accurate results (Bradburn et al. 1978). Also, while we cannot rule out the possibility of potentially different results on questions for which we did not try to assume the most desirable response (mostly value-related questions), there is little theoretical and empirical grounds to expect inconsistent findings. In general, the results thus strengthen the advantageous position of web mode over telephone and face-to-face surveys on sensitive and socially desirable topics.

Consistent with previous research are also the findings about less extreme responses in web mode compared to the telephone mode (Taylor 1999; Roster et al. 2004; Christian et al. 2007a; Dillman et al. 2009; de Leeuw et al. 2010b). In contrast to the study by Heerwegh and Loosveldt (2008), however, we also found significant differences between web and face-to-face interviewing. Since showcards were used for a majority of scale questions in the CAPI mode, the differences are therefore most likely not due to the question presentation channel, but other factors. Considering more pronounced differences of the corresponding items, impression management seemed to play an important role in heightening the extremity in both interviewer-administered modes, supporting the observations by Ye, Fulton, and Tourangeau (2011). However, significant differences were found also on other items, without clear directional patterns of extremity. Although there is some tentative indication of recency effects in telephone mode, additional experimental research will be needed to further explain causes of extremity in scale questions.

Finally, both midpoint answering and non-differentiation follow inconsistent findings between some existing empirical studies (Fricker et al. 2005; Heerwegh and Loosveldt 2008; Chang and Krosnick 2009). Large differences in non-differentiation between modes were generally in favour of web mode and were most likely caused by higher impression management tendencies of telephone and face-to-face respondents. Similarly to the study by Klausch et al. (2012), who also used a horizontal scrolling matrix

format to present scale questions, there is therefore generally little support of such satisficing-related behaviour.

While our results are certainly encouraging for web surveys, further research is needed to overcome some important limitations of the current study as briefly presented below.

4.4.3 Limitations of the study

One of the key issues of the presented study is its inability to disentangle various confounding effects. Although we performed the analysis across a large number of items, the variations in their characteristics were too small to allow for the isolated identification of some effects. For example, the majority of items was rated as susceptible to impression management using rather simple binary measure. This resulted in a strong confounding of impression management with extreme answers and, more importantly, non-differentiation.

Another important limitation is lacking within-mode experimental conditioning to more sharply identify the degree of differences attributable directly to the mode effects of web surveys. For example, scale questions were presented in web mode using a horizontal scrolling matrix, but without an experimental inclusion of alternative ordinary scale format with all items visible in a form of a table. Such experiments would enable the separate analysis of other response sets, such as acquiescence, primacy, and recency effects. However, a detected general tendency of less extreme responding suggests that answers to scale questions in web surveys are not at least strongly affected by response sets.

Our analysis focused predominantly on the big picture about the presence or absence of selected response sets in web surveys. The analysis across a large number of variables enabled an insight into the prevalence of these response effects, but at the cost of lower attention to specifics of individual items or scales. Especially beneficial would be the addition of indicators of validity and reliability of measurement. Although this could technically also be performed generically across all items, it is questionable whether the

result would give a sufficiently accurate picture. Namely, the key component of the analysis of measurement characteristics is a careful consideration of content and latent structure of the items.

Simultaneous modelling of a large number of dependent variables also carries a risk of falsely detected effects. On the other hand, a too conservative approach is likely to obscure important findings. We attempted to take a central path between these extremes by arbitrarily adjusting the threshold for interpretation of results as significant to $\alpha = 0.01$, reporting more conservative significance adjustments and at the same time including effect sizes of significant and non-significant effects.

Finally, the lack of negative effects of web mode on selected indicators of data quality may have been contributed by a specific sample of respondents from an online access panel. While this allowed us to reach a demographically diverse population, these individuals are highly used to participation in (long) web surveys and are likely to have substantial experience with the use of computers and the Internet for everyday tasks.

The key elements of an improved experimental design, that would allow us to resolve the open issues identified in our empirical study, would therefore include additional within-mode experimental manipulations, more sensitive measures of the impression management susceptibility of questions, and more variable characteristics of analysed scale questions. Application of the survey to a non-panel population would also increase the confidence in the generalizability of the results.

Conclusion

Our dissertation addressed the problem of mode effects from a very broad perspective, much broader than is usually the case in methodological literature. Because of this, its key added value lies in the systematic evaluation of factors contributing to mode effects in web surveys. It exposed and empirically illustrated the volatile nature of the problem, where characteristics of the mode present merely a foundation for the potential emergence of damaging effects. In this closing chapter we summarize the key points of the dissertation and expose some implications of our findings.

Summary of mode effects in web surveys

Web surveys as a specific mode

We began our investigation of mode effects in web surveys by thoroughly discussing essential concepts. First we placed the mode of data collection into the broader context of a data collection system, which covers all activities of operational implementation of a survey (Biemer and Lyberg 2003). Survey estimates can be affected by both how data is collected and how other survey activities (sampling, solicitation, data management, etc.) are performed. While data collection is inevitably interdependent with these processes, its distinguishing is crucial to gain an understanding of how a specific mode influences data quality.

Although the term “mode” is routinely used in everyday survey-related conversations, its understanding is surprisingly unclear. We avoided the most likely unsuccessful attempts to propose a universal definition of mode, but formed an operational definition to aid the discussion of mode effects. In line with this, we proposed the definition of

mode as a set of data collection procedures that determine the basic principles of communication and information transmission between the respondent and the survey questionnaire. We identified six inherent mode characteristics on which these basic principles are based: the main question presentation (input) channel, response (output) channel, interviewer involvement, closeness of interaction between interviewer and respondent, use of computer technology for data collection, and medium of information transmission. A web survey is then defined as a mode with a visual main input channel, an electronic output channel, no interviewer's involvement and no interaction between respondent and interviewer, the respondent's use of computer technology, and the web as an information transmission medium.

Inherent mode characteristics are basic building blocks for the implementation of data collection. By opting for a specific mode we restricted and, to certain degree, determined the range of other mode-related characteristics, herein called implementation-specific and contextual characteristics. Their nature strongly depends on how a specific survey is implemented and in what context the data collection takes place. Question processing order, availability of contextual information, participation burden, technology used by the respondent, privacy perceptions, and conferred legitimacy of the survey request are some examples of implementation-specific and contextual characteristics we addressed. Implementation and contextual variations are particularly pronounced in web surveys, which offer diverse possibilities of implementation in terms of questionnaire design, multimedia, and interactions with respondents. At the same time respondents are free to choose time, place, pace, and device for survey participation.

A high flexibility of web surveys may question the rigid grounding of mode definition on six characteristics. Furthermore, the common use of the term "web survey" goes even beyond the six mode characteristics we used to define this mode. New devices, virtual interviewers, the introduction of auditory question presentation, speech recognition technologies for answer entry, and numerous other innovations increasingly blur the border between different modes (Couper 2005; Couper 2011). As discussed by Callegaro, Lozar Manfreda, and Vehovar (2014), it may therefore be more appropriate to

speak about a family of *web-related modes* of which "standard web surveys" are only one member.

The Couper's (2011) question whether the concept of mode is becoming outmoded as a survey descriptor is thus in place. The problem is not trivial nor unimportant from the perspective of terminological differences in survey methodology, which result in the sometimes unclear and inconsistent use of mode descriptions. However, for mode effects the crucial consideration is not *which* characteristics define the chosen mode, but *how* characteristics of this mode influence the accuracy of data. In other words, for this particular purpose it is more important to understand the properties of a specific set of data collection procedures than deciding whether these procedures should be labelled as a separate mode.

Understanding and observing mode effects

We defined mode effects as all direct and indirect effects of inherent mode characteristics on the accuracy of obtained survey estimates. In the total survey error (TSE) framework, mode is thus best understood as an error source. However, a straightforward definition and schematic simplifications of the TSE easily obscure the underlying complexity of the problem. Mode influences are not limited to a direct impact of inherent mode characteristics on the surveys estimate. Their influence is mediated through implementation-specific and contextual characteristics as well as other factors.

Mode effects can therefore be contributed by other stages of the survey process and mediated by other error sources within TSE. An example of this are approaches to solicitation that may increase or decrease the respondent's perception of privacy and survey legitimacy in combination with the properties of information transmission medium. At the same time, mode can influence not only measurement errors, but also other components of TSE. For example, one likely explanation for lower response rates in web surveys lies in the lack of the interviewer's persuasiveness for participation and a lower perceived legitimacy of web as the medium (Lozar Manfreda et al. 2008). This makes the

nature of mode effects very volatile and helps explain often inconsistent results of mode effect studies.

The observation of mode effects of any mode is commonly conducted by various experimental designs based on the comparison of selected statistical parameters between two or more modes. Researchers needs to bear in mind several considerations in order to avoid a misinterpretation of the results. It is first necessary to acknowledge the possibility of confounding factors. Especially split-sample experimental designs are prone to a confounding of mode effects with unit nonresponse and non-coverage errors.

Once the results are obtained, it is crucial to distinguish among between-mode differences and mode effects. While the former is clearly an indicator of the latter, assuming well-controlled experimental conditions, it provides no explicit answer about which mode is causing the difference. To gain insight into this problem, the researcher needs to rely on a theoretical background as well as on additional within-mode empirical investigation of response effects. A significant fallacy would be also to dismiss the existence of mode effects in a specific mode if the analysis shows no between-mode differences in estimates. Mode effects exist within a specific mode and are not the result of multi-mode data collection. Absence of differences between two compared modes may well be the consequence of similar mode effects present in both modes. The added value of within-mode experimental comparisons to detect potential sources of mode effects should therefore not be neglected.

Finally, the detection of effects can significantly depend on the estimated statistical parameters. Previous research found the presence of lower effect sizes in correlational analyses than in mean comparisons, known as the form resistant hypothesis (Krosnick and Alwin 1987). Our empirical results, consistent with a similar approach by Jäckle et al. (2006), also showed a substantially higher detection of effects by analysing the distributions of all response categories to scale questions than by focusing merely on mean differences. The decision regarding selected parameters primarily depends on the researcher's interest, but simultaneous consideration of different parameters is beneficial to obtain a more thorough picture of mode effects.

Factors and consequences of mode effects in web surveys

In our dissertation we focused on the influence of mode effects on the web survey response process. This primarily embraces measurement errors and item nonresponse. Response process models offer solid theoretical grounds for understanding these effects. They deal with the respondent's information processing, deviations of the response process from optimal paths, and interactions between respondents and the questionnaire. A majority of our discussion was based on the four-stage information processing model by Tourangeau et al. (2000), extended with conceptualizations by other authors. The most important extensions to understand mode effects in web surveys include the satisficing model by Krosnick and Alwin (1987), deviations due to sensitive questions, and the model of respondent–questionnaire interaction by Redline and Dillman (2001). We reviewed factors contributing to mode effects in web surveys by bearing in mind the foundation provided by these models.

We considered studies comparing web surveys with other modes on a broad range of factors related to characteristics of web mode. The comprehensive review allowed us to extend the conceptual model of mode effects proposed by Tourangeau and colleagues (Tourangeau and Smith 1996; Tourangeau et al. 2000) to include a more thorough range of factors. Although the model presented here applies to web surveys, it can be further generalized to other modes.

The model establishes potential mediating relations between six inherent mode characteristics, implementation-specific and contextual characteristics of web mode. Numerous studies indicate that mode effects in web surveys emerge under different circumstances, where complex relations and interactions between mode characteristics are accompanied by other specific survey-related and respondent-related factors. We grouped the consequences of mode effects into four key types of deviations leading to measurement errors: context effects, objective failures, shortcutting, and impression management.

In general, many theoretical and empirical studies regard self-administration as the most critical characteristic of web mode from the perspective of mode effects. A lack of interviewers require respondents to perform the additional tasks of navigating through the questionnaire and attending to questions, as well as severely limiting the possibilities of providing additional extrinsic motivation. Especially when accompanied by a long and demanding questionnaire, unmotivated respondents, or respondents with lower cognitive abilities, this can increase the likelihood of shortcutting. However, empirical results are generally mixed and inconclusive, which was also the case in our experimental study. While we did find some indications of higher non-differentiation and midpoint answering compared to telephone and face-to-face interviewing, effect sizes were generally small and sometimes not even in the expected direction.

Other factors of mode effects in web surveys are even more evasive and restricted to specific survey implementations. Here we mention only some of the discussed issues. Some questionnaire layouts or question types can interact with the visual input channel in violating the visibility principle or increasing the task difficulty, resulting in satisficing or objective failures to answer questions. In addition, a visual presentation conveys more explicit contextual information, such as images, content of previous questions, size of input boxes, and overall design of scale questions. The respondent's inclusion of such information into the response process gives rise to context effects.

Computerization can also be the source of damaging effects, especially among respondents lacking sufficient computer experience or even having negative attitudes toward the use of computers. Further negative influences can arise from the overuse of interactive features of computerized questionnaires. A highlighting example are early findings from virtual interviewer surveys with the unclear conclusion whether a reduction of comprehension problems caused by the lack of interviewers in web surveys outweigh the problem of potentially decreased sense of privacy, described by Tourangeau and Yan (2007) as the effect of media presence.

The final example from our discussion of factors of mode effects are specifics of the web as an information transmission medium. One set of issues arises from its inability to

convey legitimacy comparable to the one with interviewer-respondent interaction. While this is in part the problem of all self-administered modes, it is further escalated by the anonymous nature of the web. Another set of issues is more technical. As the number of web-enabled devices increases, respondents are able to complete surveys in more and more diverse environments. Although beneficial from the perspective of flexibility, this also further reduces the control over the surveying situation. The consequence can be specific measurement errors caused by distracting factors from the environment or technical issues due to inappropriate questionnaire functioning on the respondent's device.

While mode effects are by definition errors with a damaging impact on the accuracy of estimates, we found it important to additionally emphasize aspects of web mode that reduce the presence of effects. When questions are sensitive, self-administration turns from a potentially dangerous facilitator of mode effects to their main suppressor. Probably the most consistent finding of empirical mode comparisons is reduced response editing due to impression management tendencies of respondents. This was also firmly confirmed by our empirical study, where web respondents were significantly less likely to select socially desirable answers on a vast majority of analysed questions.

Other characteristics of web mode can also provide significant advantages. The visual presentation of questions reduces question comprehensibility problems and allows for the immediate provision of definitions where necessary. Respondents are able to tailor the pace of the response process and do not experience external pressure for the swift provision of answers. Computer-enabled questionnaire dynamics can automatize navigation through the questionnaire, lowering the burden of participation due to self-administration. The key to enjoying these and other benefits of web surveys lies in a sufficient understanding of the potential impact of utilized questionnaire features on the survey response process.

Implications and further research

Web survey practice

A widespread awareness of mode effects is important for the practical application of web surveys particularly due to the specific nature of their utilization. The web is a driving force of the "democratization" of survey research (Tourangeau et al. 2013), where virtually anyone with basic computer skills can carry out their own survey. Many modern software tools offer user-friendly access to basic and advanced features of web questionnaires, allowing for technically simple development, but in no respect assuring its methodologically sound implementation. As stressed above, the careless use of available features can start interacting with the inherent characteristics of web mode and introduce a powerful source of potential mode effects.

The flexibility of web questionnaires brings a danger of damaging data quality with overexcitement with a variety of media-related and interactive features. The issue may soon become even more prominent with the increasing fascination with highly interactive and gamified surveys. A somewhat more conservative approach to questionnaire design, advocated for example by Couper (2008) thus seems beneficial until methodological advantages of such innovations are sufficiently proven.

Less explicit, but probably even more important, is the appropriate consideration of the respondent's motivation. As self-administration is potentially one of the most damaging sources of mode effects in web surveys, special care should be devoted to maintaining the sufficient motivation of respondents and the prevention of satisficing. A carefully presented questionnaire (Deutskens et al. 2006), motivating instructions (Kunz and Fuchs 2013), and assurance about the survey's importance (Dillman et al. 2008), are only some of the mechanisms to help establish a higher commitment of respondents. Further efforts need to be made to establish sufficient trust of respondents in order to overcome

legitimacy-related problems of the web medium. Two possible ways of tackling this include the use of incentives and ordinary mail invitations during the initial contact with the respondent (de Leeuw 2005; Lozar Manfreda et al. 2008).

Finally, the survey practitioner's perspective should avoid taking into account accuracy of the estimates as the only indicator of data quality. As already Deming (1944) pointed out, even very accurate results make little added value if delivered too late. The relevance of concepts, timeliness, accessibility to information, comparability, coherence, and completeness (Biemer and Lyberg 2003) need to be all balanced in order to provide high-quality survey data. Furthermore, research is rarely free from cost-reduction pressures. Since web surveys excel in lower costs of data collection and rapid data availability, potential threats of mode effects should also be evaluated in light of these advantages.

It is important to note that the findings of our elaboration primarily apply to surveys of individuals. We did not address the problem of business surveys, which require additional consideration of response processes at the organisational level. Bavdaž (2007) discusses a wide range of organisational factors and persons involved in the response process of business surveys apart from respondents who report answers. This introduces a large set of additional potential sources of mode effects. While the proposed model of mode effects in web surveys covers only the respondent's level, additional extensions can be implemented to foster the understanding of the problem in business surveys.

Mixed-mode surveys

Awareness of the problem of mode effects significantly increased with the growing trends of mixed-mode data collection, commonly used to compensate the weakness of individual modes at affordable costs (de Leeuw 2005). When different modes are used for data collection, mode effects can jeopardize the comparability of results as one of the key aspects of survey quality. The issue can be very critical, especially if different

modes are used for different segments of the target population, leading to unequal biases across segments.

In mixed-mode surveys, the distinction of *mode effects* from *between-mode differences* is of especially high practical importance. The key question for comparability is whether it is worth sacrificing some accuracy offered by one mode to ensure results that are more similar to the other mode. This principle is followed by *unified mode design* where questions are presented as similarly as possible across all modes (Dillman et al. 2008). Unified design does not necessary lead to reduction of data quality and may in fact help reduce mode effects in all modes used during data collection. Two examples from surveys combining telephone and web data collection are switched to forced-choice instead of check-all-that-apply format (Smyth et al. 2006b) and the use of a horizontal scrolling matrix instead of a classic table format for scale questions (Klausch et al. 2012). Although these designs are not commonly used in web surveys, research indicates their better performance in reduction of satisficing. In other instances, mode-specific design is more questionable. For example, longer definitions of concepts can easily be presented by default in a web questionnaire, but are usually impractical to be read out each time in a telephone interview. The researcher would need to decide whether to sacrifice some accuracy of web mode (e.g. by making the definition available upon request) and in this way achieve better comparability with telephone interviewing.

The unimode principle primarily refers to the comparability of presentational or structural characteristics of the question. The more important goal, however, is to achieve the comparability of cognitive operations that take place during the question answering processes. The following statement by Cobanoglu et al. (2001, 443) underlines this aspect:

> *This potential problem* [of mode effects] *may be prevented, if not completely eliminated, by applying a unimodal design which focuses on writing and presenting questions in a way that assures the receipt by respondents of a* ***common mental stimulus****.* [Emphasis added.]

Achieving common mental stimulus sometimes requires a *mode-specific* design in which the questionnaire is modified by taking into account the different capabilities of different modes (Dillman et al. 2008). Examples from web surveys include the use of interactive features to resemble interviewer's interventions, dynamic questionnaires and visual guidance to relieve the burden of navigation between the questions, motivational instructions, and so on.

In general, however, assurance of common performance of the response process across modes is very difficult. We still do not have enough knowledge of response process details to completely understand the underlying mechanisms. Furthermore, some issues are so directly influenced by inherent mode characteristics that they cannot be completely eliminated by specific survey implementations. As shown by Krysan and Couper (2003), already the mere presence of an interviewer can cause higher levels of impression management. Similarly, visual presentation of questions in web surveys (and other visual modes) *per se* provide additional contextual information.

The problem of mixed-mode surveys shows how crucial is to understand the contribution of each individual source of mode effects. Couper (2011, 897) illustrates this by an ingenious comparison to everyday life:

> *Mixing modes is much like cooking – one can't learn to combine ingredients until one understands the properties of the individual ingredients.*

In the final part we expose some directions for future research that could help further achieve these goals.

Further research efforts

Considering the high diversity of factors contributing to the emergence of mode effects in web surveys, inconsistent and inconclusive findings of empirical research are not surprising. Survey methodology as a scientific field should therefore devote more efforts into refining the strategies for coping with the problem. Our empirical study confirmed

the benefits of using different modelling approaches to identify potential mode effects and observing the presence of differences on a large number of variables. Although this may lead to certain issues with statistical inference and may even introduce some of the data mining approaches into scientific research, the importance of exploratory investigations should not be neglected. Again, a better explanatory power of studies could be achieved by introducing more within-mode experimental manipulations in addition to between-modes comparisons.

An especially powerful and underexploited potential for better understanding of mode effects comes from meta-analyses. Interactions between inherent and other mode characteristics, as well as factors related to target population, topics, questionnaire length, and other parameters of a survey project, complicate generalizations about the nature of mode effects from individual experimental studies. Meta-analyses overcome this problem by enabling the research of relations between survey characteristics and observed effect sizes across many studies. A few meta-analysis related to mode effects have already been conducted (de Leeuw 1992; Richman et al. 1999; Tourangeau et al. 2007; Ye et al. 2011; Tourangeau et al. 2013). Their important findings should encourage further efforts to extend the models with higher number of survey-related variables. The attempts to do so may, however, encounter the obstacles of restricted reporting of necessary details by authors of individual studies. This is a strong argument in favour of open science principles, particularly open access to empirical data, and standardized reporting of survey details, like *Data Documentation Initiative* (Vardigan et al. 2008).

Finally, stronger connections need to be made between theoretical conceptualizations and empirical research of the survey response processes. Existing theoretical models provide invaluable descriptions of the processes taking place during the answering of questions, but their empirical verification is relatively limited. As pointed out by Schwarz (2007) this requires an intensified cooperation between survey methodology and cognitive psychology. Computerized questionnaires also allow for the automated collection of data about the process of data collection, so called paradata (Couper 2005). A recent increase in their exploitation reveals a wide range of utilization possibilities, going far

beyond simple time-related data. For example, paradata in web surveys can help identify changes of answers, navigational paths through the questionnaire, triggering of error messages, some forms of multitasking, and more (Berzelak et al. 2012). Another promising research approach for understanding cognitive variations between different modes is the integration of cognitive probing techniques into web surveys (Behr et al. 2012). These and other innovative methodologies open new paths toward more definite answers regarding mode effects in web surveys and other modes.

Concluding remarks

The dissertation pursued to scrutinize the central thesis that mode effects in web surveys are the result of a broad set of factors related not only to the mode itself but also to specific survey implementations. We attempted to establish a more solid conceptual framework of mode effects in web surveys, provide a comprehensive analysis of forms and sources of mode effects, empirically demonstrate the volatile nature of the problem, and draw implications for further research and survey practice. Focusing on the provision of such big picture of the problem, of course, comes at the cost of being unable to dwell deep into each individual aspect. However, it gives an integrative insight into potential sources of mode effects and provides opportunities for more the concerted development of further solutions to the problem.

The important moral of our work is that the study of mode effects should not neglect issues related to the implementation of a single mode. It is wrong to expect the absence of mode effects outside mixed-mode studies, and sources of effects should primarily be sought within an individual mode. In the process of evaluating mode effects, it is important to consider a broad range of potentially contributing factors. Our conceptual model of mode effects in web surveys can provide a useful guidance for this purpose.

Although we did not conduct an extensive empirical verification of many factors of mode effects, our empirical exploration combined with the findings of other studies reassures the position of web surveys as a viable survey mode. While far from being without the

dangers of mode effects, the problem generally does not seem to be any more prominent than in traditional modes. In many cases, especially with sensitive questions, web surveys are even less prone to mode effects than interviewer-administered modes. Innovative survey implementations also show a promising potential for overcoming the issues of self-administration, which is often found to be a principle cause of mode effects in web surveys. Investing into more intensive, innovative, and strategically oriented research efforts is therefore a key driving force for the future improvements of web survey methodology and the assurance of the high quality of obtained data.

References

Aassve, Arnstein, Icek Ajzen, Francesco Billari, Christoph Bühler, Pearl Dykstra, Tineke Fokkema, Thomas Hansen, et al. 2011. *Generations and Gender Survey Questionnaire, Draft 2.4*. The Hague, NLD: GGP Project Team.

Aguinis, Herman and Christine A Henle. 2001. Empirical Assessment of the Ethics of the Bogus Pipeline. *Journal of Applied Social Psychology* 31 (2). Blackwell Publishing Ltd: 352–75. doi:10.1111/j.1559-1816.2001.tb00201.x.

Allen, Mary J. and Wendy M. Yen. 2002. *Introduction to Measurement Theory*. Long Grove, IL, USA: Waveland.

Alwin, Duane F. 2007. *Margins of Error*. Hoboken, NJ, USA: John Wiley & Sons.

Anderson, Ronald, Judith Kasper, and Martin Frankel. 1979. *Total Survey Error: Applications to Improve Health Surveys*. San Francisco, CA, USA: Jossey-Bass.

Aquilino, William S. and Leonard A. Lo Sciuto. 1990. Effects of Interview Mode on Self-Reported Drug Use. *Public Opinion Quarterly* 54 (3): 362–93. doi:10.1086/269212.

Aquilino, William S., Debra L. Wright, and Andrew J. Supple. 2000. Response Effects due to Bystander Presence in CASI and Paper-and-Pencil Surveys of Drug Use and Alcohol Use. *Substance Use and Misuse* 35 (6–8): 845–67.

Ayidiya, Stephen A. and McKee J. McClendon. 1990. Response Effects in Mail Surveys. *Public Opinion Quarterly* 54 (2): 229–47. doi:10.1086/269200.

Backor, Kristen, Saar Golde, and Norman Nie. 2007. *Estimating Survey Fatigue in Time Use Study*. Paper presented at the 2007 International Association for Time Use Research Conference, Washington, D.C., USA. Available at: http://www.atususers.umd.edu/wip2/papers_i2007/Backor.pdf (2013-10-11).

Bälter, Olle. 2005. *Using Computer Games Design to Increase Response Rates*. Paper presented at the ESF SCSS Exploratory Workshop: Internet survey methodology: Toward concerted European research efforts, Dubrovnik, HRV.

Bates, Nancy. 2001. Internet versus Mail as a Data Collection Methodology from a High Coverage Population. In *Proceedings of the Annual Meeting of the American Statistical Association*. Alexandria, VA, USA: American Statistical Association. Available at: http://www.amstat.org/sections/srms/proceedings/y2001/Proceed/00311.pdf (2013-09-14).

Bates, Scott C. and Jared M. Cox. 2008. The Impact of Computer versus Paper-Pencil Survey, and Individual versus Group Administration, on Self-Reports of Sensitive Behaviors. *Computers in Human Behavior* 24 (3): 903–16. doi:10.1016/j.chb.2007.02.021.

Bavdaž, Mojca. 2007. *Measurement Errors and the Response Process in Business Surveys*. Doctoral dissertation. Ljubljana, SVN: Univerza v Ljubljani, Ekonomska fakulteta.

Bech, Mickael and Morten Bo Kristensen. 2009. Differential Response Rates in Postal and Web-Based Surveys among Older Respondents. *Survey Research Methods* 3 (1): 1–6. Available at: https://ojs.ub.uni-konstanz.de/srm/issue/view/65 (2013-11-03).

Beebe, Timothy J., Patricia A. Harrison, James A. Mcrae, Ronald E. Anderson, and Jayne A. Fulkerson. 1998. An Evaluation of Computer-Assisted Self-Interviews in a School Setting. *Public Opinion Quarterly* 62: 623–32.

Behr, Dorotheé, Lars Kaczmirek, Wolfgang Bandilla, and Michael Braun. 2012. Asking Probing Questions in Web Surveys: Which Factors Have an Impact on the Quality of Response? *Social Science Computer Review* 30 (4): 487–98. doi:10.1177/0894439311435305.

Bender, Ralf and Stefan Lange. 2001. Adjusting for Multiple Testing—when and How? *Journal of Clinical Epidemiology* 54: 343–49.

Benjamini, Yoav and Yosef Hochberg. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 51 (1): 289–300.

Benjamini, Yoav and Daniel Yekutieli. 2001. The Control of the False Discovery Rate in Multiple Testing under Dependency. *The Annals of Statistics* 29 (4): 1165–88.

Bennink, Margot, Guy Moors, and John Gelissen. 2013. Exploring Response Differences between Face-to-Face and Web Surveys: A Qualitative Comparative Analysis of the Dutch European Values Survey 2008. *Field Methods*, January, Advance online publication. doi:10.1177/1525822X12472875.

Berzelak, Nejc, Katja Lozar Manfreda, Ana Slavec, and Vasja Vehovar. 2012. *Advanced Paradata in Web Surveys: What Can They Tell about the Response Process?* Paper presented at the American Association for Public Opinion Research 67th Annual Conference, Orlando, FL, USA.

Biemer, Paul P. 2010. Total Survey Error: Design, Implementation, and Evaluation. *Public Opinion Quarterly* 74 (5): 817–48. doi:10.1093/poq/nfq058.

Biemer, Paul P., Heather Bergdahl, Lilli Japec, Åke Pettersson, and Dennis Trewin. 2012. *A Tool for Managing Product Quality: The Case of Statistics Sweden*. Paper presented at the International Total Survey Error Workshop, Sanpoort, NLD.

Biemer, Paul P. and Rachel Caspar. 1994. Continuous Quality Improvement for Survey Operations: Some General Principles and Applications. *Journal of Official Statistics* 10 (3): 307–26. Available at: http://www.jos.nu/Articles/abstract.asp?article=103307 (2013-10-14).

Biemer, Paul P. and Lars E. Lyberg. 2003. *Introduction to Survey Quality*. Hoboken, NJ, USA: John Wiley & Sons.

Bishop, George F. 1997. Response-Order Effects in Public Opinion Surveys: THe Plausibility of Rival Hyptheses. In *Proceedings of the Survey Research Methods Section*, 1041–46. Alexandria, VA, USA: American Statistical Association. Available at: http://www.amstat.org/sections/srms/Proceedings/papers/1997_179.pdf (2013-09-22).

Bishop, George F., Hans-Jürgen Hippler, and Fritz Strack. 1988. A Comparison of Response Effects in Self-Administered and Telephone Surveys. In *Telephone Survey Methodology*, edited by Robert M. Groves, Paul P. Biemer, Lars E. Lyberg, James T. Massey, William L. Nicholls, II, and Joseph Waksberg, 321–40. Hoboken, NJ, USA: John Wiley & Sons.

Bishop, George F. and Andrew Smith. 2001. Response-Order Effects and the Early Gallup Split-Ballots. *Public Opinion Quarterly* 65 (4): 479–505. doi:10.1086/323575. Available at: rg/content/65/4/479.abstract (2013-10-14).

Bless, Herbert, Gerd Bohner, Traudel Hild, and Norbert Schwarz. 1992. Asking Difficult Questions: Task Complexity Increases the Impact of Response Alternatives. *European Journal of Social Psychology* 22: 309–12. doi:10.1002/ejsp.2420220309.

Bohner, Gerd and Nina Dickel. 2011. Attitudes and Attitude Change. *Annual Review of Psychology* 62 (January): 391–417. doi:10.1146/annurev.psych.121208.131609.

Börkan, Bengü. 2010. The Mode Effect in Mixed-Mode Surveys: Mail and Web Surveys. *Social Science Computer Review* 28 (3): 371–80. doi:10.1177/0894439309350698.

Bosnjak, Michael, Tracy L. Tuten, and Wolfgang Bandilla. 2001. Participation in Web Surveys – A Typology. *ZUMA Nachricten* 48: 7–17.

Boyer, Kenneth K., John R. Olson, Roger J. Calantone, and Eric C. Jackson. 2002. Print versus Electronic Surveys: A Comparison of Two Data Collection Methodologies. *Journal of Operations Management* 20 (4): 357–73. doi:10.1016/S0272-6963(02)00004-9.

Bradburn, Norman M., Seymour Sudman, Ed Blair, and Carol Stocking. 1978. Question Threat and Response Bias. *Public Opinion Quarterly* 42 (2): 221–34. doi:10.1086/268444.

Braunsberger, Karin, Hans Wybenga, and Roger Gates. 2007. A Comparison of Reliability between Telephone and Web-Based Surveys. *Journal of Business Research* 60 (7): 758–64. doi:10.1016/j.jbusres.2007.02.015.

Buchanan, Tom and Ulf-Dietrich Reips. 2001. *Platform-Dependent Biases in Online Research: Do Mac Users Really Think Different?* Paper presented at the General Online Research Conference 2001, Gottingen, DEU.

Buchanan, Tom and John L. Smith. 1999. Using the Internet for Psychological Research: Personality Testing on the World Wide Web. *British Journal of Psychology* 90 (1). Wiley-Blackwell: 125–45.

Burr, Michele A., Kerry Y. Levin, and Angela Beecher. 2001. *Examining Web vs. Paper Mode Effects in a Federal Government Customer Satisfaction Survey*. Paper presented at the American Association for Public Opinion Research 56th Annual Conference, Montreal, CAN.

Burton, Scot and Edward Blair. 1991. Task Conditions, Response Formulation Processes, and Response Accuracy for Behavioral Frequency Questions in Surveys. *Public Opinion Quarterly* 55 (1): 50–79. doi:10.1086/269241.

Buskirk, Trent D. and Charles Andrus. 2012. Smart Surveys for Smart Phones: Exploring Various Approaches for Conducting Online Mobile Surveys via Smartphones. *Survey Practice* 5 (1). Available at: http://surveypractice.wordpress.com/2012/02/21/smart-surveys-for-smart-phones/ (2013-04-05).

Callegaro, Mario. 2010a. Do You Know Which Device Your Respondent Has Used to Take Your Online Survey? *Survey Practice* 3 (December). Available at: http://surveypractice.wordpress.com/2010/12/08/device-respondent-has-used/ (2013-11-05).

———. 2010b. How Can Mobile Research Findings Be Integrated with Other Methodologies? Panel discussion at the Mobile Research Conference 2010, London, GBR.

———. 2013. Conference Note: From Mixed-Mode to Multiple Devices: Web Surveys, Smartphone Surveys and Apps: Has the Respondent Gone ahead of Us in Answering Surveys? *International Journal of Market Research* 55 (2): 107–10. doi:10.2501/IJMR-2013-000.

Callegaro, Mario, Katja Lozar Manfreda, and Vasja Vehovar. 2014. *Web Survey Methodology*. Thousand Oaks, CA, USA: Sage.

Callegaro, Mario, Yongwei Yang, Dennison S. Bhola, Don A. Dillman, and Tzu-Yun Chin. 2009. Response Latency as an Indicator of Optimizing in Online Questionnaires. *Bulletin de Méthodologie Sociologique* 103 (1): 5–25. doi:10.1177/075910630910300103.

Callegaro, Mario, Yongwei Yang, and Ana Villar. 2011. *A Meta-Analysis of Experiments Manipulating Progress Indicators in Online Surveys*. Paper presented at The American Association for Public Opinion Research 66th Annual Conference, Phoenix, AZ, USA.

Cannell, Charles F., Kent H. Marquis, and André Laurent. 1977. A Summary of Research Studies of Interviewing Methodology. *Data Evaluation and Methods Research*. Rockville, MD, USA. Available at: http://stacks.cdc.gov/view/cdc/12892/ (2013-12-15).

Cannell, Charles F., Peter V. Miller, and Lois Oksenberg. 1981. Research on Interviewing Techniques. In *Sociological Methodology 1981*, edited by S. Leinhardt, 389–437. San Francisco, CA, USA: Jossey-Bass.

Cantril, Hadley. 1944. *Gauging Public Opinion*. Princeton, NJ, USA: Princeton University Press.

Card, Stuart K., William K. English, and Betty Burr. 1977. *Evaluation of Mouse, Rate-Controlled Isometric Joystick, Step Keys, and Text Keys for Text Selection on a CRT*. Palo Alto, CA, USA: Xerox Palo Alto Research Center.

Carini, Robert M., John C. Hayek, George D. Kuh, John M. Kennedy, and Judith A. Ouimet. 2003. College Student Responses to Web and Paper Surveys: Does Mode Matter? *Research in Higher Education* 44 (1): 1–19.

Casey, Tristan W. and Arthur Poropat. 2014. Beauty Is More than Screen Deep: Improving the Web Survey Respondent Experience through Socially-Present and Aesthetically-Pleasing User Interfaces. *Computers in Human Behavior* 30 (January): 153–63. doi:10.1016/j.chb.2013.08.001.

Castelli, Luigi and Silvia Tomelleri. 2008. Contextual Effects on Prejudiced Attitudes: When the Presence of Others Leads to More Egalitarian Responses. *Journal of Experimental Social Psychology* 44: 679–86. doi:10.1016/j.jesp.2007.04.006.

Chang, Linchiat and Jon A. Krosnick. 2009. National Surveys Via Rdd Telephone Interviewing Versus the Internet: Comparing Sample Representativeness and Response Quality. *Public Opinion Quarterly* 73 (4): 641–78. doi:10.1093/poq/nfp075.

———. 2010. Comparing Oral Interviewing with Self-Administered Computerized Questionnaires: An Experiment. *Public Opinion Quarterly* 74 (1): 154–67. doi:10.1093/poq/nfp090.

Christian, Leah Melani and Don A. Dillman. 2004. The Influence of Graphical and Symbolic Language Manipulations on Responses to Self-Administered Questions. *Public Opinion Quarterly* 68 (1): 57–80. doi:10.1093/poq/nfh004.

Christian, Leah Melani, Don A. Dillman, and Jolene D. Smyth. 2007a. The Effects of Mode and Format on Answers to Scalar Questions in Telephone and Web Surveys. In *Advances in Telephone Survey Methodology*, edited by James M. Lepkowski, Clyde Tucker, J. Michael Brick, Edith D. de Leeuw, Lilli Japec, Paul J. Lavrakas, Michael W. Link, and Roberta L. Sangster, 250–75. Hoboken, NJ, USA: John Wiley & Sons. doi:10.1002/9780470173404.ch12.

———. 2007b. Helping Respondents Get It Right the First Time: The Influence of Words, Symbols, and Graphics in Web Surveys. *Public Opinion Quarterly* 71 (1): 113–25. doi:10.1093/poq/nfl039.

Cobanoglu, Cihan, Bill Warde, and Patrick J Moreo. 2001. A Comparison of Mail, Fax and Web-Based Survey Methods. *International Journal of Market Research* 43 (4). Warc LTD: 441–52.

Cochran, William Gemmell. 1953. *Sampling Techniques*. New York, NY, USA: John Wiley & Sons.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. London, GBR: Routledge.

Comley, Pete. 2002. Online Research Will Take off in the UK in 2003: The Case for and against. *Research Magazine*.

Conrad, Frederick G. 2011. *Interactive Interventions in Web Surveys Can Increase Response Accuracy*. Paper presented at The American Association for Public Opinion Research 66th Annual Conference, Phoenix, AZ, USA.

Conrad, Frederick G., Mick P. Couper, Roger Tourangeau, and Mirta Galesic. 2005. Interactive Feedback Can Improve the Quality of Responses in Web Surveys. In *Proceedings of the Survey Research Methods Section*, 3835–40. Alexandria, VA, USA: American Statistical Association. Available at: http://www.amstat.org/sections/srms/proceedings/y2005/Files/JSM2005-000938.pdf (2013-09-10).

Conrad, Frederick G., Mick P. Couper, Roger Tourangeau, and Andy Peytchev. 2006. Use and Non-Use of Clarification Features in Web Surveys. *Journal of Official Statistics* 22 (2): 245–69. Available at: http://www.jos.nu/Articles/abstract.asp?article=222245 (2013-11-10).

Conrad, Frederick G. and Michael F. Schober. 1998. A Conversational Approach to Computer-Administered Questionnaires. In *Proceedings of the Survey Research Methods Section*, 962–67.

Alexandria, VA, USA: American Statistical Association. Available at: http://www.amstat.org/sections/srms/proceedings/papers/1998_166.pdf (2013-08-02).

———. 1999. *A Conversational Approach to Text-Based Computer-Administered Questionnaires*. Paper presented at the ASC's 3rd International Conference on Survey and Statistical Computing, Edinburgh, GBR. Available at: http://www.websm.org/db/12/1880/rec/ (2013-08-02).

———. 2000. Clarifying Question Meaning in a Household Telephone Survey. *Public Opinion Quarterly* 64 (1): 1–28. doi:10.1086/316757.

Conrad, Frederick G., Michael F. Schober, and Tania Coiner. 2007. Bringing Features of Human Dialogue to Web Surveys. *Applied Cognitive Psychology* 21: 165–87. doi:10.1002/acp.1335.

Conrad, Frederick G., Michael F. Schober, Matthew Jans, Rachel A. Orlowski, and Danial Nielsen. 2008. *Virtual Interviewers on Mundane Non-Sensitive Topics: Dialog Cabability Affects Response Accuracy More than Visual Realism Does*. Paper presented at The American Association for Public Opinion Research 63rd Annual Conference, New Orleans, LA, USA.

Couper, Mick P. 2005. Technology Trends in Survey Data Collection. *Social Science Computer Review* 23 (4): 486–501. doi:10.1177/0894439305278972.

———. 2008. *Designing Effective Web Surveys*. Cambridge, GBR: Cambridge University Press.

———. 2011. The Future of Modes of Data Collection. *Public Opinion Quarterly* 75 (5): 889–908. doi:10.1093/poq/nfr046.

———. 2013. *Surveys on Mobile Devices: Opportunities and Challenges*. Paper presented at the Web Surveys for the General Population conference, London, GBR. Available at: http://www.cls.ioe.ac.uk/ (2013-10-06).

Couper, Mick P., Frederick G. Conrad, and Roger Tourangeau. 2007. Visual Context Effects in Web Surveys. *Public Opinion Quarterly* 71 (4): 623–34. doi:10.1093/poq/nfm044.

Couper, Mick P., Courtney Kennedy, Frederick G. Conrad, and Roger Tourangeau. 2011. Designing Input Fields for Non-Narrative Open-Ended Responses in Web Surveys. *Journal of Official Statistics* 27 (1): 65–85. Available at: http://www.jos.nu/Articles/abstract.asp?article=271065 (2013-08-15).

Couper, Mick P., Roger Tourangeau, Frederick G. Conrad, and Scott D. Crawford. 2004a. What They See Is What We Get: Response Options for Web Surveys. *Social Science Computer Review* 22 (1): 111–27. doi:10.1177/0894439303256555.

Couper, Mick P., Roger Tourangeau, and Kristin Kenyon. 2004b. Picture This! Exploring Visual Effects in Web Surveys. *Public Opinion Quarterly* 68 (2): 255–66. doi:10.1093/poq/nfh013.

Couper, Mick P., Michael W. Traugott, and Mark J. Lamias. 2001. Web Survey Design and Administration. *Public Opinion Quarterly* 65 (2): 230–53. doi:10.1086/322199.

Crawford, Scott D., Mick P. Couper, and Mark J. Lamias. 2001. Web Surveys: Perceptions of Burden. *Social Science Computer Review* 19 (2): 146–62. doi:10.1177/089443930101900202.

Cunningham, William A., Philip David Zelazo, Dominic J. Packer, and Jay J. Van Bavel. 2007. The Iterative Reprocessing Model: A Multilevel Framework for Attitudes and Evaluation. *Social Cognition* 25 (5): 736–60. Available at: http://search.ebscohost.com.nukweb.nuk.uni-lj.si/login.aspx?direct=true&db=a9h&AN=27392949&lang=sl&site=eds-live (2013-08-22).

De Leeuw, Edith D. 1992. *Data Quality in Mail, Telephone and Face to Face Surveys*. Amsterdam, NLD: TT-Publikaties.

———. 2005. To Mix or Not to Mix Data Collection Modes in Surveys. *Journal of Official Statistics* 21 (2): 233–55.

———. 2008. *The Effects of Computer-Assisted Interviewing on Data Quality*. Unpublished paper. Available at: http://igitur-archive.library.uu.nl/fss/2010-0601-200229/EdL-effect 2002.pdf (2013-09-16).

De Leeuw, Edith D., Joop Hox, and Mark Huisman. 2003. Prevention and Treatment of Item Nonresponse. *Journal of Official Statistics* 19 (2): 153–76. Available at: http://www.jos.nu/Articles/abstract.asp?article=192153 (2013-10-14).

De Leeuw, Edith D., Joop Hox, and Annette Scherpenzeel. 2010a. *Emulating Interviewers in an Online Survey: Experimental Manipulation of "Do-Not-Know" over the Phone and on the Web.* Paper presented at The American Association for Public Opinion Research 65th Annual Conference, Chicago, IL, USA. Available at: https://www.amstat.org/sections/SRMS/Proceedings/y2010/Files/400148.pdf (2013-07-19).

———. 2010b. Mode Effect or Question Wording? Measurement Error in Mixed Mode Surveys. In *Proceedings of the Survey Research Methods Section*, 5959–67. Alexandria, VA, USA: American Statistical Association. Available at: http://www.amstat.org/sections/srms/proceedings/y2010/Files/400117.pdf (2013-08-03).

Deming, William Edwards. 1944. On Errors in Surveys. *American Sociological Review* 9: 359–69.

———. 1960. *Sample Design in Business Research*. Hoboken, NJ, USA: John Wiley & Sons.

Denniston, Maxine M., Nancy D. Brener, Laura Kann, Danice K. Eaton, Timothy McManus, Tonja M. Kyle, Alice M. Roberts, Katherine H. Flint, and James G. Ross. 2010. Comparison of Paper-and-Pencil versus Web Administration of the Youth Risk Behavior Survey (YRBS): Participation, Data Quality, and Perceived Privacy and Anonymity. *Computers in Human Behavior* 26 (5). Elsevier Ltd: 1054–60. doi:10.1016/j.chb.2010.03.006.

Denscombe, Martyn. 2006. Web-Based Questionnaires and the Mode Effect: An Evaluation Based on Completion Rates and Data Contents of near-Identical Questionnaires Delivered in Different Modes. *Social Science Computer Review* 24 (2): 246–54. doi:10.1177/0894439305284522.

———. 2008. The Length of Responses to Open-Ended Questions: A Comparison of Online and Paper Questionnaires in Terms of a Mode Effect. *Social Science Computer Review* 26 (3): 359–68. doi:10.1177/0894439307309671.

———. 2009. Item Non-Response Rates: A Comparison of Online and Paper Questionnaires. *International Journal of Social Research Methodology* 12 (4): 281–91. doi:10.1080/13645570802054706.

Derouvray, Cristel and Mick P. Couper. 2002. Designing a Strategy for Reducing "no Opinion" Responses in Web-Based Surveys. *Social Science Computer Review* 20 (1): 3–9. doi:10.1177/089443930202000101.

Deutschmann, Marc and Frank Faulbaum. 2001. The Recruitment of Online Samples by CATI-Screening: Problems of Non-Response. In *The Challenge of the Internet: Proceedings of the ASC International Conference on Survey Research Methods*, edited by Andrew Westlake, Wendy Sykes, Tony Manners, and Malcolm Rig, 69–77. Chesham, GBR: Association for Survey Computing. Available at: http://www.asc.org.uk/publications/proceedings/ASC2001Proceedings.pdf#page=81 (2013-08-03).

Deutskens, Elisabeth, Ko de Ruyter, and Martin Wetzels. 2006. An Assessment of Equivalence Between Online and Mail Surveys in Service Research. *Journal of Service Research* 8 (4): 346–55. doi:10.1177/1094670506286323.

Deutskens, Elisabeth, Ko de Ruyter, Martin Wetzels, and Paul Oosterveld. 2004. Response Rate and Response Quality of Internet-Based Surveys: An Experimental Study. *Marketing Letters* 15 (1): 21–36.

Dijkstra, Wil and Yfke Ongena. 2006. Question-Answer Sequences in Survey-Interviews. *Quality and Quantity* 40: 983–1011.

Dillman, Don A. 1991. The Design and Administration of Mail Surveys. *Annual Review of Sociology* 17: 225–49.

Dillman, Don A. and Dennis Bowker. 2001. The Web Questionnaire Challenge to Survey Methodologists. In *Dimensions of Internet Science*, edited by Ulf-Dietrich Reips and Michael Bosnjak, 159–78. Lengerich, DEU: Pabs.

Dillman, Don A. and Leah Melani Christian. 2005. Survey Mode as a Source of Instability in Responses across Surveys. *Field Methods* 17 (1): 30–52. doi:10.1177/1525822X04269550.

Dillman, Don A., Glenn Phelps, Robert Tortora, Karen Swift, Julie Kohrell, Jodi Berck, and Benjamin L. Messer. 2009. Response Rate and Measurement Differences in Mixed-Mode Surveys Using Mail, Telephone, Interactive Voice Response (IVR) and the Internet. *Social Science Research* 38 (1): 1–18. doi:10.1016/j.ssresearch.2008.03.007.

Dillman, Don A., Roberta L. Sangster, John Tarnai, and Todd H. Rockwood. 1996. Understanding Differences in People's Answers to Telephone and Mail Surveys. In *Advances in Survey Research*, edited by Marc T. Braverman and Jana Kay Slater, 46–52. San Francisco, CA, USA: Evaluation Association.

Dillman, Don A., Jolene D. Smyth, and Leah Melani Christian. 2008. *Internet, Mail, and Mixed-Mode Durveys: The Tailored Design Method*. Hoboken, NJ, USA: John Wiley & Sons.

Dillman, Don A. and John Tarnai. 1991. Mode Effects of Cognitively Designed Recall Questions: A Comparison of Answers to Telephone and Mail Surveys. In *Measurement Errors in Surveys*, edited by Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman, 73–93. Hoboken, NJ, USA: John Wiley & Sons. doi:10.1002/9781118150382.ch5.

Downes-Le Guin, Theo, Reg Baker, Joanne Mechling, and Erica Ruylea. 2012. Myths and Realities of Respondent Engagement in Online Surveys. *International Journal of Market Research* 54 (5). Available at: http://www.marketstrategies.com/user_area/content_media/Online_Survey_Engagement.pdf (2013-10-14).

Duffy, Bobby. 2003. Response Order Effects – How Do People Read? *International Journal of Market Research* 45 (4): 457–66.

Eaton, Danice K., Nancy D. Brener, Laura Kann, Maxine M. Denniston, Tim McManus, Tonja M. Kyle, Alice M. Roberts, Katherine H. Flint, and James G. Ross. 2010. Comparison of Paper-and-Pencil versus Web Administration of the Youth Risk Behavior Survey (YRBS): Risk Behavior Prevalence Estimates. *Evaluation Review* 34 (2): 137–53. doi:10.1177/0193841X10362491.

Ellis, Paul D. 2010. *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation*. Cambridge, GBR: Cambridge University Press.

Epstein, J., W. D. Klinkenberg, D. Wiley, and L. McKinley. 2001. Insuring Sample Equivalence across Internet and Paper-and-Pencil Assessments. *Computers in Human Behavior* 17 (3): 339–46. doi:10.1016/S0747-5632(01)00002-4.

Esposito, James L. and Jared B. Jobe. 1991. A General Model of the Survey Interaction Process. In *Bureau of the Census Seventh Annual Research Conference Proceedings*, 537–67. Suitland, MD, USA: Bureau of the Census. Available at: http://www.bls.gov/osmr/pdf/st910040.pdf (2013-07-21).

Evan, William M. and James R. Miller. 1965. *Differential Effects on Response Bias of Computer vs. Conventional Administration of a School Science Questionnaire: A Field Experiment*. Working paper 206-66, Preliminary draft. Cambridge, MA, USA. Available at: http://dspace.mit.edu/bitstream/handle/1721.1/47186/differentialeffe00evan.pdf (2013-07-21).

Eysenck, Michael W. and Mark T. Keane. 2010. *Cognitive Psychology: A Student's Handbook*. 6th ed. Hove, GBR: Taylor & Francis Group.

Farrell, Dan and James C. Petersen. 2010. The Growth of Internet Research Methods and the Reluctant Sociologist. *Sociological Inquiry* 80 (1): 114–25.

Forsyth, Barbara and Michael L. Hubbard. 1992. A Method for Identifying Cognitive Properties of Survey Items. In *Proceedings of the Survey Research Methods Section*, 470–75. Alexandria, VA, USA: American Statistical Association. Available at: http://www.amstat.org/sections/srms/Proceedings/papers/1992_076.pdf (2013-08-23).

Fricker, Scott, Mirta Galesic, Roger Tourangeau, and Ting Yan. 2005. An Experimental Comparison of Web and Telephone Surveys. *Public Opinion Quarterly* 69 (3): 370–92.

Fu, Vincent Kang. 1998. Estimating Generalized Ordered Models. *Stata Technical Bulletin*, no. 44: 27–30.

Fuchs, Marek. 2007. Mobile Web Surveys: A Preliminary Discussion of Methodological Implications. In *Envisioning the Survey Interview of the Future*, edited by Frederick J. Conrad and Michael F. Schobber, 77–94. Hoboken, NJ, USA: John Wiley & Sons.

———. 2009. Asking for Numbers and Quantities: Visual Design Effects in Paper&pencil Surveys. *International Journal of Public Opinion Research* 21 (1). Oxford University Press / USA: 65–84.

Funke, Frederik and Ulf-Dietrich Reips. 2012. Why Semantic Differentials in Web-Based Research Should Be Made from Visual Analogue Scales and Not from 5-Point Scales. *Field Methods*, no. Advance online publication (May). doi:10.1177/1525822X12444061.

Galesic, Mirta and Michael Bosnjak. 2009. Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. *Public Opinion Quarterly* 73 (2): 349–60. doi:10.1093/poq/nfp031.

Galesic, Mirta and Roger Tourangeau. 2007. What Is Sexual Harassment ? It Depends on Who Asks! Framing Effects on Survey Responses. *Applied Cognitive Psychology* 202: 189–202. doi:10.1002/acp.

Galesic, Mirta, Roger Tourangeau, Mick P. Couper, and Frederick G. Conrad. 2008. Eye-Tracking Data: New Insights on Response Order Effects and Other Cognitive Shortcuts in Survey Responding. *Public Opinion Quarterly* 72 (5): 892–913. doi:10.1093/poq/nfn059.

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

Gerber, Eleanor R., Tracy R. Wellens, and Catherine Keeley. 1996. “Who Lives Here?”: The Use of Vignettes in Household Roster Research. In *Proceedings of the Survey Research Methods Section*, 962–67. Alexandria, VA, USA: American Statistical Association. Available at: http://www.census.gov/srd/papers/pdf/erg9601.pdf (2013-10-14).

Ghoneim, Ahmed. 2013. Gamified Surveys. Available at: http://gamifiedsurveys.com/ (2013-11-12).

Göritz, Anja. 2006. Incentives in Web Studies: Methodological Issues and a Review. *International Journal of Internet Science* 1 (1): 58–70. Available at: http://www.ijis.net/ijis1_1/ijis1_1_goeritz.pdf (2013-08-13).

Graesser, Arthur C., Zhiqiang Cai, Max M. Louwerse, and Frances Daniel. 2006. Question Understanding Aid (QUAID): A Web Facility That Tests Question Comprehensibility. *Public Opinion Quarterly* 70 (1): 3–22. doi:10.1093/poq/nfj012.

Grandcolas, Ursula, Ruth Rettie, and Kira Marusenko. 2003. Web Survey Bias: Sample or Mode Effect? *Journal of Marketing Management* 19 (5/6). Routledge: 541–61.

Gravlee, Clarence C., H. Russell Bernard, Chad R. Maxwell, and Aryeh Jacobsohn. 2013. Mode Effects in Free-List Elicitation: Comparing Oral, Written, and Web-Based Data Collection. *Social Science Computer Review* 31 (1): 119–32. doi:10.1177/0894439312455312.

Griffis, Stanley E., Thomas J. Goldsby, and Martha Cooper. 2003. Web-Based and Mail Surveys: A Comparison of Response, Data, and Cost. *Journal of Business Logistics* 24 (2). Blackwell Publishing Ltd: 237–58. doi:10.1002/j.2158-1592.2003.tb00053.x.

Grondin, Chantal and Limei Sun. 2008. 2006 Census Mode Effect Study. In *Proceedings of the Survey Research Methods Section*, 1255–62. Alexandria, VA, USA: American Statistical Association. Available at: http://www.amstat.org/sections/srms/proceedings/y2008/Files/300977.pdf (2013-12-06).

Groves, Robert M. 1979. Actors and Questions in Telephone and Personal Interview Surveys. *Public Opinion Quarterly* 43 (2): 190–205. doi:10.1086/268511.

———. 2004. *Survey Errors and Survey Costs*. Hoboken, NJ, USA: John Wiley & Sons.

———. 2005. *Total Survey Error: Past, Present, and Future*. Paper presented at the NISS Affiliates Workshop on Total Survey Error, Washington, D.C., USA.

Groves, Robert M. and Mick P. Couper. 1998. *Nonresponse in Household Interview Surveys*. Hoboken, NJ, USA: John Wiley & Sons.

Groves, Robert M., Floyd J. Fowler, Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2009. *Survey Methodology*. 2nd ed. Hoboken, NJ, USA: John Wiley & Sons.

Groves, Robert M. and Lars E. Lyberg. 2010. Total Survey Error: Past, Present, and Future. *Public Opinion Quarterly* 74 (5): 849–79. doi:10.1093/poq/nfq065.

Guidry, Kevin R. 2012. *Response Quality and Demographic Characteristics of Respondents Using a Mobile Device on a Web-Based Survey*. Paper presented at The American Association for Public Opinion Research 67th Annual Conference, Orlando, FL, USA.

Gutierrez, Christina, Tom Wells, Kumar Rao, and David Kurzynski. 2011. *Catch Them When You Can: Speeders and Their Role in Data Quality*. Paper presented at the Midwest Association of Public Opinion Research, Chicago, IL, USA. Available at: http://www.mapor.org/confdocs/absandpaps/2011/2011_slides/6d2Gutierrez.pdf (2013-11-10).

Guzy, Nathalie and Heinz Leitgöb. 2012. *Mode Effects in Online and Telephone Based Victimisation Surveys*. Paper presented at the Stockholm Criminology Symposium 2012, Stockholm, SWE.

Hansen, Morris H., William N. Hurwitz, Eli S. Marks, and W. Parker Mauldin. 1951. Response Errors in Surveys. *Journal of the American Statistical Association* 46 (254): 147–90. Available at: http://www.jstor.org/stable/2279918 (2013-07-08).

Heerwegh, Dirk. 2006. An Investigation of the Effect of Lotteries on Web Survey Response Rates. *Field Methods* 18 (2): 205–20. doi:10.1177/1525822X05285781.

Heerwegh, Dirk and Geert Loosveldt. 2008. Face-to-Face versus Web Surveying in a High-Internet-Coverage Population: Differences in Response Quality. *Public Opinion Quarterly* 72 (5): 836–46. doi:10.1093/poq/nfn045.

Heerwegh, Dirk, Tim Vanhove, Geert Loosveldt, and Koen Matthijs. 2004. *Effects of Personalization on Web Survey Response Rates and Data Quality*. Paper presented at the Sixth International Conference on Logic and Methodology (RC-33), Amsterdam, NLD. Available at: https://perswww.kuleuven.be/~u0034437/public/Files/ProceedingsRC33.pdf (2013-06-25).

Helgeson, James G. and Michael L. Ursic. 1994. The Role of Affective and Cognitive Decision-Making Process during Questionnaire Completion. *Psychology & Marketing* 11 (5). John Wiley & Sons, Inc. 493–510. Available at: http://search.ebscohost.com/login.aspx?direct=true&db=ufh&AN=11580238&lang=sl&site=eds-live (2013-08-09).

Herr, Paul M., Steven J. Sherman, and Russell H. Fazio. 1983. On the Consequences of Priming: Assimilation and Contrast Effects. *Journal of Experimental Social Psychology* 19 (4): 323–40.

Hertel, Guido, Sonja Naumann, Udo Konradt, and Bernard Batinic. 2002. Personality Assessment via Internet: Comparing Online and Paper-and-Pencil Questionnaires. In *Online Social Sciences*, edited by Bernard Batinic, Ulf-Dietrich Reips, and Michael Bosnjak, 137–53. Seattle, WA, USA: Hogrefe & Huber.

Hippler, Hans-Jürgen and Norbert Schwarz. 1998. Response Effects in Surveys. In *Telephone Survey Methodology*, edited by Robert M. Groves, Paul P. Biemer, Lars E. Lyberg, James T. Massey, II Nicholls, William L., and Joseph Waksberg, 102–22. Hoboken, NJ, USA: John Wiley & Sons.

Hochstim, Joseph R. 1967. A Critical Comparison of Three Strategies of Collecting Data from Households. *Journal of the American Statistical Association* 62 (319). American Statistical Association: 976–89. doi:10.2307/2283686.

Holland, Jennifer L. and Leah Melani Christian. 2009. The Influence of Topic Interest and Interactive Probing on Responses to Open-Ended Questions in Web Surveys. *Social Science Computer Review* 27 (2): 196–212.

Holtgraves, Thomas. 2004. Social Desirability and Self-Reports: Testing Models of Socially Desirable Responding. *Personality and Social Psychology Bulletin* 30 (2): 161–72. doi:10.1177/0146167203259930.

Jäckle, Annette, Caroline Roberts, and Peter Lynn. 2006. *Telephone versus Face-to-Face Interviewing: Mode Effects on Data Quality and Likely Causes*. ISER Working Paper 2006-41. Essex, GBR: Institute for Social & Economic Research. Available at: https://www.iser.essex.ac.uk/files/iser_working_papers/2006-41.pdf (2013-10-12).

Jenkins, Cleo R. and Don A. Dillman. 1997. Towards a Theory of Self-Administered Questionnaire Design. In *Survey Measurement and Process Quality*, edited by Lars E. Lyberg, Paul P. Biemer, Martin Collins, Edith D. de Leeuw, Cathryn Dippo, Norbert Schwarz, and Dennis Trewin, 165–96. Hoboken, NJ, USA: John Wiley & Sons. doi:10.1002/9781118490013.ch7.

Jobe, Jared B. 2003. Cognitive Psychology and Self-Reports: Models and Methods. *Quality of Life Research* 12: 219–27.

Jobe, Jared B. and Douglas J. Herrmann. 1996. Implications of Models of Survey Cognition for Memory Theory. In *Basic and Applied Memory Research: Practical Applications*, edited by Douglas J. Herrmann, Cathy McEvoy, Christopher Hertzog, Paula Hertel, and Marica K. Johnson, 193–205. Mahwah, NJ, USA: Lawrence Erlbaum Associates.

Jones, Edward E. and Harold Sigall. 1971. The Bogus Pipeline: A New Paradigm for Measuring Affect and Attitude. *Psychological Bulletin* 76 (5): 349–64.

Judd, Charles M., Roger A. Drake, James W. Downing, and Jon A. Krosnick. 1991. Some Dynamic Properties of Attitude Structures: Context-Induced Response Facilitation and Polarization. *Journal of Personality and Social Psychology* 60 (2): 193–202. doi:10.1037/0022-3514.60.2.193.

Kaczmirek, Lars. 2009. *Human-Survey Interaction: Usability and Nonresponse in Online Surveys*. Köln, DEU: Herbert von Halem.

Kelaher, Dan, Todd Nay, Brad Lawrence, Sabrina Lamar, and Carolyn M Sommerich. 2001. An Investigation of the Effects of Touchpad Location within a Notebook Computer. *Applied Ergonomics* 32 (1): 101–10. doi:http://dx.doi.org/10.1016/S0003-6870(00)00020-X.

Kennedy, Courtney and Stephen E. Everett. 2011. Use of Cognitive Shortcuts in Landline and Cell Phone Surveys. *Public Opinion Quarterly* 75 (2): 336–48. doi:10.1093/poq/nfr007.

Kerwin, Jeffrey, Kerry Levin, Andrew Wang, Stephen Cambell, and Stephanie Shipp. 2006. *A Comparison of Strategies for Reducing Item Nonresponse in a Web Survey*. Paper presented at the Joint Statistical Meetings 2006, Seattle, WA, USA.

Kiesler, Sara and Lee S. Sproull. 1986. Incentives in Web Studies: Methodological Issues and a Review. *Public Opinion Quarterly* 50 (3). American Association for Public Opinion Research: 402–13. Available at: http://search.ebscohost.com/login.aspx?direct=true&db=ufh&AN=5415005&lang=sl&site=eds-live (2013-09-10).

Kish, Leslie. 1965. *Survey Sampling*. New York, NY, USA: John Wiley & Sons.

Klassen, Robert D. and Jennifer Jacobs. 2001. Experimental Comparison of Web, Electronic and Mail Survey Technologies in Operations Management. *Journal of Operations Management* 19 (6): 713–28. doi:10.1016/S0272-6963(01)00071-7.

Klausch, Thomas, Edith D. de Leeuw, Joop Hox, Anouk Roberts, and Anneke de Jongh. 2012. *Matrix vs. Single Question Formats in Web Surveys: Results from a Large Scale Experiment*. Paper presented at the International Total Survey Error Workshop, Mannheim, DEU.

Knapp, Herschel and Stuart A. Kirk. 2003. Using Pencil and Paper, Internet and Touch-Tone Phones for Self-Administered Surveys: Does Methodology Matter? *Computers in Human Behavior* 19 (1): 117–34. doi:10.1016/S0747-5632(02)00008-0.

Knäuper, Bärbel. 1999. The Impact of Age and Education on Response Order Effects in Attitude Measurement. *Public Opinion Quarterly* 63 (3): 347–70. Available at: http://poq.oxfordjournals.org/content/63/3/347.short (2013-11-06).

Knäuper, Bärbel, Robert F. Belli, Daniel H. Hill, and A. Regula Herzog. 1997. Question Difficulty and Respondents Cognitive Ability: The Effect on Data Quality. *Journal of Official Statistics* 13 (2): 181–99. Available at: http://www.isr.umich.edu/src/smp/Electronic Copies/68.pdf (2013-09-16).

Knowles, Eric S. and Christopher A. Condon. 1999. Why People Say "Yes": A Dual-Process Theory of Acquiescence. *Journal of Personality and Social Psychology* 72 (2): 379–86.

Kreuter, Frauke, Stanley Presser, and Roger Tourangeau. 2008. Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity. *Public Opinion Quarterly* 72 (5): 847–65. doi:10.1093/poq/nfn063.

Krosnick, Jon A. 1991. Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology* 5 (3). John Wiley & Sons, Inc. 213–36.

———. 1999. Survey Research. *Annual Review of Psychology* 50: 537–67. Available at: http://www.stanford.edu/dept/communication/faculty/krosnick/docs/1999/1999 AnnRevSurvey.pdf (2013-08-24).

Krosnick, Jon A. and Duane F. Alwin. 1987. An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement. *Public Opinion Quarterly* 51 (2): 201–19.

Krosnick, Jon A. and Leandre R. Fabrigar. 1997. Designing Rating Scales for Effective Measurement in Surveys. In *Survey Measurement and Process Quality*, edited by Lars E. Lyberg, Paul P. Biemer, Martin Collins, Edith D. de Leeuw, Cathryn Dippo, Norbert Schwarz, and Dennis Trewin, 141–64. Hoboken, NJ, USA: John Wiley & Sons. doi:10.1002/9781118490013.ch6.

Krosnick, Jon A., Allyson L. Holbrook, Matthew K. Berent, Richard T. Carson, W. Michael Hanemann, Raymond J. Kopp, Robert Cameron Mitchell, et al. 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. doi:10.1086/341394.

Krysan, Maria and Mick P. Couper. 2003. Race in the Live and the Virtual Interview: Racial Deference, Social Desirability and Activation Effects in Attitude Surveys. *Social Psychology Quarterly* 66 (4): 364–83.

———. 2006. Race of Interviewer Effects: What Happens on the Web? *International Journal of Internet Science* 1 (1): 17–28.

Kunz, Tanja and Marek Fuchs. 2013. *Reducing Response Order Effects in Check-All-That-Apply Questions by Use of Dynamic Tooltip Instructions*. Paper presented at the General Online Research 2013 conference, Mannheim, DEU.

Kuusela, Vesa and Veijo Notkola. 2005. Survey Quality and Mobile Phones. Technical report. Helsinki, FIN.

Kwak, Nojin and Barry Radler. 2002. A Comparison between Mail and Web Surveys : Response Pattern, Respondent Profile, and Data Quality. *Journal of Official Statistics* 18 (2): 257–73.

Kypri, Kypros and Stephen J. Gallagher. 2003. Incentives to Increase Participation in an Internet Survey of Alcohol Use: A Controlled Experiment. *Alcohol and Alcoholism* 38 (5): 437–41. doi:10.1093/alcalc/agg107.

Layne, Benjamin H., Joseph R. Decristoforo, and Dixie Mcginty. 1999. Electronic versus Traditional Student Ratings of Instruction. *Research in Higher Education* 40 (2): 221–32.

Lenzner, Timo, Lars Kaczmirek, and Mirta Galesic. 2011. Seeing through the Eyes of the Respondent: An Eye-Tracking Study on Survey Question Comprehension. *International Journal of Public Opinion Research* 23 (3): 361–73. doi:10.1093/ijpor/edq053.

Lenzner, Timo, Lars Kaczmirek, and Alwine Lenzner. 2010. Cognitive Burden of Survey Questions and Response Times: A Psycholinguistic Experiment. *Applied Cognitive Psychology* 24 (7). John Wiley & Sons, Inc. 1003–20. doi:10.1002/acp.1602.

Lessler, Judith T. 1984. Measurement Errors in Surveys. In *Surveying Subjective Phenomena*, edited by Charles F. Turner and Elizabeth Martin, 2:405–40. New York, NY, USA: Russell Sage Foundation.

Linville, Patricia W., Gregory W. Fischer, and Peter Salovey. 1989. Perceived Distributions of the Characteristics of in-Group and out-Group Members: Empirical Evidence and a Computer Simulation. *Journal of Personality and Social Psychology* 57 (2). American Psychological Association: 165–88. doi:10.1037/0022-3514.57.2.165.

Long, John Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA, USA: Sage.

Lorenc, Boris. 2010. Item Nonresponse Analysis for a Mixed-Mode Survey. In *Official Statistics – Methodology and Applications in Honour of Daniel Thorburn*, edited by Michael Carlson, Hans Nyquist, and Mattias Villani, 117–36. Stockholm, SWE: Department of Statistics, Stockholm University.

Lorenz, Frederick O. and Vern D. Ryan. 1996. Experiments in General/specific Questions: Comparing Results of Mail and Telephone Survey. In *Proceedings of the Survey Research Methods Section*, 611–13. Alexandria, VA, USA: American Statistical Association. Available at: http://www.amstat.org/sections/srms/Proceedings/papers/1996_104.pdf (2013-10-15).

Lozar Manfreda, Katja, Zenel Batagelj, and Vasja Vehovar. 2002. Design of Web Survey Questionnaires: Three Basic Experiments. *Journal of Computer-Mediated Communication* 7 (3). doi:10.1111/j.1083-6101.2002.tb00149.x.

Lozar Manfreda, Katja, Michael Bosnjak, Nejc Berzelak, Iris Haas, and Vasja Vehovar. 2008. Web Surveys versus Other Survey Modes: A Meta-Analysis Comparing Response Rates. *International Journal of Market Research* 50 (1): 79–104.

Lozar Manfreda, Katja and Vasja Vehovar. 2002a. Do Mail and Web Surveys Provide Same Results? *Development in Social Science Methodology* 18: 149–69. Available at: http://mrvar.fdv.uni-lj.si/pub/mz/mz18/lozar1.pdf (2013-12-12).

———. 2002b. Mode Effects in Web Surveys. In *Proceedings of the Survey Research Methods Section*, 2172–77. Alexandria, VA, USA: American Statistical Association. Available at: http://www.amstat.org/sections/srms/Proceedings/y2002/Files/JSM2002-000972.pdf (2013-12-12).

Malhotra, Neil. 2008. Completion Time and Response Order Effects in Web Surveys. *Public Opinion Quarterly* 72 (5): 914–34. doi:10.1093/poq/nfn050.

Mangunkusumo, Resiti T., Joop S. Duisterhout, Nijs De Graaff, Erik J. Maarsingh, Harry J. De Koning, and Hein Raat. 2006. Internet versus Paper Mode of Health and Health Behavior Questionnaires in Elementary Schools: Asthma and Fruit as Examples. *Journal of School Health* 76 (2). Blackwell Publishing Inc: 80–86. doi:10.1111/j.1746-1561.2006.00072.x.

Martin, Jean, Colm O'Muircheartaigh, and John Curtice. 1993. The Use of CAPI for Attitude Surveys: An Experimental Comparison with Traditional Methods. *Journal of Official Statistics* 9 (3): 641–61.

Martin, Peter. 2011. *What Makes a Good Mix? Chances and Challenges of Mixed Mode Data Collection in the ESS.* Working Paper No. 02. London, GBR. Available at: http://www.europeansocialsurvey.org (2013-05-14).

Matzat, Uwe, Chris Snijders, and Wouter van der Horst. 2009. Effects of Different Types of Progress Indicators on Drop-out Rates in Web Surveys. *Social Psychology* 40 (1): 43–52. doi:10.1027/1864-9335.40.1.43.

Maynard, Rebecca, Nianbo Dong, and Irma Perez-Johnson. 2007. *Issues in Calculating Average Effect Sizes in Meta-Analyses*. Paper presented at rountable Application of Effect Sizes in Research on Children and Families, Washington, D.C., USA. Available at: http://www.acf.hhs.gov/programs/opre/resource/issues-in-calculating-average-effect-sizes-in-meta-analyses-rebecca (2013-12-18).

McCabe, Sean Esteban, Mick P. Couper, James A. Cranford, and Carol J. Boyd. 2006. Comparison of Web and Mail Surveys for Studying Secondary Consequences Associated with Substance Use: Evidence for Minimal Mode Effects. *Addictive Behaviors* 31 (1): 162–68. doi:10.1016/j.addbeh.2005.04.018.

McCarty, John A. and L. J. Shrum. 2000. The Measurement of Personal Values in Survey Research: A Test of Alternative Rating Procedures. *Public Opinion Quarterly* 64 (3): 271–98. doi:10.1086/317989.

McCullagh, Peter. 1980. Regression Models for Ordinal Data. *Journal of the Royal Statistical Society: Series B (Methodological)* 42 (2): 109–42.

McDonald, Heath and Stewart Adam. 2003. A Comparison of Online and Postal Data Collection Methods in Marketing Research. *Marketing Intelligence & Planning* 21 (2): 8–95. doi:10.1108/02634500310465399.

Millar, Morgan M. and Don A. Dillman. 2012. Encouraging Survey Response via Smartphones: Effects on Respondents' Use of Mobile Devices and Survey Response Rates. *Survey Practice* 5 (3).

Miller, Elizabeth T., Dan J. Neal, Lisa J. Roberts, John S. Baer, Sally O. Cressler, Jane Metrik, and G. Alan Marlatt. 2002. Test-Retest Reliability of Alcohol Measures: Is There a Difference between Internet-Based Assessment and Traditional Methods? *Psychology of Addictive Behaviors* 16 (1): 56–63. doi:10.1037//0893-164X.16.1.56.

Mohorko, Anja, Edith D. de Leeuw, and Joop Hox. 2013a. Coverage Bias in European Telephone Surveys: Dvelopments of Landline and Mobile Phone Coverage across Countries and over Time. *Survey Methods: Insights from the Field*. Available at: http://surveyinsights.org/?p=828 (2013-12-25).

———. 2013b. Internet Coverage and Coverage Bias in Europe: Developments Across Countries and Over Time. *Journal of Official Statistics* 29 (4): 609–622. doi: 10.2478/jos-2013-0042

Mooney, Geraldine M., Barbara A. Rogers, and Deborah H. Trunzo. 2003. *Web Experiment: Examining the Effects of Error Prompting on Item Nonresponse and Survey Nonresponse*. Paper presented at The American Association for Public Opinion Research 58th Conference, Nashville, TN, USA.

Näher, Anatol-Fiete and Ivar Krumpal. 2012. Asking Sensitive Questions: The Impact of Forgiving Wording and Question Context on Social Desirability Bias. *Quality & Quantity* 46 (5). Springer Science & Business Media B.V. 1601–16. doi:10.1007/s11135-011-9469-2.

Narayan, Sowmya and Jon A. Krosnick. 1996. Education Moderates Some Response Effects in Attitude Measurement. *Public Opinion Quarterly* 60 (1): 58–88. doi:10.1086/297739.

Nederhof, Anton J. 1985. Methods of Coping with Social Desirability Bias: A Review. *European Journal of Social Psychology* 15 (3). John Wiley & Sons, Inc. 263–80.

Newson, Roger and The ALSPAC Study Team. 2003. Multiple–test Procedures and Smile Plots. *The Stata Journal* 9 (2): 109–32. Available at: http://www.stata-journal.com/article.html?article=st0035 (2013-10-10).

NIDI. 2011. Generations and Gender Programme. Available at: http://www.ggp-i.org/ (2013-12-10).

Novick, Melvin R. 1966. The Axioms and Principal Results of Classical Test Theory. *Journal of Mathematical Psychology* 3 (1): 1–18. doi:10.1016/0022-2496(66)90002-2.

O'Connell, Ann A. 2006. *Logistic Regression Models for Ordinal Response Variables*. Quantitative Applications in the Social Sciences. Thousand Oaks, CA, USA: SAGE.

O'Neil, Michael J. 1979. Estimating the Nonresponse Bias due to Refusals in Telephone Surveys. *Public Opinion Quarterly* 43 (2): 218–32. doi:10.1086/268513.

Paulhus, Delroy L. 2002. Socially Desirable Responding : The Evolution of a Construct. In *The Role of Constructs in Psychological and Educational Measurement*, edited by Henry I. Braun, Douglas N. Jackson, and David E. Wiley, 49–69. Mahwah, NJ, USA: Lawrence Erlbaum Associates.

Pealer, Lisa N., Robert M. Weiler, R. Morgan Pigg, Jr., David Miller, and Steve M. Dorman. 2001. The Feasibility of a Web-Based Surveillance System to Collect Health Risk Behavior Data from College Students. *Health Education & Behavior* 28 (5): 547–59. doi:10.1177/109019810102800503.

Perfetti, Charles A., Nicole Landi, and Jane Oakhill. 2005. The Acquisition of Reading Comprehension Skills. In *The Neural Basis of Reading*, edited by Piers Cornelissen, Peter Hansen, Morten Kringelbach, and Ken Pugh, 227–47. Oxford, GBR: Oxford University Press.

Perneger, Thomas V. 1998. What's Wrong with Bonferroni Adjustment. *British Medical Journal* 316: 1236–38.

Peterson, Bercedis and Frank E Harrell. 1990. Partial Proportional Odds Models for Ordinal Response Variables. *Applied Statistics* 39 (2). Wiley-Blackwell: 205–17.

Peytchev, Andy, Frederick G. Conrad, Mick P. Couper, and Roger Tourangeau. 2010. Increasing Respondents' Use of Definitions in Web Surveys. *Journal of Official Statistics* 26 (4): 633–50. Available at: http://www.jos.nu/Articles/abstract.asp?article=264633 (2013-10-18).

Peytchev, Andy, Mick P. Couper, Sean Esteban McCabe, and Scott D. Crawford. 2006. Web Survey Design: Paging versus Scrolling. *Public Opinion Quarterly* 70 (4): 596–607.

Peytchev, Andy and Scott D. Crawford. 2005. A Typology of Real-Time Validations in Web-Based Surveys. *Social Science Computer Review* 23 (2): 235–49. doi:10.1177/0894439304273272.

Peytchev, Andy and Craig A. Hill. 2010. Experiments in Mobile Web Survey Design: Similarities to Other Modes and Unique Considerations. *Social Science Computer Review* 28 (3): 319–35. doi:10.1177/0894439309353037.

Platek, Richard and Carl-Erik Särndal. 2001. Can a Statistician Deliver ? *Journal of Official Statistics* 17 (1): 1–20. Available at: http://www.jos.nu/Articles/abstract.asp?article=17101 (2013-06-18).

Porter, Stephen R. and Michael E. Whitcomb. 2007. Mixed-Mode Contacts in Web Surveys: Paper Is Not Necessarily Better. *Public Opinion Quarterly* 71 (4): 635–48. doi:10.1093/poq/nfm038.

Rasinski, Kenneth A., David Mingay, and Norman M. Bradburn. 1994. Do Respondents Really "Mark All That Apply" on Self-Administered Questionnaires. *Public Opinion Quarterly* 58 (3): 400–408. doi:10.1086/269434.

Redline, Cleo D. and Don A. Dillman. 2001. The Influence of Alternative Visual Designs on Respondents' Performance with Branching Instructions in Self-Administered Questionnaires. In *Survey Nonresponse*, edited by Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, 179–93. Hoboken, NJ, USA: John Wiley & Sons.

Redline, Cleo D., Don A. Dillman, Lisa Carley-Baxter, and Robert H. Creecy. 2005. Factors That Influence Reading and Comprehension of Branching Instructions in Self-Administered Questionnaires. *Allgemeines Statistiches Archiv* 89: 21–38. doi:10.1007/s101820500189.

Revilla, Melanie A. 2013. Measurement Invariance and Quality of Composite Scores in a Face-to-Face and a Web Survey. *Survey Research Methods* 7 (1): 17–28. Available at: https://ojs.ub.uni-konstanz.de/srm/article/view/5098 (2013-06-18).

Rice, John A. 2007. *Mathematical Statistics and Data Analysis*. 2nd ed. Belmont, CA, USA: Thomson Brooks/Cole.

Richardson, John T. E. 2011. Eta Squared and Partial Eta Squared as Measures of Effect Size in Educational Research. *Educational Research Review* 6 (2): 135–47. doi:10.1016/j.edurev.2010.12.001.

Richman, Wendy L., Sara Kiesler, Suzanne Weisb, and Frith Drasgow. 1999. A Meta-Analytic Study of Social Desirability Distortion in Computer-Administered Questionnaires, Traditional Questionnaires, and Interviews. *Journal of Applied Psychology* 84 (5): 754–75.

Rivara, Frederick P., Thomas D. Koepsell, Jin Wang, Dennis. Durbin, Kenneth M. Jaffe, Monica Vavilala, Andrea Dorsch, Maria Roper-Caldbeck, Eileen Houseknecht, and Nancy Temkin. 2011. Comparison of Telephone with World Wide Web-Based Responses by Parents and Teens to a Follow-up Survey after Injury. *Health Services Research* 46 (3). Wiley-Blackwell: 964–81. doi:10.1111/j.1475-6773.2010.01236.x.

Roese, Neal J. and David W. Jamieson. 1993. Twenty Years of Bogus Pipeline Research: A Critical Review and Meta-Analysis. *Psychological Bulletin* 114 (2): 363–75.

Roster, Catherine A., Robert D. Rogers, Gerald Albaum, and Darin Klein. 2004. A Comparison of Response Characteristics from Web and Telephone Surveys. *International Journal of Market Research* 46 (3): 359–74.

Rothman, Kenneth J. 1990. No Adjustments Are Needed for Multiple Comparisons. *Epidemiology* 1 (1): 43–46. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2081237 (2013-08-15).

Russo, J. Edward and Barbara Anne Dosher. 1983. Strategies for Multiattribute Binary Choice. *Journal of Experimental Psychology* 9 (4): 676–96.

Sander, Joan E., Frederick G. Conrad, Paul A. Mullin, and Douglas J. Herrmann. 1992. Cognitive Modelling of the Survey Interview. In *Proceedings of the Survey Research Methods Section*, 818–23. Alexandria, VA, USA: American Statistical Association. Available at: http://www.amstat.org/sections/srms/Proceedings/papers/1992_138.pdf (2013-10-14).

Saris, Willem E. and Irmtraud Gallhofer. 2007. Estimation of the Effects of Measurement Characteristics on the Quality of Survey Responses. *Survey Research Methods* 1 (1): 29–43.

Sayre, J. 1939. A Comparison of Three Indices of Attitude toward Radio Advertising. *Journal of Applied Psychology* 23 (1): 23–33. doi:10.1037/h0055029.

Schaeffer, Nora Cate. 1980. Evaluating Race-of-Interviewer Effects in a National Survey. *Sociological Methods & Research* 8 (4): 400–419. doi:10.1177/004912418000800403.

Schober, Michael F. and Frederick G. Conrad. 1997. Does Conversational Interviewing Reduce Survey Measurement Error? *Public Opinion Quarterly* 61 (4): 576–602. doi:10.1086/297818.

Schuman, Howard and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys*. Thousand Oaks, CA, USA: Sage.

Schwarz, Norbert. 2007. Cognitive Aspects of Survey Methodology. *Applied Cognitive Psychology* 287: 277–87. doi:10.1002/acp.

Schwarz, Norbert and Herbert Bless. 2007. Mental Construal Processes: The Inclusion/exclusion Model. In *Assimilation and Contrast in Social Psychology*, edited by Diederik A. Stapel and Jerry Suls, 119–41. Philadelphia, PA, USA: Psychology Press.

Schwarz, Norbert and Hans-Jürgen Hippler. 1995. Subsequent Questions May Influence Answers to Preceding Questions in Mail Surveys. *Public Opinion Quarterly* 59 (1): 93–97. doi:10.1086/269460.

Schwarz, Norbert and Daphna Oyserman. 2001. Asking Questions about Behavior: Cognition, Communication, and Questionnaire Construction. *American Journal of Evaluation* 22 (2): 127–60. Available at: http://personal.stevens.edu/~ysakamot/730/paper/schwarz2001.pdf (2013-09-17).

Schwarz, Norbert, Fritz Strack, Hans-Jürgen Hippler, and George F. Bishop. 1991. The Impact of Administration Mode on Response Effects in Survey Measurement. *Applied Cognitive Psychology* 5 (3). John Wiley & Sons, Inc. 193–212.

Shin, Eunjnung, Timothy P. Johnson, and Kumar Rao. 2012. Survey Mode Effects on Data Quality: Comparison of Web and Mail Modes in a U.S. National Panel Survey. *Social Science Computer Review* 30 (2): 212–28. doi:10.1177/0894439311404508.

Smith, Tom W. 1995. Little Things Matter: A Sampler of How Differences in Questionnaire Format Can Affect Survey Responses. In *Proceedings of the Survey Research Methods Section*, 1046–51. Alexandria, VA, USA: American Statistical Association. Available at: http://www.amstat.org/sections/srms/Proceedings/papers/1995_182.pdf (2013-05-08).

———. 2011. Refining the Total Survey Error Perspective. *International Journal of Public Opinion Research* 23 (4): 464–84. doi:10.1093/ijpor/edq052.

Smither, James W., Alan G. Walker, and Michael K. T. Yap. 2004. An Examination of the Equivalence of Web-Based versus Paper-and-Pencil Upward Feedback Ratings: Rater- and Ratee-Level Analyses. *Educational and Psychological Measurement* 64 (1): 40–61. doi:10.1177/0013164403258429.

Smyth, Jolene D., Leah Melani Christian, and Don A. Dillman. 2008. Does "Yes or No" on the Telephone Mean the Same as "Chek-All-That-Apply" on the Web? *Public Opinion Quarterly* 72 (1): 103–13. doi:10.1093/poq/nfn005.

Smyth, Jolene D., Don A. Dillman, and Leah Melani Christian. 2009a. Context Effects in Internet Surveys: New Issues and Evidence. In *Oxford Handbook of Internet Psychology*, edited by Adam Joinson, Katelyn McKenna, Tom Postmes, and Ulf-Dietrich Reips, 429–46. Oxford, GBR: Oxford University Press. doi:10.1093/oxfordhb/9780199561803.013.0027.

Smyth, Jolene D., Don A. Dillman, Leah Melani Christian, and Mallory McBride. 2009b. Open-Ended Questions in Web Surveys: Can Incrasing the Size of Answer Boxes and Providing Extra Verbal Instructions Improve Response Quality? *Public Opinion Quarterly* 73 (2): 325–37. doi:10.1093/poq/nfp029.

Smyth, Jolene D., Don A. Dillman, Leah Melani Christian, and Michael J. Stern. 2006a. Effects of Using Visual Design Principles to Group Response Options in Web Surveys. *International Journal of Internet Science* 1 (1): 6–16.

———. 2006b. Comparing Check-All and Forced-Choice Question Formats in Web Surveys. *Public Opinion Quarterly* 70 (1): 66–77. doi:10.1093/poq/nfj007.

Stanton, Jeffrey M. 1998. An Empirical Assessment of Data Collection Using the Internet. *Personnel Psychology* 51 (709–725).

StataCorp. 2012. Stata/IC 11.2 (64-Bit). Computer software. College Station, TX, USA: StataCorp.

Strack, Fritz and Leonard L. Martin. 1987. Thinking, Judging, and Communicating: A Process Account of Context Effects in Attitude Surveys. In *Social Information Processing and Survey Methodology*, edited by Hans-Jürgen Hippler, Norbert Schwarz, and Seymour Sudman, 123–48. New York, NY, USA: Springer.

Strack, Fritz, Leonard L. Martin, and Norbert Schwarz. 1988. Priming and Communication: Social Determinants of Information Use in Judgments of Life Satisfaction. *European Journal of Social Psychology* 18 (5). John Wiley & Sons, Inc. 429–42.

Strack, Fritz, Norbert Schwarz, and Michaela Wänke. 1991. Semantic and Pragmatic Aspects of Context Effects in Social and Psychological Research. *Social Cognition* 9 (1). Guilford Publications Inc. 111–25. doi:10.1521/soco.1991.9.1.111.

Sudman, Seymour, Norman M. Bradburn, and Norbert Schwarz. 1996. *Thinking about Answers*. San Francisco, CA, USA: Jossey-Bass.

Tarnai, John and Don A. Dillman. 1992. Questionnaire Context as a Source of Response Differences in Mail versus Telephone Surveys. In *Context Effects in Social and Psychological Research*, edited by Norbert Schwarz and Seymour Sudman, 115–29. New York, NY, USA: Springer-Verlag.

Taylor, Humphrey. 1999. Does Internet Research Work? *International Journal of Market Research* 42 (1). Warc LTD: 51–63.

Thurstone, Louis Leon. 1927. A Law of Comparative Judgment. *Psychological Review* 34: 273–86. Available at: http://mpg.ndlab.net/wp-content/uploads/2009/07/thurstone94law.pdf (2013-08-11).

Toepoel, Vera, Marcel Das, and Arthur van Soest. 2008. Effects of Design in Web Surveys: Comparing Trained and Fresh Respondents. *Public Opinion Quarterly* 72 (5): 985–1007. doi:10.1093/poq/nfn060.

———. 2009a. Design of Web Questionnaires: The Effect of Layout in Rating Scales. *Journal of Official Statistics* 25 (4): 509–28.

Toepoel, Vera, Marcel Das, and Arthur Van Soest. 2009b. Design of Web Questionnaires: The Effects of the Number of Items per Screen. *Field Methods* 21 (2): 200–213. doi:10.1177/1525822X08330261.

Tomsic, Margie L., Darwin D. Hendel, and Roman P. Matross. 2000. *A World Wide Web Response to Student Satisfaction Surveys: Comparisons Using Paper and Internet Formats*. Paper presented at the 40th Annual meeting of the Association of Institutional Research, Cincinnati, OH, USA.

Tourangeau, Roger. 1984. Cognitive Sciences and Survey Methods. In *Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines*, edited by Thomas B. Jabine, Miron L. Straf, Judith M. Tanur, and Roger Tourangeau, 73–100. Washington, D.C., USA: National Academy Press.

Tourangeau, Roger, Frederick G. Conrad, and Mick P. Couper. 2013. *The Science of Web Surveys*. New York, NY, USA: Oxford University Press.

Tourangeau, Roger, Mick P. Couper, and Frederick G. Conrad. 2004. Spacing, Position, and Order: Interpretive Heuristics for Visual Features of Survey Questions. *Public Opinion Quarterly* 68 (3): 368–93.

———. 2007. Color, Labels, and Interpretive Heuristics for Response Scales. *Public Opinion Quarterly* 71 (1): 91–112. doi:10.1093/poq/nfl046.

Tourangeau, Roger and Kenneth A. Rasinski. 1988. Cognitive Processes Underlying Context Effects in Attitude Measurement. *Psychological Bulletin* 103 (3): 299–314.

Tourangeau, Roger, Lance J. Rips, and Kenneth A. Rasinski. 2000. *The Psychology of Survey Response*. Cambridge, GBR: Cambridge University Press.

Tourangeau, Roger and Tom W. Smith. 1996. Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context. *Public Opinion Quarterly* 60 (2): 275–304. doi:10.1086/297751.

Tourangeau, Roger and Ting Yan. 2007. Sensitive Questions in Surveys. *Psychological Bulletin* 133 (5): 859–83. doi:10.1037/0033-2909.133.5.859.

Truell, Allen D., James E. Bartlett, and Melody W. Alexander. 2002. Response Rate, Speed, and Completeness. *Behaviour Research Methods, Instruments, & Computers* 34 (1): 46–49.

Turner, Herbert M. and Robert M. Bernard. 2006. Calculating and Synthesizing Effect Sizes. *Contemporary Issues in Communication Science and Disorders* 33: 42–55. Available at: http://www.researchgate.net/publication/239576588_Calculating_and_Synthesizing_Effect_Sizes (2013-10-07).

Tuten, Tracy L., Mirta Galesic, and Michael Bosnjak. 2004. Effects of Immediate versus Delayed Notification of Prize Draw Results on Response Behavior in Web Surveys: An Experiment. *Social Science Computer Review* 22 (3): 377–84. doi:10.1177/0894439304265640.

Uhan, Samo. 1998. *Prava in Neprava Mnenja: Vpliv Konteksta v Raziskovanju Javnega Mnenja*. Ljubljana, SVN: Fakulteta za družbene vede.

Vannieuwenhuyze, Jorre T. A. and Geert Loosveldt. 2012. Evaluating Relative Mode Effects in Mixed-Mode Surveys: Three Methods to Disentangle Selection and Measurement Effects. *Sociological Methods & Research* 42 (1): 82–104. doi:10.1177/0049124112464868.

Vardigan, Mary, Pascal Heus, and Wendy Thomas. 2008. Data Documentation Initiative: Toward a Standard for the Social Sciences. *The International Journal of Digital Curation* 3 (1).

Vehovar, Vasja and Zenel Batagelj. 1996. *The Methodological Issues in WWW Surveys*. Paper presented at the International Conference on Computer-Assisted Survey Information Collection, San Antonio, TX, USA.

Vehovar, Vasja, Eva Belak, Zenel Batagelj, and Sanja Čikić. 2004. Mobile Phone Surveys: The Slovenian Case Study. *Metodološki Zvezki* 1 (1): 1–19. Available at: http://mrvar.fdv.uni-lj.si/pub/mz/mz1.1/vehovar.pdf (2013-08-15).

Vehovar, Vasja, Nejc Berzelak, and Katja Lozar Manfreda. 2010. Mobile Phones in an Environment of Competing Survey Modes: Applying Metric for Evaluation of Costs and Errors. *Social Science Computer Review* 28 (3): 303–18.

Vehovar, Vasja and Katja Lozar Manfreda. 2008. Overview: Online Surveys. In *The Handbook of Online Research Methods*, edited by Nigel G. Fielding, Raymond M. Lee, and Blank Grant, 177–94. Thousand Oaks, CA, USA: Sage.

Velez, Pauline and Steven D. Ashworth. 2007. The Impact of Item Readability on the Endorsement of the Midpoint Response in Surveys. *Survey Research Methods* 1 (2): 69–74.

Wang, Qing and Huiyou Chang. 2010. Multitasking Bar: Prototype and Evaluation of Introducing the Task Concept into a Browser. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 103–12. New York, NY, USA: Association for Computing Machinery. doi:10.1145/1753326.1753343.

Weisberg, Herbert F. 2005. *The Total Survey Error Approach*. Chicago, IL, USA: The University of Chicago Press.

Williams, Richard. 2006. Generalized Ordered Logit/partial Proportional Odds Models for Ordinal Dependent Variables. *The Stata Journal* 6 (1): 58–82. Available at: http://www.stata-journal.com/article.html?article=st0097 (2014-01-08).

Willis, Gordon B., Patricia Royston, and Deborah Bercini. 1991. The Use of Verbal Report Methods in the Development and Testing of Survey Questionnaires. *Applied Cognitive Psychology* 5 (3). John Wiley & Sons, Inc. 251–67.

Wilson, Timothy D and Sara D Hodges. 1992. Attitudes as Temporary Constructions. In *The Construction of Social Judgments*, edited by Leonard L. Martin and Abraham Tesser, 37–65. Hillsdale, NJ, USA: Lawrence Erlbaum Associates.

Wine, Jennifer, Melissa Cominole, Ruth Heuer, and John Riccobono. 2006. *Challenges of Designing and Implementing Multimode Instruments*. Paper presented at Second International Conference on Telephone Survey Methodology, Miami, FL, USA. Available at: http://www.rti.org/pubs/TSM2006_Wine_paper.pdf (2013-11-10).

Winship, Christopher and Robert D. Mare. 1984. Regression Models with Ordinal Variables. *American Sociological Review* 49: 512–25. Available at: http://scholar.harvard.edu/files/cwinship/files/asr_1984.pdf (2014-01-08).

Wolfe, Edward W., Patrick D. Converse, and Frederick L. Oswald. 2008. Item-Level Nonresponse Rates in an Attitudinal Survey of Teachers Delivered via Mail and Web. *Journal of Computer-Mediated Communication* 14 (1). Blackwell Publishing Inc: 35–66. doi:10.1111/j.1083-6101.2008.01430.x.

Wygant, Steve and Ron Lindorf. 1999. Surveying Collegiate Net Surfers -- Web Methodology or Mythology. *Quirk's Marketing Research Review*, no. July. Available at: http://www.quirks.com/articles/a1999/19990706.aspx (2013-08-12).

Yan, Ting, Frederick G. Conrad, Roger Tourangeau, and Mick P. Couper. 2011. Should I Stay or Should I Go: The Effects of Progress Feedback, Promised Task Duration, and Length of Questionnaire on Completing Web Surveys. *International Journal of Public Opinion Research* 23 (2). Oxford University Press / USA: 131–47. doi:10.1093/ijpor/edq046.

Yan, Ting and Roger Tourangeau. 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. doi:10.1002/acp.1331.

Ye, Cong, Jenna Fulton, and Roger Tourangeau. 2011. More Positive or More Extreme? A Meta-Analysis of Mode Differences in Response Choice. *Public Opinion Quarterly* 75 (2): 349–65. doi:10.1093/poq/nfr009.

Zarkovich, Slobodan S. 1966. *Quality of Statistical Data*. Roma, ITA: Food and Agriculture Organization of the United Nations.

Zeldman, Jeffrey and Ethan Marcotte. 2009. *Designing with Web Standards*. 3rd ed. Upper Saddle River, NJ, USA: Pearson.

Subject index

Author index

Appendices

Appendix A Questions used in analyses

The original question wordings used for the face-to-face interviewing are presented. Only the minimum required adjustments were made for other modes.

Gender (1.01) Sex of respondent.

1 male / 2 female

Year of birth (1.02) In what month and year were you born?

Education (1.07) What is the highest level of education you have successfully completed?

Health evaluation (7.02) How is your health in general? **(Health in general)**

1 very good / 2 good / 3 fair / 4 bad / 5 very bad

Personality (7.05) Below follow 15 statements about characteristics that may or may not apply to you. Please indicate the extent to which you think each characteristic applies to you, on a one to seven scale, where one refers to "Does not apply" and seven to "Applies perfectly". Do not spent too much time on each statement, but indicate the category that you immediately feels fit you the best.

1 does not apply / 2 / 3 / 4 / 5 / 6 / 7 applies perfectly

a) Is sometimes rude to others **(Rude to others)**
b) Does a thorough job **(Does thorough job)**
c) Is talkative **(Talkative)**
d) Worries a lot **(Worries a lot)**
e) Is original, comes up with new ideas **(Original)**
f) Has a forgiving nature **(Forgiving nature)**
g) Tends to be lazy **(Lazy)**
h) Is outgoing, sociable **(Outgoing, sociable)**
i) Gets nervously easily **(Easily nervous)**
j) Values artistic, aesthetic experiences **(Values artistic, aesthetic)**
k) Is considerate and kind to almost everyone **(Considerate and kind)**
l) Does things efficiently **(Efficient)**
m) Is reserved **(Reserved)**
n) Is relaxed, handles stress well **(Relaxed)**
o) Has an active imagination **(Active imagination)**

Sense of control (7.06) nFor the next five statements, please indicate the extent to which you agree or disagree by using a scale from 1 to 5, where 1=Strongly disagree, 2= Disagree, 3= Neither disagree nor agree, 4= Agree, 5= Strongly agree.

1 strongly agree / 2 agree / 3 neither agree nor disagree / 4 disagree / 5 strongly disagree

a) There is really no way I can solve some of the problems I have **(Cannot solve own problems)**
b) Sometimes I feel that I'm being pushed around in life **(Feel pushed around)**
c) I have little control over the things that happen to me **(Little control over things)**
d) I often feel helpless in dealing with the problems of life **(Feel helpless with life problems)**
e) There is little I can do to change many of the important things in my life **(Can change little important in life)**

Happiness (7.07) Taking all things together, how happy would you say you are? Please use this card. **(General happiness)**

0 Extremely unhappy / 1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / 9 / 10 Extremely happy / 97 DK

Loneliness (7.08) I am going to read out six statements about your current experiences. Please indicate for each of them to what extent they have applied to you recently.

1 yes / 2 more or less / 3 no

a) There are plenty of people that I can lean on in case of trouble **(Plenty people to lean on)**
b) I experience a general sense of emptiness **(General sense of emptiness)**
c) I miss having people around **(Miss having people around)**
d) There are many people that I can count on completely **(Many people to count on)**
e) Often, I feel rejected **(Feel rejected)**
f) There are enough people that I feel close to **(Enough close people)**

Depression (7.09) Please tell me how frequently did you experience the following during the previous week.

1 seldom or never / 2 sometimes / 3 often / 4 most or all of the time

a) I felt that I could not shake off the blues even with help from my family or friends **(Could not shake off blues)**
b) I felt depressed **(Felt depressed)**
c) I thought my life had been a failure **(Thought life is failure)**
d) I felt fearful **(Felt fearful)**
e) I felt lonely **(Felt lonely)**
f) I had crying spells **(Had crying spells)**
g) I felt sad **(Felt sad)**

Income adequacy (10.02) A household may have different sources of income and more than one household member may contribute to it. Thinking of your household's total monthly income, is your household able to make ends meet … **(Making ends meet)**

1 with great difficulty / 2 with difficulty / 3 with some difficulty / 4 fairly easily / 5 easily / 6 very easily

Affordable goods and services (10.03) There are some things many people cannot afford even if they would like them. Can I just check whether your household can afford these, supposing you wanted them?

1 yes / 2 no

a) keeping your home adequately warm **(Warm home)**
b) paying for a week's annual holiday away from home **(One-week holidays)**
c) replacing any worn-out furniture **(Furniture replacement)**
d) buying new, rather than second-hand clothes **(New clothes)**
e) eating meat, chicken or fish every second day **(Meat every second day)**
f) having friends or family for a drink or meal at least once a month **(Monthly dining out)**

Payment inability in 12 months (10.04) Has your household been in arrears at any time during the past 12 months, that is, unable to pay as scheduled any of the following?

1 yes / 2 no

a) rent for accommodation **(Accommodation rent)**
b) mortgage payments **(Mortgage)**
c) utility bills, such as for electricity, water, gas **(Utilities)**
d) purchase instalments or other loan repayments **(Loans)**

Religiosity (11.03) Regardless of whether you belong to a particular religion, how religious would you say you are? Please express your religiosity on a scale of 0 to 10 where 0 means 'Not at all religious' and 10 means 'Very religious'. **(Religiosity level)**

0 Not at all religious / 1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / 9 / 10 Very religious / 97 DK

Importance of religious ceremonies (11.04)

I am going to read out some statements about religious ceremonies and I would like you to tell me to what extent do you agree or disagree with each one.

1 strongly agree / 2 agree / 3 neither agree nor disagree / 4 disagree / 5 strongly disagree

a) It is important for an infant to be registered in the appropriate religious ceremony **(Infant registered in religious ceremony)**
b) It is important for people who marry in registry offices to have a religious wedding too **(Religious wedding)**
c) It is important for a funeral to include a religious ceremony **(Religious funeral)**

Planning for future (11.07) Do you generally plan for your future or do you just take each day as it comes? Please express your opinion on a scale of 0 to 10 where 0 means "I plan for my future as much as possible" and 10 means "I just take each day as it comes". **(Planning for future)**

0 I plan for future as much as possible / 1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / 9 / 10 I just take each day as it comes / 97 DK

Marriage and children (11.08) To what extent do you agree or disagree with each of the following statements?

1 strongly agree / 2 agree / 3 neither agree nor disagree / 4 disagree / 5 strongly disagree

a) Marriage is an outdated institution. **(Marriage outdated)**
b) It is all right for an unmarried couple to live together even if they have no interest in marriage. **(Living unmarried together all right)**
c) Marriage is a lifetime relationship and should never be ended. **(Marriage should not end)**
d) It is all right for a couple with an unhappy marriage to get a divorce even if they have children. **(Divorce having children all right)**
e) A woman has to have children in order to be fulfilled. **(Children needed to fulfil woman)**
f) A man has to have children in order to be fulfilled. **(Children needed to fulfil man)**
g) A child needs a home with both a father and a mother to grow up happily. **(Mother and father needed for happy child)**
h) A woman can have a child as a single parent even if she doesn't want to have a stable relationship with a man. **(Single woman having a child)**
i) Homosexual couples should have the same rights as heterosexual couples do. **(Same rights for homosexual couples)**

Family risks responsibilities (11.09) There are widely varying views on how we should care for people in our society. Please indicate for each of the topics mentioned whether you think (your own opinion) it is mainly the task for society, the family or for both.

1 mainly a task for society / 2 more a task for society than for the family / 3 a task equally for both society and the family / 4 more a task for the family than for society / 5 mainly a task for the family

a) Care for older persons in need of care at their home **(Care for older)**
b) Care for pre-school children **(Care for pre-school children)**
c) Care for schoolchildren during after-school hours **(Care for schoolchildren after school)**
d) Financial support for older people who live below subsistence level **(Financial support for older)**
e) Financial support for younger people with children who live below subsistence level **(Financial support for younger with children)**

Childcare responsibilities (11.10) To what extent do you agree or disagree with each of the following statements?

1 strongly agree / 2 agree / 3 neither agree nor disagree / 4 disagree 5 strongly disagree

a) Grandparents should look after their grandchildren if the parents of these grandchildren are unable to do so **(Grandparents should help childcare)**
b) Parents ought to provide financial help for their adult children when the children are having financial difficulties **(Parents should financially help adult children)**
c) If their adult children were in need, parents should adjust their own lives in order to help them **(Parents should adapt life to help adult children)**

Elderly-care responsibilities (11.11) I am going to read out some statements about who should take care of an elderly parent. I would like you to say to what extent you agree or disagree with them, choosing your answer from the card.

1 strongly agree / 2 agree / 3 neither agree nor disagree / 4 disagree / 5 strongly disagree

a) Children should take responsibility for caring for their parents when parents are in need **(Children should care for parents)**
b) Children should adjust their working lives to the needs of their parents **(Children should adjust work to parents' needs)**
c) Children ought to provide financial help for their parents when their parents are having financial difficulties **(Children should financially help parents)**
d) Children should have their parents to live with them when parents can no longer look after themselves **(Children should live with parents for care)**

Gender roles (11.12) To what extent do you agree with the following statements?

1 strongly agree / 2 agree / 3 neither agree nor disagree / 4 disagree / 5 strongly disagree

a) Work is good, but what most women really want is a home and children **(Women really want home and children)**
b) Being a housewife is just as fulfilling as working **(Being housewife fulfilling)**
c) It is the task of a man to earn money and that of a woman to look after the home and the family **(Man's task earning, woman's family)**
d) It is not good if the man stays at home and cares for the children and woman goes out to work **(Not good if woman works, man cares for children)**
e) The relationship between a working woman and her children can be just as close as that of a non-working mother **(Working woman same relation with child)**
f) A pre-school child will probably suffer if his/her mother works **(Pre-school child suffers if mother works)**
g) All in all family life suffers if the woman works full-time **(Family life suffers if woman works full-time)**
h) Family life often suffers because men concentrate too much on their work **(Family life suffers because men too concentrated on work)**

Overall survey experience (12.01) Overall how did you feel about completing this questionnaire? **(Feeling about survey participation)**

1 very unpleasant / 2 / 3 / 4 / 5 very enjoyable

Survey feedback (12.02) I will read you five questions about the survey you've just participated in. Please answer each question using a five point scale, where 1 means 'definitely not' and 5 means 'definitely yes'.

1 definitely not / 2 / 3 / 4 / 5 definitely yes

a) Was it difficult to answer the questions? **(Questions difficult)**

b) Were the questions clear? **(Questions clear)**
c) Did the questions made you think? **(Questions made think)**
d) Was the topic interesting? **(Topic interesting)**
e) Was the questionnaire too long? **(Questionnaire too long)**

General opinion about surveys (12.05) I will read you three statements that pertain to surveys in general and not merely to the survey you have just participated in.

1 strongly agree / 2 agree / 3 neither agree nor disagree / 4 disagree / 5 strongly disagree

a) Surveys are important for science, politics and the economy. **(Surveys important)**
b) Surveys only keep me from doing more important things. **(Surveys prevent doing other things)**
c) In surveys I have the opportunity to articulate your own opinion. **(Surveys enable own opinion articulation)**

Appendix B Impression management coding

The instructions are presented in the original form submitted to three expert evaluators. We only used the coding of potential for overclaiming as an indicator of impression manager, as explained in section 0, page 163.

Intrusiveness

Does the question inquire into topics that are inappropriate in everyday conversation, e.g. when talking to a stranger in a waiting room?

1 - no, the topic is very casual
2, 3, 4
5 - yes, the topic would be extremely inappropriate in such a situation

Threat of disclosure

Does (at least) one of the possible answers ask the respondent to admit to holding opinions or acting in ways that are not in accordance with generally accepted norms?

1 - no
2 - yes, weak/moderate norm
3 - yes, strong norm

Instruction to rater:

Please do read **each answer alternative** and carefully consider it. A typical inappropriate time saving strategy is to read the question and immediately produce the rating, ignoring the answer alternatives altogether. When this strategy is taken, it becomes very likely that the rater will overlook that one of the possible answers asks the respondent to admit to breaking a social norm.

If **no answer alternatives are** given (e.g. if the question asks "how often do you…") consider whether answers from the lower end (e.g. "never"), middle part and higher end ("very often") of the possible range are in violation of a social norm.

When thinking whether a particular answer violates a norm, consider the **sanctions** that could result from giving the answer. If resulting sanctions could be of a formal kind (as in admitting to tax fraud, for example) the rating should be (3). If the sanctions would be informal (as in most of the cases), e.g. disapproval, consider the gravity of the norm that is broken. Breaking strong norms (e.g. taking drugs) would result in strong informal sanctions, thus the rating (3) should be given. The rating (2) is appropriate when weaker norms are in question and only mild disapproval can be expected.

When the survey question pertains to a person other than the respondent (e.g. asking about the respondent's partner's employment), the respondent is acting as a **proxy**. Admitting that the partner is currently employed without a contract, for example, still constitutes a threat of disclosure, even though the target of sanctions would not be the respondent.

If **only one** possible answer is deemed threatening, then the whole questionnaire item should be rated as threatening, even if all other answer alternatives are non-threatening.

Sub-rating for threat of disclosure

(rated only if a (2) or (3) was given on threat of disclosure)

Was the positive rating (2 or 3) of the threat of disclosure based on answer alternatives that all pertain to very small proportions of the population?

0 - no, at least one of the threatening answers pertains to a substantial proportion
1 - yes all threatening answers pertain to small proportions of the population

Instruction to rater:

Often the respondent's choosing a particular answer alternative, positions him or her in a minority group, e.g. homosexuals. This is certainly within the scope of the threat of disclosure as some minority groups are discriminated against. Sometimes, however, the threatening answer alternative pertains to a minority that constitutes an **insignificantly low proportion** of the population. The question "Is your mother still alive?" includes the answer alternative "I don't know". Choosing this answer alternative would certainly break the norm of good family relations, but it is unlikely that a significant proportion of respondents would be in a situation where this answer applies.

The purpose of this additional rating is to identify those items that were rated as (2) or (3) on threat of disclosure solely on the basis of such answer alternatives that would not be threatening to the great majority of respondents. As a rule of thumb, the rating (1) should be given on this sub-variable if all threatening answer alternatives pertain to **no more than a few percent** of the population.

Potential for overclaiming

(portraying oneself in an overly positive manner)

Does (at least) one of the possible answers allow the respondent to portray him/herself in a more favorable light by claiming to hold opinions or act in ways that are generally considered desirable?

1 - no
2 - yes, it allows the respondent to portray him/herself in a **somewhat** more favorable light
3 - yes, it allows the respondent to portray him/herself in a **very** favorable light

Instruction to rater:

The same rules apply for proxy respondents as for threat of disclosure. For example, even if the respondent is answering about the **partner's** education this constitutes a potential for overclaiming.

When rating the potential for overclaiming, do not imagine **special groups** that could have specific values. Indeed, the respondent could e.g. express more intolerance toward foreigners, if he or she were asked about this in a group of neo-nazis. We are not interested in values held by any specific groups but instead **generally** considered desirable. This is an important aspect of how the potential for overclaiming differs from the threat of disclosure which often relates to minority groups.

Appendix C Additional tables from analyses

Appendix C.1 OLS regression models of between-mode differences

Item (model statistics)	b_0	CATI		CAPI	
		b	$\|t\|$	b	$\|t\|$
Health evaluation (7.02)					
Health in general ($F = 15.90^{\#\#}$, $R^2_{adj} = 0.108$, $n = 617$)	1.312	-0.165	2.25^{*}	-0.187	2.55^{*}
Personality (7.05)					
a) Rude to others ($F = 2.93^{*}$, $R^2_{adj} = 0.016$, $n = 616$)	3.224	0.134	0.80	0.017	0.10
b) Does thorough job ($F = 5.97^{\#\#}$, $R^2_{adj} = 0.039$, $n = 615$)	5.483	0.200	1.62	0.470	$3.82^{\#\#}$
c) Talkative ($F = 5.08^{\#\#}$, $R^2_{adj} = 0.032$, $n = 616$)	5.478	0.269	1.84	0.451	$3.10^{\#}$
d) Worries a lot ($F = 1.71$, $R^2_{adj} = 0.006$, $n = 616$)	3.984	-0.081	0.48	-0.095	0.57
e) Original ($F = 1.63$, $R^2_{adj} = 0.005$, $n = 616$)	4.905	0.161	1.27	0.217	1.72
f) Forgiving nature ($F = 1.77$, $R^2_{adj} = 0.006$, $n = 615$)	5.131	0.020	0.14	0.208	1.43
g) Lazy ($F = 4.41^{\#}$, $R^2_{adj} = 0.027$, $n = 616$)	4.726	-0.105	0.55	0.388	2.05^{*}
h) Outgoing, sociable ($F = 9.09^{\#\#}$, $R^2_{adj} = 0.062$, $n = 615$)	5.586	0.498	$3.51^{\#\#}$	0.668	$4.72^{\#\#}$
i) Easily nervous ($F = 2.99^{*}$, $R^2_{adj} = 0.016$, $n = 616$)	3.985	-0.219	1.32	-0.385	2.32^{*}
j) Values artistic, aesthetic experience ($F = 7.81^{\#\#}$, $R^2_{adj} = 0.053$, $n = 615$)	3.965	-0.034	0.21	0.579	$3.57^{\#\#}$
k) Considerate and kind ($F = 2.45^{*}$, $R^2_{adj} = 0.012$, $n = 616$)	5.620	0.012	0.11	0.293	2.51^{*}
l) Efficient ($F = 3.99^{\#}$, $R^2_{adj} = 0.024$, $n = 616$)	6.031	-0.107	1.02	0.164	1.57
m) Reserved ($F = 3.28^{**}$, $R^2_{adj} = 0.018$, $n = 616$)	3.060	-0.068	0.40	-0.416	2.47^{*}
n) Relaxed ($F = 2.48^{*}$, $R^2_{adj} = 0.012$, $n = 614$)	4.654	0.467	$3.31^{\#}$	0.320	2.27^{*}
o) Active imagination ($F = 1.40$, $R^2_{adj} = 0.003$, $n = 616$)	5.385	-0.160	1.18	0.097	0.71

(*Table continued on next page*)

Item (model statistics)	b_0	CATI		CAPI	
		b	$\|t\|$	b	$\|t\|$
Sense of control (7.06)					
a) Cannot solve own problems ($F = 3.58^{\#}$, $R^2_{adj} = 0.021$, $n = 616$)	3.704	0.354	$3.11^{\#}$	0.407	$3.58^{\#\#}$
b) Feel pushed around ($F = 4.06^{\#}$, $R^2_{adj} = 0.024$, $n = 615$)	3.512	0.287	2.55^{*}	0.435	$3.86^{\#\#}$
c) Little control over things ($F = 2.42^{*}$, $R^2_{adj} = 0.011$, $n = 615$)	3.885	-0.017	0.16	0.045	0.43
d) Feel helpless with life problems ($F = 2.33^{*}$, $R^2_{adj} = 0.011$, $n = 615$)	3.538	0.102	1.01	0.229	2.28^{*}
e) Can change little important in life ($F = 3.48^{\#}$, $R^2_{adj} = 0.020$, $n = 611$)	4.207	0.056	0.54	0.016	0.16
Happiness (7.07)					
General happiness ($F = 3.03^{*}$, $R^2_{adj} = 0.016$, $n = 618$)	7.830	0.359	2.30^{*}	0.397	2.55^{*}
Loneliness (7.08) *Logistic regression, recoded to* 0: "no", 1: "yes" *or* "more or less"					
a) Plenty of people to lean on ($\chi^2_{(5)} = 4.26$, $n = 615$)	n/a	1.132	0.42	1.018	0.06
b) General sense of emptiness ($\chi^2_{(5)} = 19.05^{\#}$, $n = 618$)	n/a	0.548	2.30^{*}	0.510	2.57^{*}
c) Miss having people around ($\chi^2_{(5)} = 7.92$, $n = 617$)	n/a	0.664	1.88	0.838	0.84
d) Many people to count on ($\chi^2_{(5)} = 3.76$, $n = 617$)	n/a	1.211	0.75	1.182	0.67
e) Feel rejected ($\chi^2_{(5)} = 7.32$, $n = 617$)	n/a	0.529	2.37^{*}	0.584	2.06^{*}
f) Enough close people ($\chi^2_{(5)} = 6.60$, $n = 616$)	n/a	1.579	1.09	2.830	2.11^{*}
Depression (7.09)					
a) Could not shake off blues ($F = 6.80^{\#\#}$, $R^2_{adj} = 0.045$, $n = 615$)	1.343	-0.284	$5.01^{\#\#}$	-0.015	0.26
b) Felt depressed ($F = 5.98^{\#\#}$, $R^2_{adj} = 0.039$, $n = 617$)	1.238	-0.220	$4.33^{\#\#}$	-0.082	1.61
c) Thought life is a failure ($F = 5.81^{\#\#}$, $R^2_{adj} = 0.038$, $n = 617$)	1.373	-0.241	$4.77^{\#\#}$	-0.167	$3.31^{\#}$
d) Felt fearful ($F = 10.45^{\#\#}$, $R^2_{adj} = 0.071$, $n = 617$)	1.628	-0.364	$6.63^{\#\#}$	-0.181	$3.30^{\#}$
e) Felt lonely ($F = 7.57^{\#\#}$, $R^2_{adj} = 0.051$, $n = 617$)	1.621	-0.313	$5.42^{\#\#}$	-0.165	$2.86^{\#}$
f) Had crying spells ($F = 16.08^{\#\#}$, $R^2_{adj} = 0.110$, $n = 616$)	1.392	-0.240	$5.38^{\#\#}$	-0.099	2.22^{*}
g) Felt sad ($F = 11.12^{\#\#}$, $R^2_{adj} = 0.076$, $n = 616$)	1.742	-0.351	$6.01^{\#\#}$	-0.126	2.15^{*}

(*Table continued on next page*)

Item (model statistics)	b_0	CATI		CAPI	
		b	$\|t\|$	b	$\|t\|$
Income adequacy (10.02)					
Making ends meet ($F = 10.51^{\#\#}$, $R^2_{adj} = 0.072$, $n = 617$)	3.777	0.227	1.97*	0.435	3.77##
Religiosity(11.03)					
Religiosity level ($F = 1.008$, $R^2_{adj} = 0.000$, $n = 617$)	4.521	-0.236	0.74	-0.403	1.26
Importance of religious ceremonies (11.04)					
a) Infant registered in religious ceremony ($F = 1.62$, $R^2_{adj} = 0.015$, $n = 618$)	3.337	0.276	2.05*	0.127	0.94
b) Religious wedding ($F = 2.85^{*}$, $R^2_{adj} = 0.015$, $n = 616$)	3.730	0.403	3.32#	0.160	1.32
c) Religious funeral ($F = 1.78$, $R^2_{adj} = 0.006$, $n = 617$)	3.453	0.318	2.42*	0.151	1.15
Planning for future (11.07)					
Planning for future ($F = 6.54^{\#\#}$, $R^2_{adj} = 0.043$, $n = 616$)	4.251	-0.296	1.21	-0.903	3.69##
Marriage and children (11.08)					
a) Marriage outdated ($F = 0.75$, $R^2_{adj} = 0.000$, $n = 618$	3.180	-0.006	0.05	0.011	0.09
b) Living unmarried together all right ($F = 1.96$, $R^2_{adj} = 0.008$, $n = 615$)	1.884	-0.040	0.38	-0.284	2.71**
c) Marriage should not end ($F = 3.43^{\#}$, $R^2_{adj} = 0.019$, $n = 617$)	3.425	0.017	0.14	0.176	1.43
d) Divorce having children all right ($F = 5.74^{\#\#}$, $R^2_{adj} = 0.037$, $n = 617$)	1.690	0.273	2.80#	0.036	0.37
e) Children needed to fulfil woman ($F = 4.12^{\#}$, $R^2_{adj} = 0.025$, $n = 617$)	4.405	0.072	0.65	0.092	0.83
f) Children needed to fulfil man ($F = 5.35^{\#\#}$, $R^2_{adj} = 0.034$, $n = 617$)	4.462	0.079	0.74	0.171	1.60
g) Mother and father needed for happy child ($F = 6.20^{\#\#}$, $R^2_{adj} = 0.041$, $n = 617$)	3.369	-0.125	0.95	-0.046	0.35
h) Woman w/o stable relationship with man having a child ($F = 5.70^{\#\#}$, $R^2_{adj} = 0.037$, $n = 617$)	2.487	-0.217	2.00*	-0.385	3.55##
i) Same rights for homosexual couples ($F = 3.00^{*}$, $R^2_{adj} = 0.016$, $n = 616$)	2.570	-0.011	0.08	-0.202	1.42
Family risks responsibilities (11.09)					
a) Care for older ($F = 0.91$, $R^2_{adj} = 0.000$, $n = 618$)	3.245	-0.010	0.11	0.021	0.24

(*Table continued on next page*)

Item (model statistics)	b_0	CATI		CAPI	
		b	$\|t\|$	b	$\|t\|$
b) Care for pre-school children ($F = 1.62$, $R^2_{adj} = 0.005$, $n = 618$)	3.584	0.100	1.00	0.149	1.48
c) Care for schoolchildren after school ($F = 1.14$, $R^2_{adj} = 0.001$, $n = 618$)	3.697	0.042	0.41	0.077	0.76
d) Financial support for older ($F = 2.76^{*}$, $R^2_{adj} = 0.014$, $n = 617$)	1.908	-0.197	1.97*	0.076	0.76
e) Financial support for younger with children ($F = 1.86$, $R^2_{adj} = 0.007$, $n = 615$)	1.968	-0.060	0.58	0.048	0.46
Childcare responsibilities (11.10)					
a) Grandparents should help childcare ($F = 0.62$, $R^2_{adj} = 0.000$, $n = 617$)	2.899	0.133	1.16	0.116	1.01
b) Parents should financially help adult children ($F = 0.51$, $R^2_{adj} = 0.000$, $n = 618$)	2.914	0.079	0.79	0.043	0.42
c) Parents should adapt life to help adult children ($F = 4.50^{\#\#}$, $R^2_{adj} = 0.028$, $n = 616$)	3.544	0.127	1.20	0.005	0.05
Elderly-care responsibilities (11.11)					
a) Children should care for parents ($F = 1.29$, $R^2_{adj} = 0.002$, $n = 616$)	2.014	-0.036	0.40	-0.064	0.71
b) Children should adjust work to parents' needs ($F = 6.72^{\#\#}$, $R^2_{adj} = 0.044$, $n = 617$)	3.119	0.108	1.12	0.277	2.88$^{\#}$
c) Children should financially help parents ($F = 5.99^{\#\#}$, $R^2_{adj} = 0.039$, $n = 616$)	2.074	0.307	3.35$^{\#}$	0.143	1.56
d) Children should live with parents for care ($F = 1.83$, $R^2_{adj} = 0.007$, $n = 617$)	3.340	0.044	0.46	0.197	2.06*
Gender roles (11.12)					
a) Women really want home and children ($F = 11.52^{\#\#}$, $R^2_{adj} = 0.079$, $n = 618$)	3.794	-0.181	1.71*	-0.311	2.94$^{\#}$
b) Being housewife fulfilling ($F = 11.26^{\#\#}$, $R^2_{adj} = 0.077$, $n = 617$)	3.410	0.002	0.01	-0.286	2.47*
c) Man's task earning, woman's family ($F = 10.01^{\#\#}$, $R^2_{adj} = 0.068$, $n = 618$)	4.557	0.118	1.31	0.261	2.92$^{\#}$
d) Not good if woman works, man cares for children ($F = 5.24^{\#\#}$, $R^2_{adj} = 0.033$, $n = 618$)	4.121	-0.004	0.03	-0.320	2.71**
e) Working woman same relation with child ($F = 6.17^{\#\#}$, $R^2_{adj} = 0.040$, $n = 618$)	2.025	-0.235	2.29*	-0.415	4.06$^{\#\#}$

(*Table continued on next page*)

Item (model statistics)	b_0	CATI		CAPI	
		b	$\|t\|$	b	$\|t\|$
f) Pre-school child suffers if mother works ($F = 6.99^{\#\#}$, $R^2_{adj} = 0.046$, $n = 617$)	4.388	-0.022	0.22	0.067	0.67
g) Family life suffers if mother works ($F = 8.92^{\#\#}$, $R^2_{adj} = 0.060$, $n = 617$)	4.609	0.133	1.40	0.184	1.93
h) Family life suffers because men too concentrated on work. ($F = 3.74^{\#}$, $R^2_{adj} = 0.022$, $n = 614$)	3.107	0.375	$3.34^{\#}$	0.199	1.77
Overall survey experience (12.01)					
Feeling about survey participation ($F = 2.89^{*}$, $R^2_{adj} = 0.015$, $n = 618$)	3.966	-0.072	0.89	0.179	2.24^{*}
Survey feedback (12.02)					
a) Questions difficult ($F = 1.96$, $R^2_{adj} = 0.008$, $n = 616$)	1.797	-0.099	1.05	-0.276	$2.92^{\#}$
b) Questions clear ($F = 1.78$, $R^2_{adj} = 0.006$, $n = 618$)	4.278	-0.082	0.81	0.017	0.17
c) Questions made think ($F = 6.06^{\#\#}$, $R^2_{adj} = 0.039$, $n = 618$)	3.795	-0.513	$4.09^{\#\#}$	-0.626	$5.00^{\#\#}$
d) Topic interesting ($F = 5.12^{\#\#}$, $R^2_{adj} = 0.032$, $n = 618$)	4.288	-0.163	1.98^{*}	0.081	0.97
e) Questionnaire too long ($F = 16.65^{\#\#}$, $R^2_{adj} = 0.113$, $n = 616$)	1.785	0.847	$7.13^{\#\#}$	-0.073	0.62
General opinion about surveys (12.05)					
a) Surveys important ($F = 2.74^{*}$, $R^2_{adj} = 0.014$, $n = 616$)	1.722	-0.012	0.13	-0.152	1.74
b) Surveys prevent doing other things ($F = 2.29^{*}$, $R^2_{adj} = 0.010$, $n = 616$)	4.353	-0.014	0.18	0.140	1.78
c) Surveys enable own opinion articulation ($F = 2.69^{*}$, $R^2_{adj} = 0.014$, $n = 616$)	1.532	-0.064	0.71	-0.180	2.00^{*}

Notes:

- Control variables: gender, age, and higher education.
- $^{*}p < 0.05$, $^{**}p < 0.01$, $^{\#}p < \alpha_{yek} \approx 0.0052$, $^{\#\#}p < \alpha_{bnf} \approx 0.0006$

Appendix C.2 Partial proportional odds models of between-mode differences

The table shows odds ratios and predicted changes in answer distributions due to mode after controlling for gender, age, and higher-level education. Web is defined as the reference mode.

Item (model test statistics)	CATI		CAPI	
	Odds ratio	Predicted CATI-Web	Odds ratio	Predicted CAPI-Web
Health evaluation (7.02)				
Health in general ($W_{(5)} = 67.17^{\#\#}$, $n = 617$)	0.644*		0.594#	
1 very good		0.088		0.105
2 good		-0.013		-0.018
3 fair		-0.066		-0.076
4 bad		-0.010		-0.011
5 very bad		n/a		n/a
Personality (7.05)				
a) Rude to others ($W_{(15)} = 47.38^{\#\#}$, $n = 616$)				
1 does not apply	0.768	0.058	0.678	0.087
2	1.342	-0.127	0.963	-0.078
3	1.468	-0.003	1.209	-0.043
4	1.841*	-0.006	1.687	-0.031
5	1.023	0.077	1.190	0.052
6	0.821	0.007	1.892	-0.013
7 applies perfectly	-	-0.005	-	0.026
b) Does thorough job	*Model not estimated* [a].			
c) Talkative	*Model not estimated* [a].			
d) Worries a lot ($W_{(15)} = 42.24^{\#\#}$, $n = 616$)	0.944		0.917	
1 does not apply		0.004		0.006
2		0.007		0.010
3		0.003		0.005
4		-0.001		-0.001
5		-0.004		-0.006
6		-0.005		-0.008
7 applies perfectly		-0.004		-0.006
e) Original	*Model not estimated* [a].			
f) Forgiving nature ($W_{(15)} = 38.79^{\#}$, $n = 615$)				
1 does not apply	0.589	0.006	1.769	-0.004
2	0.573	0.029	0.726	0.020
3	0.549	0.024	0.568	0.040
4	0.862	-0.033	1.249	-0.091
5	1.443	-0.115	1.746#	-0.096
6	1.328	0.035	1.619*	0.037
7 applies perfectly	-	0.053	-	0.094

(*Table continued on next page*)

Item (model test statistics)	CATI		CAPI	
	Odds ratio	Predicted CATI-Web	Odds ratio	Predicted CAPI-Web
g) Lazy				
($W_{(10)} = 37.40^{\#\#}$, $n = 616$)	0.893		1.415*	
1 does not apply		0.012		-0.032
2		0.011		-0.033
3		0.005		-0.017
4		-0.001		-0.003
5		-0.006		0.014
6		-0.008		0.025
7 applies perfectly		-0.013		0.046
h) Outgoing, sociable				
($W_{(5)} = 42.15^{\#\#}$, $n = 615$)	1.730#		2.292##	
1 does not apply		-0.007		-0.009
2		-0.022		-0.029
3		-0.037		-0.051
4		-0.044		-0.065
5		-0.024		-0.045
6		0.033		0.039
7 applies perfectly		0.099		0.159
i) Easily nervous				
($W_{(5)} = 16.97^{\#}$, $n = 616$)	0.801		0.631**	
1 does not apply		0.031		0.070
2		0.022		0.041
3		0.001		-0.004
4		-0.011		-0.025
5		-0.020		-0.040
6		-0.014		-0.026
7 applies perfectly		-0.009		-0.016
j) Values artistic, aesthetic experience				
($W_{(10)} = 54.48^{\#\#}$, $n = 615$)	1.022		1.886##	
1 does not apply		-0.001		-0.022
2		-0.001		-0.035
3		-0.002		-0.052
4		-0.001		-0.037
5		-0.000		-0.003
6		-0.002		0.061
7 applies perfectly		-0.002		0.087
k) Considerate and kind				
($W_{(10)} = 26.05^{\#}$, $n = 616$)			1.611**	
1 does not apply	0.142	0.017		-0.001
2	0.248*	0.016		-0.003
3	0.400*	0.021		-0.011
4	0.890	-0.040		-0.033
5	1.405	-0.093		-0.060
6	1.315	0.027		0.014
7 applies perfectly	-	0.052		0.094

(*Table continued on next page*)

Item (model test statistics)	CATI		CAPI	
	Odds ratio	Predicted CATI-Web	Odds ratio	Predicted CAPI-Web
l) Efficient				
($W_{(5)} = 14.16^{*}$, $n = 616$)	0.852		1.237	
1 does not apply		0.000		-0.000
2		0.002		-0.002
3		0.003		-0.004
4		0.010		-0.012
5		0.020		-0.027
6		-0.004		-0.001
7 applies perfectly		-0.031		-0.045
m) Reserved				
($W_{(10)} = 29.27^{\#}$, $n = 616$)	0.963		0.655*	
1 does not apply		0.004		0.055
2		0.004		0.038
3		0.002		0.010
4		-0.001		-0.017
5		-0.003		-0.036
6		-0.003		-0.034
7 applies perfectly		-0.002		-0.017
n) Relaxed				
($W_{(5)} = 12.69^{*}$, $n = 614$)	1.790$^{\#}$		1.463*	
1 does not apply		-0.010		-0.007
2		-0.028		-0.020
3		-0.038		-0.026
4		-0.049		-0.032
5		-0.011		-0.002
6		0.064		0.043
7 applies perfectly		0.071		0.043
o) Active imagination				
($W_{(5)} = 5.82$, $n = 616$)	0.839		1.142	
1 does not apply		0.001		-0.001
2		0.006		-0.004
3		0.009		-0.006
4		0.018		-0.013
5		0.009		-0.009
6		-0.013		0.008
7 applies perfectly		-0.031		0.025
Sense of control (7.06)				
a) Cannot solve own problems				
($W_{(8)} = 31.09^{\#\#}$, $n = 616$)			2.062$^{\#\#}$	
1 strongly agree	0.889	0.006		-0.027
2 agree	1.500	-0.063		-0.063
3 neither agree nor disagree	1.659*	-0.061		-0.072
4 disagree	2.411$^{\#\#}$	-0.076		0.007
5 strongly disagree	-	0.193		0.155
b) Feel pushed around				
($W_{(11)} = 45.46^{\#\#}$, $n = 615$)			2.261$^{\#\#}$	
1 strongly agree	0.585	0.022		-0.018
2 agree	1.565	-0.088		-0.088
3 neither agree nor disagree	1.488*	-0.025		-0.065
4 disagree	2.140$^{\#\#}$	-0.078		-0.010
5 strongly disagree	-	0.168		0.182

(*Table continued on next page*)

Item (model test statistics)	CATI		CAPI	
	Odds ratio	Predicted CATI-Web	Odds ratio	Predicted CAPI-Web
c) Little control over things ($W_{(20)} = 58.30^{\#\#}$, $n = 615$)				
1 strongly agree	0.192[#]	0.049	0.141[#]	0.067
2 agree	0.687	-0.010	0.773	-0.042
3 neither agree nor disagree	1.035	-0.047	1.086	-0.044
4 disagree	1.206	-0.028	1.638[*]	-0.082
5 strongly disagree	-	0.035	-	0.099
d) Feel helpless with life problems ($W_{(5)} = 10.63$, $n = 615$)	1.260		1.507[*]	
1 strongly agree		-0.006		-0.010
2 agree		-0.018		-0.031
3 neither agree nor disagree		-0.028		-0.049
4 disagree		0.007		0.006
5 strongly disagree		0.045		0.083
e) Can change little important in life ($W_{(8)} = 26.40^{\#}$, $n = 611$)	1.202		1.058	
1 strongly agree		-0.006		-0.002
2 agree		-0.012		-0.004
3 neither agree nor disagree		-0.020		-0.006
4 disagree		0.000		0.001
5 strongly disagree		0.038		0.011
Happiness (7.07)				
General happiness	*Model not estimated* [a)].			
Loneliness (7.08)				
a) Plenty people to lean on ($W_{(8)} = 75.76^{\#\#}$, $n = 615$)				
1 yes	0.238[##]	0.322	0.405[##]	0.200
2 more or less	0.891	-0.310	0.974	-0.197
3 no	-	-0.013	-	-0.003
b) General sense of emptiness ($W_{(6)} = 27.25^{\#\#}$, $n = 618$)			1.988[**]	
1 yes	0.560	0.024		-0.016
2 more or less	1.834[*]	-0.115		-0.085
3 no	-	0.091		0.101
c) Miss having people around ($W_{(5)} = 8.20$, $n = 617$)	1.531[*]		1.179	
1 yes		-0.039		-0.017
2 more or less		-0.050		-0.019
3 no		0.089		0.036
d) Many people to count on ($W_{(7)} = 38.70^{\#\#}$, $n = 617$)			0.646[*]	
1 yes	0.452[##]	0.187		0.101
2 more or less	0.744	-0.140		-0.035
3 no	-	-0.046		-0.066

(*Table continued on next page*)

Item (model test statistics)	CATI		CAPI	
	Odds ratio	Predicted CATI-Web	Odds ratio	Predicted CAPI-Web
e) Feel rejected				
($W_{(5)} = 7.32$, $n = 617$)	1.871*		1.706*	
1 yes		-0.025		-0.022
2 more or less		-0.064		-0.056
3 no		0.089		0.078
f) Enough close people				
($W_{(5)} = 39.11$##, $n = 616$)	0.378##		0.510#	
1 yes		0.194		0.142
2 more or less		-0.146		-0.105
3 no		-0.048		-0.037
Depression (7.09)				
a) Could not shake off blues				
($W_{(5)} = 34.43$##, $n = 615$)	0.218##		0.933	
1 seldom or never		0.237		0.015
2 sometimes		-0.189		-0.011
3 often		-0.038		-0.003
4 most or all of the time		-0.011		-0.001
b) Felt depressed				
($W_{(5)} = 27.78$##, $n = 617$)	0.295##		0.741	
1 seldom or never		0.166		0.053
2 sometimes		-0.130		-0.040
3 often		-0.029		-0.010
4 most or all of the time		-0.007		-0.002
c) Thought life is failure	*Model not estimated* [a].			
d) Felt fearful				
($W_{(5)} = 50.48$##, $n = 617$)	0.197##		0.555#	
1 seldom or never		0.306		0.135
2 sometimes		-0.252		-0.106
3 often		-0.044		-0.024
4 most or all of the time		-0.009		-0.005
e) Felt lonely				
($W_{(5)} = 38.01$##, $n = 617$)	0.251##		0.510#	
1 seldom or never		0.252		0.144
2 sometimes		-0.181		-0.100
3 often		-0.065		-0.041
4 most or all of the time		-0.006		-0.004
f) Had crying spells	*Model not estimated* [a].			
g) Felt sad				
($W_{(7)} = 55.92$##, $n = 616$)	0.273##		0.659*	
1 seldom or never		0.295		0.101
2 sometimes		-0.233		-0.073
3 often		-0.059		-0.026
4 most or all of the time		-0.004		-0.002

(*Table continued on next page*)

Item (model test statistics)	CATI		CAPI	
	Odds ratio	Predicted CATI-Web	Odds ratio	Predicted CAPI-Web
Income adequacy (10.02)				
Making ends meet				
($W_{(5)} = 51.94^{\#\#}$, $n = 617$)	1.367		$1.950^{\#\#}$	
1 with great difficulty		-0.017		-0.032
2 with difficulty		-0.027		-0.053
3 with some difficulty		-0.029		-0.071
4 fairly easily		0.023		0.037
5 easily		0.037		0.084
6 very easily		0.014		0.034
Affordable goods and services (10.03)				
a) Warm home				
($\chi^2_{(5)} = 2.75$, $n = 617$)	0.878		0.736	
1 yes		0.005		-0.010
2 no		-0.005		0.010
b) One-week holidays				
($\chi^2_{(5)} = 39.54^{\#\#}$, $n = 617$)	0.718		$0.467^{\#}$	
1 yes		0.05		0.10
2 no		-0.05		-0.10
c) Furniture replacement				
($\chi^2_{(5)} = 60.06^{\#\#}$, $n = 614$)	0.666		$0.546^{\#}$	
1 yes		0.085		0.123
2 no		-0.085		-0.123
d) New clothes				
($\chi^2_{(5)} = 46.26^{\#\#}$, $n = 616$)	0.590		0.700	
1 yes		0.055		0.040
2 no		-0.055		-0.040
e) Meat every second day				
($\chi^2_{(5)} = 27.52^{\#\#}$, $n = 616$)	0.562		0.552^{*}	
1 yes		0.064		0.066
2 no		-0.064		-0.066
f) Monthly dining out				
($\chi^2_{(5)} = 37.02^{\#\#}$, $n = 617$)	$0.452^{\#}$		$0.354^{\#\#}$	
1 yes		0.102		0.123
2 no		-0.102		-0.123
Payment inability in 12 months (10.04)				
a) Accommodation rent				
($\chi^2_{(5)} = 6.11$, $n = 615$)	0.459		1.514	
1 yes		0.032		-0.010
2 no		-0.032		0.010
b) Mortgage				
($\chi^2_{(5)} = 1.07$, $n = 612$)	0.636		0.986	
1 yes		0.011		0.000
2 no		-0.011		0.000
c) Utilities				
($\chi^2_{(5)} = 18.32^{\#}$ $n = 616$)	1.751		$2.651^{\#}$	
1 yes		-0.070		-0.106
2 no		0.070		0.106

(*Table continued on next page*)

Item (model test statistics)	CATI		CAPI	
	Odds ratio	Predicted CATI-Web	Odds ratio	Predicted CAPI-Web
d) Loans				
($\chi^2_{(5)} = 18.15^{\#}$ $n = 613$)	1.294		3.627$^{\#\#}$	
1 yes		-0.033		-0.116
2 no		0.033		0.116
Religiosity(11.03)				
Religiosity level				
($W_{(14)} = 27.87^{*}$, $n = 617$)			0.763	
0 Not at all religious	1.046	-0.008		0.055
1	0.831	0.051		0.009
2	0.880	-0.012		0.003
3	0.950	-0.019		-0.003
4	0.863	0.023		-0.001
5	0.892	-0.011		-0.009
6	0.633*	0.054		-0.008
7	0.894	-0.063		-0.013
8	0.675	0.016		-0.012
9	0.673	-0.016		-0.012
10 Very religious	-	-0.016		-0.012
Importance of religious ceremonies (11.04)				
a) Infant registered in religious ceremony				
($W_{(8)} = 20.91^{\#}$, $n = 618$)				
1 strongly agree			1.224	-0.014
2 agree	0.727	0.028		-0.024
3 neither agree nor disagree	1.412	-0.091		-0.012
4 disagree	1.582*	-0.050		0.004
5 strongly disagree	1.820$^{\#}$	-0.031		0.046
	-	0.143		
b) Religious wedding				
($W_{(5)} = 19.05^{\#}$, $n = 616$)	2.120$^{\#\#}$		1.241	
1 strongly agree		-0.038		-0.014
2 agree		-0.048		-0.016
3 neither agree nor disagree		-0.071		-0.020
4 disagree		-0.028		-0.003
5 strongly disagree		0.185		0.053
c) Religious funeral				
($W_{(8)} = 19.13^{*}$, $n = 617$)			1.237	
1 strongly agree	0.787	0.020		-0.015
2 agree	1.433	-0.084		-0.025
3 neither agree nor disagree	1.843$^{\#}$	-0.085		-0.014
4 disagree	1.721$^{\#}$	0.022		0.005
5 strongly disagree	-	0.127		0.048

(*Table continued on next page*)

Item (model test statistics)	CATI		CAPI	
	Odds ratio	Predicted CATI-Web	Odds ratio	Predicted CAPI-Web
Planning for future (11.07)				
Planning for future ($W_{(5)} = 29.86^{\#\#}$, $n = 616$)	0.817		0.537$^{\#\#}$	
0 I plan f. fut. as much as possible		0.001		0.035
1		0.012		0.042
2		0.015		0.045
3		0.010		0.026
4		0.003		0.003
5		-0.006		-0.029
6		-0.008		-0.026
7		-0.010		-0.029
8		-0.012		-0.035
9		-0.003		-0.009
10 I just take each day as it comes		-0.009		-0.022
Marriage and children (11.08)				
a) Marriage outdated ($W_{(11)} = 37.90^{\#\#}$, $n = 618$)			1.115	
1 strongly agree	0.848	0.016		-0.010
2 agree	0.733	0.036		-0.007
3 neither agree nor disagree	0.878	-0.020		-0.011
4 disagree	1.909$^{\#}$	-0.154		0.009
5 strongly disagree	-	0.122		0.018
b) Living unmarried together all right ($W_{(8)} = 31.52^{\#\#}$, $n = 617$)			0.513$^{\#\#}$	
1 strongly agree	0.578$^{\#}$	0.133		0.163
2 agree	1.206	-0.171		-0.054
3 neither agree nor disagree	1.345	0.011		-0.067
4 disagree	1.262	0.017		-0.022
5 strongly disagree	-	0.010		-0.019
c) Marriage should not end ($W_{(11)} = 42.33^{\#\#}$, $n = 617$)				
1 strongly agree	0.441*	0.066	0.757	0.018
2 agree	0.675	-0.005	1.000	-0.018
3 neither agree nor disagree	1.144	-0.094	1.831$^{\#}$	-0.142
4 disagree	1.760**	-0.084	1.382	0.077
5 strongly disagree	-	0.118	-	0.064
d) Divorce having children all right	*Model not estimated* [a)].			
e) Children needed to fulfil woman ($W_{(8)} = 30.78^{\#\#}$, $n = 617$)	1.266		1.227	
1 strongly agree		-0.007		-0.006
2 agree		-0.023		-0.020
3 neither agree nor disagree		-0.022		-0.019
4 disagree		0.000		0.001
5 strongly disagree		0.053		0.046

(*Table continued on next page*)

Item (model test statistics)	CATI		CAPI	
	Odds ratio	Predicted CATI-Web	Odds ratio	Predicted CAPI-Web
f) Children needed to fulfil man ($W_{(8)} = 34.46^{\#\#}$, $n = 617$)			1.424*	
1 strongly agree	0.411	0.023		-0.005
2 agree	0.825	0.001		-0.032
3 neither agree nor disagree	1.241	-0.072		-0.040
4 disagree	1.481*	-0.040		-0.002
5 strongly disagree	-	0.088		0.079
g) Mother and father needed for happy child ($W_{(8)} = 42.90^{\#\#}$, $n = 617$)				
1 strongly agree			0.950	
2 agree	0.539#	0.115		0.008
3 neither agree nor disagree	0.808	-0.064		0.004
4 disagree	0.986	-0.048		0.002
5 strongly disagree	1.413	-0.042		0.005
	-	0.040		0.040
h) Woman w/o stable relationship with man having a child ($W_{(5)} = 28.56^{\#\#}$, $n = 617$)	0.675*		0.531##	
1 strongly agree		0.074		0.125
2 agree		0.019		0.020
3 neither agree nor disagree		-0.038		-0.063
4 disagree		-0.038		-0.057
5 strongly disagree		-0.017		-0.026
i) Same rights for homosexual couples ($W_{(11)} = 35.59^{\#\#}$, $n = 616$)	0.993			
1 strongly agree		0.001	1.051	-0.009
2 agree		0.000	0.726	0.088
3 neither agree nor disagree		-0.000	0.803	-0.030
4 disagree		-0.001	0.587*	0.026
5 strongly disagree		-0.001	-	-0.075
Family risk responsibility scale (11.09)				
a) Care for older ($W_{(8)} = 10.88$, $n = 618$)	1.026			
1 mainly a task for society		-0.001	1.564	-0.021
2 more a task for society than for the family		-0.002	1.910	-0.050
3 a task equally for both society and the family		-0.002	0.849	-0.103
4 more a task for the family than for society				
5 mainly a task for the family		0.003	0.669	-0.004
		0.002	-	-0.027
b) Care for pre-school children ($W_{(8)} = 19.77^{*}$, $n = 618$)	1.233		1.315	
1 mainly a task for society		-0.008		-0.010
2 more a task for society than for the family		-0.009		-0.011
3 a task equally for both society and the family		-0.035		-0.046
4 more a task for the family than for society				
5 mainly a task for the family		0.015		0.018
		0.037		0.050

(*Table continued on next page*)

Item (model test statistics)	CATI		CAPI	
	Odds ratio	Predicted CATI-Web	Odds ratio	Predicted CAPI-Web
c) Care for schoolchildren after school ($W_{(8)} = 13.15$, $n = 618$)	1.087		1.143	
1 mainly a task for society		-0.004		-0.006
2 more a task for society than for the family		-0.005		-0.008
3 a task equally for both society and the family		-0.012		-0.020
4 more a task for the family than for society		0.008		0.012
5 mainly a task for the family		0.013		0.021
d) Financial support for older ($W_{(5)} = 15.01^{*}$, $n = 617$)	0.712		1.180	
1 mainly a task for society		0.079		-0.036
2 more a task for society than for the family		0.000		-0.005
3 a task equally for both society and the family		-0.061		-0.029
4 more a task for the family than for society		-0.012		0.007
5 mainly a task for the family		-0.006		0.003
e) Financial support for younger with children ($W_{(11)} = 28.00^{\#}$, $n = 615$)				
1 mainly a task for society	0.927			
2 more a task for society than for the family		0.016	0.926	0.016
3 a task equally for both society and the family		0.002	1.386	-0.095
4 more a task for the family than for society		-0.012	0.953	0.084
5 mainly a task for the family		-0.004	0.789	0.002
		-0.002	-	-0.006
Childcare responsibilities (11.10)				
a) Grandparents should help childcare ($W_{(11)} = 26.31^{\#}$, $n = 617$)				
1 strongly agree	0.521^{*}	0.062	0.567	0.052
2 agree	1.124	-0.088	0.967	-0.044
3 neither agree nor disagree	1.305	-0.030	1.479	-0.093
4 disagree	$4.220^{\#\#}$	-0.058	$3.500^{\#}$	-0.006
5 strongly disagree	-	0.114	-	0.091
b) Parents should financially help adult children ($W_{(8)} = 13.92$, $n = 618$)				
1 strongly agree	1.164		1.035	
2 agree		-0.012		-0.003
3 neither agree nor disagree		-0.022		-0.005
4 disagree		-0.007		-0.002
5 strongly disagree		0.019		0.004
		0.009		0.002

(*Table continued on next page*)

Item (model test statistics)	CATI		CAPI	
	Odds ratio	Predicted CATI-Web	Odds ratio	Predicted CAPI-Web
c) Parents should adapt life to help adult children ($W_{(14)} = 48.09^{\#\#}$, $n = 616$)				
1 strongly agree				
2 agree	0.373*	0.055	0.313#	0.070
3 neither agree nor disagree	1.236	-0.100	1.121	-0.095
4 disagree	1.205	0.005	0.892	0.048
5 strongly disagree	3.138#	-0.053	2.309*	-0.082
	-	0.093	-	0.059
Elderly-care responsibilities (11.11)				
a) Children should care for parents ($W_{(14)} = 29.82^{**}$, $n = 616$)				
1 strongly agree	0.517#	0.114	0.531**	0.109
2 agree	1.179	-0.147	1.186	-0.143
3 neither agree nor disagree	1.512	0.004	1.246	0.020
4 disagree	5.576	0.006	2.970	0.004
5 strongly disagree	-	0.023	-	0.010
b) Children should adjust work to parents' needs ($W_{(11)} = 49.76^{\#\#}$, $n = 617$)				
1 strongly agree			1.698#	
2 agree	0.949	0.002		-0.017
3 neither agree nor disagree	0.765	0.032		-0.035
4 disagree	1.490*	-0.131		-0.076
5 strongly disagree	1.425	0.053		0.059
	-	0.043		0.069
c) Children should financially help parents ($W_{(11)} = 55.27^{\#\#}$, $n = 616$)				
1 strongly agree				
2 agree	0.643	0.047	0.708	0.035
3 neither agree nor disagree	2.319##	-0.249	1.880#	-0.187
4 disagree	2.597#	0.079	1.210	0.133
5 strongly disagree	4.136	0.094	2.503	0.004
	-	0.029	-	0.014
d) Children should live with parents for care ($W_{(17)} = 42.92^{\#\#}$, $n = 617$)				
1 strongly agree				
2 agree	1.107	-0.003	0.793	0.008
3 neither agree nor disagree	0.616	0.055	0.709	0.026
4 disagree	1.145	-0.085	1.465	-0.129
5 strongly disagree	1.830	-0.027	3.123##	-0.045
	-	0.060		0.139
Gender roles (11.12)				
a) Women really want home and children ($W_{(11)} = 68.33^{\#\#}$, $n = 618$)				
1 strongly agree				
2 agree	0.173#	0.062	0.095##	0.116
3 neither agree nor disagree	1.027	-0.067	0.799	-0.075
4 disagree	0.629*	0.110	0.561#	0.089
5 strongly disagree	0.749	-0.070	0.800	-0.102
	-	-0.036	-	-0.028

(*Table continued on next page*)

Item (model test statistics)	CATI		CAPI	
	Odds ratio	Predicted CATI-Web	Odds ratio	Predicted CAPI-Web
b) Being housewife fulfilling				
($W_{(8)} = 62.14^{\#\#}$, $n = 617$)	0.992		0.645*	
1 strongly agree		0.001		0.066
2 agree		0.001		0.039
3 neither agree nor disagree		-0.000		-0.023
4 disagree		-0.001		-0.049
5 strongly disagree		-0.001		-0.033

(Table continued on next page.)

c) Man's task earning, woman's family				
($W_{(11)} = 65.60^{\#\#}$, $n = 618$)			1.802#	
1 strongly agree	0.370	0.019		-0.005
2 agree	0.821	-0.008		-0.019
3 neither agree nor disagree	1.109	-0.030		-0.068
4 disagree	1.772#	-0.113		-0.043
5 strongly disagree	-	0.131		0.135
d) Not good if woman works, man cares for children				
($W_{(8)} = 40.51^{\#\#}$, $n = 618$)				
1 strongly agree	1.060			
2 agree		-0.004	0.507*	0.057
3 neither agree nor disagree		-0.004	0.444##	0.082
4 disagree		-0.006	0.627*	-0.025
5 strongly disagree		0.004	1.129	-0.135
		0.010	-	0.022
e) Working woman same relation with child				
($W_{(8)} = 44.34^{\#\#}$, $n = 618$)				
1 strongly agree			0.400##	
2 agree	0.443##	0.177		0.202
3 neither agree nor disagree	0.783	-0.130		-0.052
4 disagree	0.921	-0.038		-0.071
5 strongly disagree	0.759	0.001		-0.052
	-	-0.011		-0.027
f) Pre-school child suffers if mother works				
($W_{(8)} = 45.61^{\#\#}$, $n = 617$)				
1 strongly agree			1.267	
2 agree	0.463	0.029		-0.005
3 neither agree nor disagree	0.655	0.017		-0.015
4 disagree	0.887	-0.022		-0.023
5 strongly disagree	1.616*	-0.124		-0.004
		0.100		0.047
g) Family life suffers if woman works full-time				
($W_{(8)} = 57.55^{\#\#}$, $n = 617$)				
1 strongly agree			1.379	
2 agree	0.385	0.021		-0.004
3 neither agree nor disagree	0.701	0.013		-0.020
4 disagree	1.084	-0.048		-0.031
5 strongly disagree	1.930#	-0.131		-0.014
		0.145		0.068

(*Table continued on next page*)

Item (model test statistics)	CATI		CAPI	
	Odds ratio	Predicted CATI-Web	Odds ratio	Predicted CAPI-Web
h) Family life suffers because men too concentrated on work, ($W_{(11)} = 41.01^{\#\#}$, $n = 614$)				
1 strongly agree	0.585	0.031	0.378*	0.070
2 agree	1.933#	-0.183	1.137	-0.101
3 neither agree nor disagree	2.162##	-0.017	1.767**	-0.089
4 disagree	2.699**	0.092	3.382#	0.017
5 strongly disagree	-	0.077	-	0.104
Overall survey experience (12.01)				
Feeling about survey participation ($W_{(5)} = 20.68^{\#}$, $n = 618$)	0.858		1.769#	
1 very unpleasant		0.002		-0.006
2		0.003		-0.009
3		0.022		-0.067
4		0.002		-0.045
5 very enjoyable		-0.030		0.127
Survey feedback (12.02)				
a) Questions difficult ($W_{(5)} = 10.25$, $n = 616$)	0.739		0.577#	
1 definitely not		0.071		0.124
2		-0.023		-0.044
3		-0.027		-0.047
4		-0.017		-0.029
5 definitely yes		-0.003		-0.005
b) Questions clear ($W_{(11)} = 28.27^{\#}$, $n = 618$)				
1 definitely not	0.147	0.026	0.125	0.031
2	0.847	-0.010	1.340	-0.056
3	0.629	0.050	0.795	0.055
4	1.097	-0.088	1.234	-0.081
5 definitely yes	-	0.022	-	0.051
c) Questions made think ($W_{(5)} = 27.28^{\#\#}$, $n = 618$)	0.505##		0.436##	
1 definitely not		0.066		0.084
2		0.055		0.068
3		0.047		0.052
4		-0.045		-0.060
5 definitely yes		-0.123		-0.143
d) Topic interesting ($W_{(5)} = 23.63^{\#\#}$, $n = 618$)	0.775		1.338	
1 definitely not		-0.003		-0.003
2		0.005		-0.005
3		0.029		-0.028
4		0.024		-0.035
5 definitely yes		-0.062		0.071
e) Questionnaire too long ($W_{(8)} = 81.34^{\#\#}$, $n = 616$)				
1 definitely not	2.401##	-0.193	0.808	0.052
2	3.841##	-0.123		-0.011
3	3.579##	0.099		-0.019
4	4.631##	0.118		-0.017
5 definitely yes	-	0.099		-0.006

Item (model test statistics)	CATI		CAPI	
	Odds ratio	Predicted CATI-Web	Odds ratio	Predicted CAPI-Web
General opinion about surveys (12.05)				
a) Surveys important	*Model not estimated*[a)].			
b) Surveys prevent doing other things	*Model not estimated* [a)].			
c) Surveys enable own opinion articulation ($W_{(11)} = 51.39^{\#\#}$, $n = 616$)				
1 strongly agree				
2 agree	0.461[##]	0.184	0.388[##]	0.226
3 neither agree nor disagree	1.263	-0.217	1.178	-0.249
4 disagree	3.883[**]	-0.028	1.719	0.007
5 strongly disagree	3.256	0.041	2.130	0.006
	-	0.021	-	0.010

Notes:

a) Model failed to converge, presumably due to low or zero cell frequencies.

- Control variables: gender, age, and higher education.
- $^{*}\, p < 0.05$, $^{**}\, p < 0.01$, $^{\#}\, p < \alpha_{yek} \approx 0.0078$, $^{\#\#}\, p < \alpha_{bnf} \approx 0.0006$

Appendix C.3 Odds ratios of extreme and midpoint responses

	Lower extreme		Midpoint		Upper extreme	
Item	CATI	CAPI	CATI	CAPI	CATI	CAPI
7.02	1.692*	2.129#	0.615	0.767	1.000	1.000
7.05A	1.289	1.462	0.934	0.690	0.757	1.698
7.05B	0.474	1.000	0.492*	0.509	1.511	2.227##
7.05C	1.518	0.994	0.786	0.443*	1.306	1.886#
7.05D	1.404	1.912	1.602	1.082	0.676	1.017
7.05E	1.000	1.000	0.786	1.019	1.323	1.464
7.05F	1.507	0.487	0.718	0.291**	1.270	1.579*
7.05G	1.405	1.341	0.740	0.773	0.713	1.842*
7.05H	0.975	1.411	0.487*	0.670	1.333	2.121#
7.05I	1.639	2.058**	0.637	0.517*	0.621	0.720
7.05J	1.623	0.636	0.716	0.983	1.250	1.898*
7.05K	4.572	1.000	0.540	0.529	1.160	1.305
7.05L	1.000	1.000	0.817	0.349*	0.880	1.074
7.05M	1.382	2.107	0.849	0.919	0.466	0.655
7.05N	0.293	0.278	0.668	0.676	1.507	1.164
7.05O	1.000	1.000	0.865	0.946	0.947	1.071
7.06A	1.183	0.646	0.662	0.605*	2.659##	2.453##
7.06B	1.902	0.628	0.788	0.525*	2.486##	2.823##
7.06C	2.234	3.232*	0.726	0.748	1.209	1.636*
7.06D	1.030	0.484	0.720	1.005	1.486	1.795**
7.06E	1.966	1.278	0.778	1.087	1.341	1.331
7.07	1.000	1.000	0.325*	0.430	1.301	1.464
7.08A	4.383##	2.492##	0.239##	0.425##	0.883	0.982
7.08B	1.707	0.379	0.393**	0.567*	1.824*	1.961*
7.08C	0.587	0.959	0.786	0.811	1.505	1.193
7.08D	2.395##	1.759**	0.455##	0.637*	0.826	0.846
7.08E	0.666	0.640	0.518*	0.596	1.890*	1.713*
7.08F	2.707##	1.925#	0.369##	0.628*	0.633	0.353*
7.09A	4.646##	1.089	na	na	0.332	0.655
7.09B	3.319##	1.321	na	na	1.000	1.000
7.09C	3.208##	1.622*	na	na	1.000	1.000
7.09D	5.120##	1.806**	na	na	0.315	1.000
7.09E	4.105##	2.088#	na	na	1.198	1.210
7.09F	7.985##	1.721*	na	na	1.067	1.000
7.09G	3.574##	1.499*	na	na	1.067	1.000
10.02	0.745	0.854	na	na	2.505	2.067
11.03	1.061	1.600*	1.038	1.283	0.926	1.501
11.04A	1.802	1.409	0.735	0.941	1.846#	1.271
11.04B	0.865	0.639	0.450**	0.621	2.267##	1.143
11.04C	1.300	0.830	0.596*	0.946	1.751**	1.269
11.07	6.113*	13.296##	1.463	1.300	0.734	0.773
11.08A	1.680	1.674	0.734	0.649	2.108#	1.359

	Lower extreme		Midpoint		Upper extreme	
Item	CATI	CAPI	CATI	CAPI	CATI	CAPI
11.08B	1.798**	2.074##	0.933	0.426**	1.559	0.869
11.08C	2.419*	1.442	0.593*	0.424##	1.784**	1.423
11.08D	1.142	1.655*	1.420	0.831	0.939	1.000
11.08E	10.177*	7.801	0.589*	0.795	1.590*	1.367
11.08F	8.265*	4.658	0.637	0.800	1.549*	1.560*
11.08G	2.039#	1.217	0.789	1.216	1.438	0.931
11.08H	2.207#	2.299##	0.929	0.910	1.007	0.475
11.08I	1.114	1.030	1.170	0.855	1.142	0.617
11.09A	2.425*	1.051	0.926	1.483	1.502	0.789
11.09B	1.412	1.043	0.778	0.859	1.702*	1.540
11.09C	1.815	1.357	1.096	1.171	1.597	1.281
11.09D	1.390	0.974	0.809	1.272	0.541	0.725
11.09E	1.106	1.086	0.941	1.431	1.028	0.858
11.10A	1.806	1.632	0.884	0.651*	4.106##	3.371#
11.10B	1.565	1.319	0.908	1.070	2.258	2.711*
11.10C	2.736*	2.667*	1.018	1.230	3.206#	2.310*
11.11A	1.928**	1.862**	1.020	1.108	5.980	2.913
11.11B	1.223	0.735	0.579**	0.771	2.062*	3.009##
11.11C	1.567	1.383	1.395	1.766**	4.374	2.542
11.11D	1.065	1.169	0.711	0.573**	1.853	3.123##
11.12A	5.792**	10.548##	1.623*	1.448	0.775	0.821
11.12B	1.361	2.330#	1.225	0.972	0.961	0.585
11.12C	3.089	0.842	0.961	0.851	1.858#	2.009##
11.12D	2.653*	3.672#	0.755	0.797	1.298	1.229
11.12E	2.415##	2.784##	0.619	0.333**	1.043	0.991
11.12F	3.878*	2.150	0.778	0.675	1.823**	1.617*
11.12G	3.535	1.372	0.844	1.270	2.092##	1.543*
11.12H	1.756	2.629*	0.888	0.591*	2.832**	3.554#
12.01	0.565	2.901	1.090	0.590	0.864	1.969#
12.02A	1.456	1.671*	1.006	0.702	1.659	0.554
12.02B	7.184	8.250	2.137	2.282*	1.102	1.233
12.02C	3.189#	4.127##	1.202	0.887	0.646	0.618*
12.02D	1.618	1.067	1.990*	1.300	0.995	1.604*
12.02E	0.429##	1.362	1.754*	0.814	10.287##	3.098
12.05A	1.553*	1.861#	0.963	0.540*	8.003	2.044
12.05B	0.636	1.000	1.047	0.635	1.379	1.985##
12.05C	2.137##	2.557##	0.779	1.095	3.307	1.917

Notes:

- Control variables: gender, age, and higher education.
- Significance levels for extremes: $^{*}\,p < 0.05$, $^{**}\,p < 0.01$, $^{\#}\,p < \alpha_{yek} \approx 0.003$, $^{\#\#}\,p < \alpha_{bnf} \approx 0.0006$
- Significance levels for midpoints: $^{*}\,p < 0.05$, $^{**}\,p < 0.01$, $^{\#\#}\,p < \alpha_{bnf} \approx 0.0007$

Appendix C.4 Proportion of extreme responses

Shaded cells present likely direction of impression management answer where it was possible to identify it (see page 165). Bolded values mark the mode with the most frequent selection of a specific scale value.

	Lower extreme			Midpoint			Upper extreme		
Item	Web	CATI	CAPI	Web	CATI	CAPI	Web	CATI	CAPI
7.02	26.6	+8.9*	+12.2#	23.4	-5.9	-2.5	0.0	0.0	0.0
7.05A	30.7	+7.3	+11.5	11.2	-1.2	-3.8	3.3	-0.8	+2.1
7.05B	0.0	+0.5	+1.0	13.6	-6.6*	-6.3	24.4	+9.1	+17.1##
7.05C	1.4	+0.6	+0.1	16.0	-3.0	-8.2*	26.8	+6.2	+13.5#
7.05D	5.2	+2.3	+4.5	16.4	+6.6	+0.1	8.5	-2.0	+0.7
7.05E	0.0	0.0	+0.5	21.6	-4.1	-0.2	11.3	+3.7	+4.7
7.05F	0.9	+0.6	-0.4	14.1	-3.6	-9.7**	22.5	+5.5	+10.7*
7.05G	8.9	+4.1	+3.7	17.4	-3.9	-3.3	12.7	-3.2	+8.7*
7.05H	0.9	+0.1	+0.6	16.9	-7.9*	-4.8	21.6	+5.5	+14.8#
7.05I	12.7	+7.3	+11.2**	20.2	-6.7	-9.0*	4.7	-1.7	-1.3
7.05J	3.8	+2.2	-1.4	18.8	-4.8	-0.3	11.7	+3.3	+9.3*
7.05K	0.5	+1.5	-0.5	10.8	-4.8	-5.0	24.4	+3.6	+5.7
7.05L	0.0	+0.5	0.0	8.9	-1.4	-5.5*	29.6	-2.1	+1.5
7.05M	10.8	+3.7	+8.7**	18.3	-2.3	-0.3	5.6	-2.6	-1.2
7.05N	3.3	-2.3	-2.3	19.2	-5.6	-5.1	12.2	+5.4	+2.4
7.05O	0.0	+1.5	0.0	16.9	-1.9	-0.4	23.9	-0.9	+0.9
7.06A	4.7	+1.3	-1.3	23.9	-6.4	-7.4*	22.5	+21.5##	+18.8##
7.06B	2.8	+2.7	-0.8	19.2	-3.2	-7.0*	23.9	+20.6##	+22.4##
7.06C	1.9	+2.6	+4.9*	23.5	-5.0	-4.0	23.9	+4.1	+9.3*
7.06D	2.8	+0.2	-1.3	23.0	-5.0	+0.4	22.5	+8.0	+11.6**
7.06E	2.4	+2.6	+1.0	20.2	-3.2	+2.1	26.4	+5.6	+4.7
7.07	1.4	-1.4	-1.4	7.9	-4.9*	-4.0	8.4	+2.6	+3.7
7.08A	30.8	+31.7##	+17.7##	55.6	-30.6##	-17.9##	13.6	-1.1	+0.1
7.08B	3.3	+2.2	-1.8	20.0	-11.0**	-6.9*	76.7	+8.8*	+8.7*
7.08C	12.1	-4.6	0.0	22.4	-3.9	-3.5	65.4	+8.6	+3.5
7.08D	35.0	+19.5##	+11.1**	44.9	-16.9##	-9.5*	20.1	-2.6	-1.7
7.08E	5.1	-1.6	-1.7	16.8	-7.3*	-5.6	78.0	+9.0*	+7.4*
7.08F	60.1	+19.4##	+12.7#	32.4	-16.9##	-8.1*	7.5	-2.5	-4.6*
7.09A	66.4	+23.5##	+1.4	na	na	na	1.4	-0.9	-0.4
7.09B	74.8	+15.2##	+3.8	na	na	na	1.9	-1.9	-1.9
7.09C	75.2	+15.3##	+7.8*	na	na	na	2.3	-2.3	-2.3
7.09D	55.1	+30.9##	+14.8*	na	na	na	1.4	-0.9	-1.4
7.09E	59.8	+26.2##	+15.0#	na	na	na	0.5	0.0	0.0
7.09F	77.1	+18.4##	+6.4*	na	na	na	0.5	0.0	-0.5
7.09G	45.3	+28.2##	+8.4*	na	na	na	0.5	0.0	-0.5
10.02	5.6	-1.1	-0.3	na	na	na	3.3	+4.2	+2.5

(*Table continued on next page*)

	Lower extreme			Midpoint			Upper extreme		
Item	Web	CATI	CAPI	Web	CATI	CAPI	Web	CATI	CAPI
10.03A	96.3	+0.2	+0.8	na	na	na	3.7	-0.2	-0.8
10.03B	79.0	+3.5	+7.9**	na	na	na	21.0	-3.5	-7.9**
10.03C	61.2	+6.5	+9.0**	na	na	na	38.8	-6.5	-9.0**
10.03D	85.4	+4.1	+2.0	na	na	na	14.6	-4.1	-2.0
10.03E	83.6	+5.4	+5.7*	na	na	na	16.4	-5.4	-5.7*
10.03F	79.4	+9.1**	+10.9##	na	na	na	20.6	-9.1**	-10.9**
10.04A	2.8	+3.3	-0.9	na	na	na	97.2	-3.3	+0.9
10.04B	1.9	+1.1	0.0	na	na	na	98.1	-1.1	0.0
10.04C	18.2	-6.6	-9.9	na	na	na	81.8	+6.6	+9.9#
10.04D	17.0	-3.5	-11.6##	na	na	na	83.0	+3.5	+11.6##
11.03	23.7	+1.3	+9.0*	11.2	+0.8	+2.9	3.7	-0.2	+1.7
11.04A	6.5	+4.5	+2.2	22.8	-4.8	0.0	33.5	+15.0#	+4.8
11.04B	7.0	-1.0	-2.6	23.9	-11.4**	-7.4	41.3	+20.2##	+2.4
11.04C	7.9	+2.1	-1.1	24.8	-8.3*	-0.5	31.8	+13.2**	+4.6
11.07	0.9	+4.6*	+10.3##	17.8	+6.7	+4.0	4.2	-0.7	-0.3
11.08A	7.4	+4.6	+4.3	29.3	-5.3	-7.5	18.1	+13.9#	+4.7
11.08B	36.4	+13.9#	+16.8##	19.2	-0.6	-9.9**	3.3	+1.7	-0.4
11.08C	6.1	+6.4*	+1.7	30.8	-9.8*	-14.3##	24.3	+12.7**	+6.8
11.08D	35.8	+3.9	+11.8*	16.7	+4.9	-2.6	0.0	+4.0	+4.4
11.08E	0.5	+4.5*	+3.4	24.3	-8.3*	-2.5	31.3	+9.7*	+5.1
11.08F	0.5	+3.5*	+1.9	23.4	-6.9	-2.0	31.8	+9.2*	+8.0*
11.08G	19.2	+12.8#	+3.6	18.7	-3.2	+4.6	11.7	+4.3	-1.0
11.08H	18.2	+15.3#	+15.8##	23.4	-1.4	-1.1	5.1	-0.1	-2.7
11.08I	25.7	+2.3	-0.3	18.7	+2.3	-2.1	19.6	+2.4	-5.5
11.09A	3.7	+4.8*	+0.2	56.7	-1.2	+8.8	7.0	+3.5	-1.2
11.09B	3.3	+1.2	+0.1	43.7	-6.2	-4.4	18.6	+9.4*	+8.1
11.09C	3.3	+2.2	+1.1	39.1	+2.9	+4.1	16.3	+7.2	+3.1
11.09D	33.6	+7.4	-1.1	33.2	-3.7	+6.1	2.8	-1.3	-0.9
11.09E	29.1	+1.6	+0.5	31.9	-0.2	+9.4	2.8	+0.2	-0.4
11.10A	7.9	+6.1	+4.8	35.8	-2.8	-8.5*	4.2	+11.3##	+9.0#
11.10B	7.0	+4.0	+2.7	40.5	-2.0	+2.2	3.3	+3.7	+5.0*
11.10C	3.8	+6.2*	+6.4*	35.2	+0.3	+5.1	5.2	+9.3#	+5.5*
11.11A	17.3	+11.3**	+10.4**	19.6	+1.0	+2.2	0.5	+2.5	+1.0
11.11B	3.7	+0.8	-0.8	43.0	-13.0**	-6.1	8.9	+8.6*	+14.4##
11.11C	10.3	+4.3	+2.8	28.0	+8.2	+14.2**	0.9	+3.1	+1.5
11.11D	3.3	+0.2	+0.6	46.3	-8.3	-12.8**	8.4	+6.6	+13.9##
11.12A	1.4	+6.1**	+12.2##	30.7	+10.8*	+9.1	16.7	-3.2	-3.6
11.12B	12.1	+4.0	+13.1#	24.2	+3.9	-0.4	10.7	-0.1	-4.4
11.12C	0.9	+2.1	+0.1	17.2	-1.2	-1.7	37.2	+13.8#	+13.3##
11.12D	3.7	+5.8*	+8.9#	27.9	-4.9	-3.6	21.4	+4.1	+2.4
11.12E	24.7	+19.3##	+20.9##	17.2	-5.7	-10.4**	3.3	+0.2	+0.1
11.12F	1.4	+4.1*	+2.0	18.2	-3.7	-4.6	25.2	+12.3**	+7.8*
11.12G	0.9	+2.6	+0.6	13.1	-1.6	+4.4	29.9	+15.6##	+7.5
11.12H	4.7	+3.3	+7.1*	29.6	-2.1	-9.0*	5.2	+7.8**	+10.0#

(*Table continued on next page*)

	Lower extreme			Midpoint			Upper extreme		
Item	Web	CATI	CAPI	Web	CATI	CAPI	Web	CATI	CAPI
12.01	0.9	-0.4	+1.5	19.1	+0.9	-7.0	26.5	-2.5	+15.2[#]
12.02A	57.7	+9.3	+11.9[*]	12.6	-0.1	-3.8	0.9	+0.6	-0.4
12.02B	0.5	+2.5	+3.4	4.7	+4.8	+5.0[*]	53.5	+3.0	+6.2
12.02C	5.1	+9.9[#]	+12.9[##]	23.7	+3.3	-1.4	27.0	-7.5	-8.6[*]
12.02D	0.9	+0.6	+0.1	10.2	+7.3[*]	+2.9	44.7	+0.8	+11.6[*]
12.02E	42.7	-17.2[##]	+8.8	16.9	+9.1[*]	-2.8	1.4	+11.6[##]	+3.5
12.05A	35.2	+10.8[*]	+13.8[#]	16.4	-0.4	-6.2[*]	0.5	+3.0	+0.5
12.05B	0.0	+1.0	+1.5	13.1	-0.1	-4.4	37.1	+7.9	+16.8[##]
12.05C	34.3	+17.7[##]	+21.5[##]	13.6	-2.6	+1.4	0.9	+2.1	+1.0

Appendix C.5 Logistic regression models of between-mode differences in complete non-differentiation

Scale (model statistics)	CATI		CAPI	
	OR	$\|z\|$	OR	$\|z\|$
(7.05) Personality	*No complete non-differentiation, model not estimated.*			
(7.06) Sense of control ($\chi^2_5 = 2.61, n = 608$)	1.016	0.06	1.351	1.16
(7.08) Loneliness ($\chi^2_5 = 3.70, n = 614$)	0.638	0.49	0.298	1.04
(7.09) Depression ($\chi^2_5 = 59.08^{\#\#}, n = 613$)	4.722	$7.16^{\#\#}$	2.052	$3.34^{\#\#}$
(10.03) Affordable goods and services ($\chi^2_5 = 52.22^{\#\#}, n = 612$)	1.309	1.29	1.743	2.63^{**}
(10.04) Payment inability ($\chi^2_5 = 21.15^{\#\#}, n = 604$)	1.667	2.09^{*}	2.754	$3.78^{\#\#}$
(11.04) Importance of religious ceremonies ($\chi^2_5 = 19.10^{\#\#}, n = 616$)	0.596	2.57^{**}	0.528	$3.17^{\#\#}$
(11.08) Marriage and children	*No complete non-differentiation in CATI and CAPI, model not estimated.*			
(11.09) Family risks responsibilities ($\chi^2_5 = 11.21^{*}, n = 615$)	0.813	0.63	0.995	0.02
(11.10) Childcare responsibilities ($\chi^2_5 = 9.66, n = 615$)	0.598	2.35^{*}	1.037	0.18
(11.11) Elderly-care responsibilities ($\chi^2_5 = 7.36, n = 615$)	0.347	2.34^{*}	0.806	0.61
(11.12) Gender roles ($\chi^2_4 = 4.48, n = 414$)	*No complete non-differentiation in CATI.*		0.368	1.18
(12.02) Survey feedback	*No complete non-differentiation in web and CATI, model not estimated.*			
(12.05) General opinion about surveys ($\chi^2_5 = 2.70, n = 616$)	1.297	0.42	1.053	0.08

Notes:

- Control variables: gender, age, and higher education.
- $^{*}p < 0.05$, $^{**}p < 0.01$, $^{\#}p < \alpha_{yek} \approx 0.005$, $^{\#\#}p < \alpha_{bnf} \approx 0.004$

Appendix C.6 OLS regression models of between-mode differences in the level of differentiation

Scale (model statistics)	b_0	CATI		CAPI	
		b	$\|t\|$	b	$\|t\|$
(7.05) Personality ($F = 4.18^{\#\#}$, $R^2_{adj} = 0.026$, $n = 608$)	0.875	-0.023	2.37*	-0.002	0.26
(7.06) Sense of control ($F = 0.68$, $R^2_{adj} = 0.000$, $n = 608$)	0.512	0.021	0.73	-0.023	0.79
(7.08) Loneliness ($F = 0.97$, $R^2_{adj} = 0.000$, $n = 614$)	0.817	-0.001	0.09	0.010	0.62
(7.09) Depression ($F = 14.21^{\#\#}$, $R^2_{adj} = 0.097$, $n = 613$)	0.464	-0.222	7.89$^{\#\#}$	-0.084	3.00$^{\#\#}$
(10.03) Affordable goods and services ($F = 12.69^{\#\#}$, $R^2_{adj} = 0.087$, $n = 612$)	0.275	-0.050	1.36	-0.096	2.62**
(10.04) Payment inability ($F = 4.09^{\#\#}$, $R^2_{adj} = 0.025$, $n = 607$)	0.244	-0.069	2.16*	-0.124	3.87$^{\#\#}$
(11.04) Importance of religious ceremonies ($F = 4.28^{\#\#}$, $R^2_{adj} = 0.026$, $n = 616$)	0.362	0.084	2.34*	0.119	3.31$^{\#\#}$
(11.08) Marriage and children ($F = 1.47$, $R^2_{adj} = 0.004$, $n = 613$)	0.849	-0.011	0.93	-0.003	0.23
(11.09) Family risks responsibilities ($F = 3.81^{\#\#}$, $R^2_{adj} = 0.022$, $n = 615$)	0.720	0.036	1.39	-0.008	0.32
(11.10) Childcare responsibilities ($F = 3.23^{\#}$, $R^2_{adj} = 0.018$, $n = 615$)	0.525	0.119	3.34$^{\#\#}$	0.016	0.44
(11.11) Elderly-care responsibilities ($F = 1.33$, $R^2_{adj} = 0.003$, $n = 615$)	0.641	0.050	2.17*	0.009	0.37
(11.12) Gender roles ($F = 2.92^{*}$, $R^2_{adj} = 0.015$, $n = 613$)	0.719	0.057	3.36$^{\#\#}$	0.047	2.78$^{\#}$
(12.02) Survey feedback ($F = 0.90$, $R^2_{adj} = 0.000$, $n = 614$)	0.768	0.018	1.38	0.004	0.27
(12.05) General opinion about surveys ($F = 0.62$, $R^2_{adj} = 0.000$, $n = 616$)	0.820	-0.016	0.77	-0.024	1.20

Notes:

- Control variables: gender, age, and higher education.
- $^{*}p < 0.05$, $^{**}p < 0.01$, $^{\#}p < \alpha_{yek} \approx 0.008$, $^{\#\#}p < \alpha_{bnf} \approx 0.004$

Razširjeni povzetek v slovenskem jeziku

UVOD

Spletne ankete postajajo eden najbolj razširjenih načinov anketiranja, predvsem zaradi hitrosti zbiranja podatkov, naprednih funkcionalnosti anketnih vprašalnikov ter nižjih stroškov zbiranja podatkov. Vse večje zahteve po nižanju stroškov raziskovalnega dela v vladnem, javnem in zasebnem sektorju spodbujajo tudi kombinirane načine anketiranja, v katerih je spletno zbiranje podatkov uporabljeno skupaj s tradicionalnimi načini, običajno telefonskim, osebnim ali poštnim anketiranjem.

Odločitev za določen način anketiranja ima pomembne implikacije za kakovost anketnih podatkov. Ta poleg točnosti ocen vključuje tudi druge kriterije, kot so relevantnost konceptov, pravočasnost in točnost objave rezultatov, dostopnost in jasnost informacij, primerljivost, koherentnost ter celostnost (Biemer in Lyberg 2003). V disertaciji smo obravnavali vlogo načina anketiranja pri nastanku napak v ocenah opazovanih parametrov, za katerega se je v anketni metodologiji uveljavil izraz »učinek načina anketiranja« (Aquilino in Lo Sciuto 1990; Dillman in Tarnai 1991).

Namen in cilji disertacije

Namen disertacije je bil izdelati celostno obrazložitev učinkov načina anketiranja v spletnih anketah med posamezniki. Bolj kot na empirično obravnavo problema smo se osredotočili na njegove konceptualne vidike in podrobno elaboracijo dosedanjih raziskav. To pripomore k boljšemu razumevanju mehanizmov anketnih napak, ki nastanejo zaradi uporabe spletnega načina zbiranja podatkov. Disertacija je osnovana na naslednji osrednji tezi:

Specifične značilnosti spletnega načina anketiranja vplivajo na anketirančev proces odgovarjanja na anketna vprašanja in lahko povzročijo nastanek učinkov načina

anketiranja. Pojav, obliko in obseg teh učinkov pogojuje vrsta dejavnikov, povezanih s specifično implementacijo določene ankete.

Za uresničitev postavljenega namena smo sledili štirim glavnim ciljem:

1. Vzpostavitev utemeljenega konceptualnega okvira za razumevanje učinkov načina anketiranja v spletnih anketah.
2. Izdelava celostnega in integriranega pregleda oblik in virov učinkov načina anketiranja v spletnih anketah.
3. Empirični prikaz kompleksnih odnosov med različnimi dejavniki, ki prispevajo k nastanku teh učinkov.
4. Izpostavitev implikacij za nadaljnje raziskovanje učinkov načina anketiranja in pomembnost problema za anketno prakso.

Disertacija z doseganjem zgornjih ciljev odpira veliko novih vprašanj, vendar hkrati ponuja trdno teoretično osnovo za razumevanje vzrokov učinkov načina anketiranja v spletnih anketah. S tem oblikuje tudi izhodišča za nadaljnje metodološke raziskave, ki bodo omogočala učinkovitejše reševanje problema v prihodnosti.

OPREDELITEV POJMOV

Čeprav sta pojma »način anketiranja« in »učinek načina anketiranja« pomemben del metodološke terminologije, je njuna pomenska raba v literaturi pogosto nekonsistentna in nejasna. Razpravo učinkov načina anketiranja v spletnih anketah smo zato pričeli s pregledom in opredelitvijo pojmov, ki so ključni za umestitev proučevanega problema v kontekst anketnih napak. Naš cilj ni bil predlagati univerzalne definicije teh pojmov, temveč vzpostaviti terminološki okvir za teoretično in empirično obravnavo razlik med načini anketiranja.

Način anketiranja

Izraz »način anketiranja« se običajno intuitivno uporablja za opis pristopa k zbiranju podatkov (na primer poštna, telefonska, osebna ali spletna anketa). Objavljenih je razmeroma malo poskusov sistematičnega iskanja odgovora na vprašanje o značilnostih,

ki opredeljujejo in razlikujejo načine. To je postalo zlasti problematično z razmahom številnih novih načinov anketiranja v zadnjih dveh desetletjih, ki ga je v spodbudil predvsem razvoj informacijsko-komunikacijskih tehnologij (Couper 2011).

Za konceptualizacijo načinov anketiranja je najprej smiselno umestiti način zbiranja podatkov v širši kontekst anketnega procesa, ki ga sestavlja več faz priprave in implementacije ankete (Groves in drugi 2009). Biemer in Lyberg (2003) skupek teh operativnih faz implementacije ankete imenujeta *sistem zbiranja podatkov*. Način anketiranja predstavlja le en del tega sistema, zato ga obravnavamo ločeno od načina vzorčenja, vabljenja anketirancev in drugih faz anketnega procesa.

Avtorji se pri obravnavi načinov anketiranja opirajo na različne značilnosti postopkov za zbiranje podatkov. Na osnovi pregleda relevantne literature (de Leeuw 1992; 2005; Tourangeau in drugi 2000; Biemer in Lyberg 2003; Groves in drugi 2009; Couper 2011) smo identificirali šest inherentnih značilnosti načina anketiranja:

- **Medij za prenos informacij** je orodje ali storitev za prenos vprašanj in odgovorov (podatkov) med anketirancem in raziskovalcem. Vzpostavi se lahko kot osebna komunikacija, po telefonu, pošti oz. drugem načinu fizičnega prenosa vprašalnika, po e-pošti ali spletu.
- **Glavni kanal za predstavitev vprašanj (vhodni kanal)** določa način anketirančevega sprejemanja informacij; lahko je slušni, vidni ali kombinacija obeh.
- **Kanal za podajanje odgovora (izhodni kanal)** služi anketirancu za odgovarjanje na vprašanja. To lahko poteka ustno ali pisno. Zaradi posebnosti nadalje ločimo tudi elektronski izhodni kanal, pri katerem anketiranec odgovor poda z uporabo miške, tipkovnice ali druge elektronske naprave.
- **Vključenost anketarja med zbiranjem podatkov** se nanaša na prisotnost in vlogo anketarja med anketiranjem. Anketar je lahko izvajalec ankete ali zgolj prisostvuje anketirančevemu samostojnemu izpolnjevanju vprašalnika. Pri samoanketiranju anketar ni prisoten.

- **Bližina interakcije med anketarjem in anketirancem** je povezana z vključenostjo anketarja in opisuje vrsto interakcije med obema akterjema. Ta interakcija je lahko osebna ali oddaljena (npr. pri telefonskih anketah). Posebna oblika interakcije je virtualni anketar; pri tem dejanska (živa) oseba ni vključena v proces zbiranja podatkov.
- **Uporaba računalniške tehnologije za zbiranje podatkov** se nanaša na anketirančevo ali anketarjevo uporabo kakršnekoli vrste računalniške tehnologije med anketiranjem.

Ključna lastnost inherentnih značilnosti načina anketiranja je njihova nespremenljivost, ne glede na različne implementacije ankete in kontekste, v katerih anketiranje poteka. Druge značilnosti načina anketiranja so delno omejene z inherentnimi značilnostmi, vendar lahko variirajo glede na specifičen način izvedbe ankete, obnašanje in lastnosti vključenih akterjev (anketarjev in anketirancev) ter različne dejavnike družbenega in posameznikovega konteksta anketiranja. Primeri teh izvedbeno specifičnih in kontekstualnih značilnosti so mesto nadzora nad potekom anketiranja, ki je lahko na strani anketiranca ali anketarja, okolje anketiranja, stopnja anketirančeve obremenjenosti zaradi sodelovanja v anketi, stopnja zaznane zasebnosti, hitrost poteka anketiranja in drugi.

Na tej osnovi smo opredelili način anketiranja kot skupek postopkov za zbiranje podatkov, ki določajo osnovna načela komunikacije in prenosa informacij med anketirancem in anketnim vprašalnikom. Ta načela so utemeljena z inherentnimi značilnostmi, ki diferencirajo posamezne načine.

Spletne ankete so opredeljene z naslednjimi inherentnimi značilnostmi: splet kot medij za prenos informacij, vizualni vhodni in elektronski izhodni kanal, odsotnost anketarja in vsake oblike interakcije z anketarjem (samoanketiranje) ter anketiranec kot uporabnik računalniške tehnologije za zbiranje podatkov.

Učinki načina anketiranja

V literaturi se pojavljata dve različni razumevanji učinkov načina anketiranja. V širšem smislu se učinki načina anketiranja nanašajo na vplive celotnega sistema zbiranja podatkov na dobljene ocene. To vključuje napake zaradi načina anketiranja in tudi napake, nastale zaradi specifičnega vzorčenja, rekrutiranja anketirancev, obdelave podatkov in drugih faz anketnega procesa. V disertaciji obravnavamo učinke načina anketiranja v ožjem smislu, in sicer vpliv samega načina zbiranja podatkov.

Skladno z našo razpravo o značilnostih načinov anketiranja razumemo učinek načina kot vse neposredne in posredne vplive inherentnih značilnosti načina anketiranja na dobljene anketne ocene. Pri razlagi problema zato ne upoštevamo le inherentnih značilnosti načina anketiranja, temveč tudi njihove kompleksne odnose z izvedbeno specifičnimi in kontekstualnimi značilnostmi ter drugimi dejavniki.

Za razumevanje problema je pomembno tudi razlikovanje med učinki načina anketiranja in razlikami v ocenah med načini. Čeprav je razlika v ocenah med dvema ali več načini anketiranja eden ključnih indikatorjev prisotnosti učinkov, to ne sme voditi do sklepa, da se problem ne pojavlja v anketah, ki uporabljajo le en način anketiranja. Obravnavani učinki so namreč posledica značilnosti določenega načina anketiranja, in ne hkratne uporabe več načinov.

V disertaciji smo se omejili na vplive načina anketiranja na napako merjenja, t. j. razkorak med poročano in dejansko vrednostjo spremenljivke (Biemer in Lyberg 2003). Čeprav lahko način anketiranja vpliva tudi na druge anketne napake (npr. napako zaradi neodgovora), je napaka merjenja najbolj tipična in najkompleksnejša posledica tega problema.

ELABORACIJA UČINKOV NAČINA ANKETIRANJA V SPLETNIH ANKETAH

Obravnava učinkov načina anketiranja v spletnih anketah in njihovega vpliva na točnost ocen predstavlja osrednji prispevek disertacije. Ker smo se osredotočili na napake merjenja, je za obravnavo problema ključno razumevanje vplivov značilnosti načina anketiranja na anketirančevo izvedbo procesov odgovarjanja na anketna vprašanja.

Proučevanje problema smo zato utemeljili na teorijah oziroma modelih, ki opisujejo relevantne kognitivne procese. Na tej osnovi smo z izčrpno evalvacijo in integracijo obstoječih empiričnih študij izdelali konceptualni model učinkov načina anketiranja v spletnih anketah.

Proces odgovarjanja na anketna vprašanja

Najbolj razdelan in najširše uporabljan model anketirančeve obdelave informacij pri odgovarjanju na anketna vprašanja je predstavil Tourangeau (1984; Tourangeau in drugi 2000). Model opisuje štiri kognitivne stopnje, ki jih mora anketiranec celostno izvesti za podajanje točnega odgovora na anketno vprašanja:

1. razumevanje vprašanja,
2. pridobivanje relevantnih informacij iz spomina,
3. presojanje pridobljenih informacij ter
4. oblikovanje in podajanje odgovora.

Pri tem smo obravnavali tudi sorodne pristope drugih avtorjev (Cannell in drugi 1981; Strack in Martin 1987; Willis in drugi 1991; Forsyth in Hubbard 1992; Schwarz in Oyserman 2001), ki kljub nekaterim specifičnim poudarkom kažejo visoko stopnjo medsebojne skladnosti.

Modeli procesa odgovarjanja izpostavljajo potencialno veliko zahtevnost odgovarjanja na anketna vprašanja. Samoanketiranje, ki je ena od značilnosti spletnih anket, od anketiranca zahteva še izvedbo dodatnih nalog – samostojno osredotočenje na posamezno vprašanje, interpretacijo vizualnih elementov vprašalnika ter upoštevanje navodil za navigacijo po vprašalniku (Jenkins in Dillman 1997; Redline in Dillman 2001).

Odkloni v procesu odgovarjanja in posledično netočni odgovori se pojavijo, kadar anketiranec potrebnih operacij ne izvede celostno. To je lahko posledica anketirančeve objektivne nezmožnosti ali nezadostne motivacije. Slednji vidik obravnava model zadovoljevanja, ki sta ga razvila Krosnick in Alwin (1987; Krosnick 1991). Avtorja trdita, da posameznik pri odgovarjanju na anketna vprašanja išče ravnovesje med potrebnim trudom in zadostno točnostjo odgovorov. O šibkem zadovoljevanju govorimo, kadar

anketiranec izvede kognitivne korake procesa odgovarjanja le do stopnje, ki po njegovem mnenju pripelje do zadosti točnega odgovora. Pri močnem zadovoljevanju pa anketiranec nekatere korake procesa odgovarjanja v celoti izpusti in poda odgovor, za katerega meni, da se bo zdel točen izvajalcu ankete. Avtorja kot dejavnike zadovoljevanja navajata zahtevnost naloge ter anketirančevo sposobnost in motivacijo.

Tovrstni odkloni lahko rezultirajo v nekaterih tipičnih merskih napakah, kot so učinki zaporedja odgovorov, težnja k strinjanju, nediferenciacija odgovorov na lestvice, nevsebinski odgovori (npr. »ne vem«), naključno izbiranje odgovorov, ne glede na vsebino, in drugi (Krosnick, 1991).

Posebna vrsta odklonov v procesu odgovarjanja lahko nastane pri občutljivih vprašanjih, na primer vprašanjih o dohodku, spolnosti, zdravju, uporabi prepovedanih drog in podobno (Bradburn in drugi 1978; Kreuter in drugi 2008). Če anketiranec ocenjuje svoj odgovor kot družbeno manj sprejemljiv, se poveča verjetnost za oblikovanje odgovora, ki je po mnenju anketiranca družbeno bolj zaželen in mu omogoča višjo stopnjo upravljanja z vtisom (Nederhof 1985; Paulhus 2002).

Kompleksnost dejavnikov učinkov načina anketiranja v spletnih anketah

Učinke načina anketiranja v spletnih anketah smo proučevali v navezavi z značilnostmi spletnih anket s procesom odgovarjanja na anketna vprašanja. Pri tem smo se opirali na številne objavljene raziskave, ki obravnavajo posamezne značilnosti spletnih anket, z njimi povezane napake merjenja ter primerjave z drugimi načini anketiranja. Tukaj povzemamo le zaključne ugotovitve elaboracije in navajamo nekatere najpomembnejše vire.

Učinki načina anketiranja v spletnih anketah so vedno posledica določene kombinacije značilnosti načina in zunanjih dejavnikov, povezanih s specifično izvedbo ankete in lastnostmi anketirancev. Inherentnih značilnosti spletnih anket tako ne smemo obravnavati kot determinirajočega dejavnika za nastanek učinkov. Način anketiranja namreč predstavlja zgolj osnovo različnih možnosti za izdelavo anketnega projekta.

Inherentne značilnosti spletnih anket ponujajo fleksibilne možnosti za uporabo tega načina anketiranja. To se neposredno izraža v veliki raznolikosti izvedbeno specifičnih in kontekstualnih značilnosti. V nadaljevanju navajamo primere takšnih raznolikosti v spletnih anketah, ki lahko prispevajo k nastanku ali omejitvi učinkov načina anketiranja:

- **Komunikacijske kanale** spletnih anket lahko sestavljajo besede, neverbalni simboli in grafični elementi.
- **Mesto nadzora** nad potekom izpolnjevanja vprašalnika je pri samoanketiranju popolnoma prepuščeno anketirancu, čeprav ga lahko delno omejimo z interaktivnimi in dinamičnimi funkcionalnostmi spletnih vprašalnikov. To anketirancu omogoča tudi nadzor nad **hitrostjo poteka anketiranja**.
- Vizualna predstavitev vprašanj in samoanketiranje omogočata anketirancu fleksibilno **zaporedje obdelave vprašanj**. Skupaj z drugimi vizualnimi elementi vprašalnika (npr. slikami) so anketirancu tako na voljo eksplicitne **kontekstualne informacije**, ki jih lahko vključi v proces odgovarjanja na anketna vprašanja.
- Sodelovanje v spletni anketi je omogočeno z uporabo različnih vrst **računalniške tehnologije**. Tehnologija, ki jo anketiranec uporablja, določa način prikaza vprašalnika in vpliva na kakovost uporabniške izkušnje pri izpolnjevanju vprašalnika.
- **Uporaba medija** za prenos informacij je pomembno povezana s posameznikovimi in družbenimi stališči. Nezaupljivost do informacij na spletu in strah pred varnostnimi grožnjami izrazito omejujeta možnosti za prepričevanje o **legitimnosti ankete.** To je nadalje oteženo zaradi neosebnosti kot posledice odsotnosti anketarja. Večjo legitimnost je mogoče doseči z ustrezno vizualno predstavitvijo ankete.
- Odsotnost anketarja zaradi samoanketiranja zagotavlja visoko stopnjo **zaznane zasebnosti**. Nekatere izvedbeno specifične in kontekstualne značilnosti pa lahko povzročijo njeno znižanje. To je lahko zlasti posledica uporabe naprednih interaktivnih vprašalnikov, ki posnemajo vlogo anketarja, nezadostnega zagotovila zasebnosti ali anketirančeve nezaupljivosti do ankete.

- **Obremenjenost anketiranca** je močno odvisna od inherentnih in drugih značilnosti načina ter lastnosti posameznega anketiranca. Poleg kognitivnih zahtev spletnega anketiranja (predvsem ustreznih bralnih sposobnosti) lahko obremenjenost anketiranca povečujejo tudi odsotnost pomoči in motiviranja anketarja, uporaba računalniške tehnologije, interaktivne funkcionalnosti vprašalnika, motnje iz okolja in številni drugi dejavniki. Ali in koliko to povečuje zahtevnost sodelovanja v anketi, je odvisno tudi od anketirančevih sposobnosti, znanja in predhodnih izkušenj.

S proučevanjem tovrstnih odnosov med značilnostmi spletnih anket smo izdelali konceptualni model njihovega vpliva na različne vrste merskih napak: učinke zaradi vpliva kontekstualnih informacij na odgovore, objektivno nezmožnost anketiranca za podajanje točnega odgovora, iskanje bližnjic v procesu odgovarjanja (zadovoljevanje ali neodgovor na postavko) ter upravljanje z vtisom. Model smo izdelali kot razširitev in prilagoditev sheme, ki so jo oblikovali Tourangeau in drugi (2000).

Dejanski pojav učinkov načina anketiranja lahko pogojujejo tudi drugi dejavniki, ki niso neposredno povezani z načinom, na primer vsebina vprašanj, zaporedje odgovorov pri posameznih vprašanjih, motivacija ter sposobnost anketiranca in drugo (Bennink in drugi 2013).

Občutljivost spletnih anket na učinke načina anketiranja

Kompleksnost dejavnikov, ki prispevajo k nastanku učinkov načina anketiranja, otežuje oceno kritičnosti problema v spletnih anketah. Velikost učinkov je močno odvisna tudi od vrste ocenjevanega parametra. Običajno so ocene parametrov na nivoju posamezne spremenljivke (npr. povprečja ali deleži) občutljivejše na učinke načina anketiranja kot ocene korelacij med več spremenljivkami (Krosnick in Alwin 1987; Jäckle in drugi 2006). Študije kažejo na primerljive ali boljše merske lastnosti spletnih anket v primerjavi z osebnim (Miller in drugi 2002) in telefonskim anketiranjem (Roster in drugi 2004; Braunsberger in drugi 2007; Chang in Krosnick 2009).

Izdelani elaborat izpostavlja samoanketiranje kot najsplošnejši vir učinkov načina anketiranja v spletnih anketah. Odsotnost anketarja lahko vodi do povečanja obremenitve anketiranca in hkrati izrazito zmanjšuje možnosti njegovega zunanjega motiviranja. S tem se pri anketirancu povečuje verjetnost za iskanje bližnjic v procesu odgovarjanja, kar vodi do manj celostne izvedbe procesa in posledično do nižanja točnosti odgovorov. To postavlja samoanketiranje v položaj posredniškega dejavnika številnih učinkov načina anketiranja.

Množica pogostih težav se pojavlja tudi zaradi neustrezne uporabe vizualnega kanala za predstavitev vprašanj in funkcionalnosti računalniško posredovanih vprašalnikov. Pretirana uporaba interaktivnosti in vizualnih elementov lahko povzroči merske napake, ki jih v drugih načinih anketiranja ni. Funkcionalnosti, ki posnemajo vlogo anketarja, lahko pripeljejo celo do učinkov, podobnih vplivom anketarja v osebnih in telefonskih anketah. Previdna in nekoliko konservativna uporaba takšnih elementov vprašalnika, ki jo predlaga na primer Couper (2008), je zato ključna za zmanjševanje potencialno negativnih vplivov na kakovost podatkov.

Samoanketiranje, računalniško posredovani vprašalniki in vizualna predstavitev vprašanj hkrati paradoksalno ponujajo nekatere najizrazitejše prednosti spletnih anket. Njihova ustrezna uporaba omogoča pridobivanje točnejših odgovorov na občutljiva vprašanja, nadzorovanje kakovosti podatkov in številne možnosti za zbiranje podatkov. Ustrezno razumevanje učinkov načina anketiranja v spletnih anketah je zato bistveno za izkoriščanje vseh potencialov tega načina zbiranja podatkov in preprečevanje specifičnih napak.

ANALIZA UČINKOV NAČINA ANKETIRANJA V PILOTSKI ANKETI GGS

V empiričnem delu disertacije smo se osredotočili na opazovanje učinkov načina anketiranja pri vprašanjih v obliki lestvic. Študija je namenjena predvsem ilustraciji nestalne narave obravnavnih učinkov, in sicer z opazovanjem njihove konsistentnosti na večjem številu spremenljivk. S tem prispeva nova spoznanja za boljše razumevanje pogosto nepojasnjenih razlik med spletnim in drugimi načini anketiranja. Z uporabo različnih statističnih metod za opazovanje razlik med načini smo izpostavili tudi

pomembnost primerjave ocen različnih parametrov za pridobivanje natančnejšega vpogleda v obravnavane učinke.

Hipoteze

Na osnovi obravnavane literature smo oblikovali štiri splošne hipoteze; prvi dve se nanašata na odsotnost, drugi dve na prisotnost negativnih vplivov spletnega načina anketiranja.

H_1: Spletni anketiranci manj težijo k upravljanju z vtisom. Anketirančeve zaznave zasebnosti v anketni situaciji pomembno določajo verjetnost upravljanja z vtisom. Samoanketiranje, kot ena od značilnosti spletnih anket, ponuja več zasebnosti kot načini s prisotnostjo anketarja. Zato lahko pričakujemo, da bodo spletni anketiranci manj zadržani pri izbiranju odgovorov, ki jih potencialno zaznavajo kot manj primerne v določenem družbenem kontekstu.

H_2: Spletni anketiranci manj verjetno izbirajo odgovore na zgornjem ali spodnjem koncu lestvice. Številne raziskave so odkrile manjšo prisotnost skrajnih odgovorov v spletnih anketah v primerjavi z drugimi načini anketiranja, predvsem telefonskimi in osebnimi anketami. Čeprav so teoretične razlage tega pojava različne in necelovite, so najpogosteje izpostavljeni dejavniki samoanketiranje, vizualna predstavitev vprašanj in manjše upravljanje z vtisom.

H_3: Spletni anketiranci pogosteje izbirajo srednje vrednosti lestvic. Izbiranje srednjih vrednosti na lestvicah je pogosta oblika zadovoljevanja. Občutljivost spletnih anket na ta problem je najpogosteje pripisana samoanketiranju, ki mesto nadzora nad potekom anketiranja skoraj popolnoma prenaša na anketiranca. Samoanketiranje hkrati močno omejuje možnosti za povečevanje zunanje motivacije anketiranca. Oboje povečuje verjetnost neoptimalne izvedbe procesov odgovarjanja, zlasti pri dolgih vprašalnikih.

H_4: Spletni anketiranci se pogosteje poslužujejo nediferenciacije odgovorov kot ene izmed strategij zadovoljevanja. Izbiranje enakih ali skoraj enakih vrednosti lestvice pri vseh postavkah je ena izmed oblik močnega zadovoljevanja. Podobno kot zgoraj tudi tu

pričakujemo večjo incidenco teh pojavov v spletni anketi kot v telefonski ali osebni anketi.

Metodologija

Za analizo učinkov načina anketiranja in preverjanje postavljenih hipotez smo uporabili eksperimentalno izvedbo ankete *Generations and Gender Survey* (GGS). Anketni vprašalnik sestavlja 340 vprašanj, pri čemer se nekatera izmed njih ponovijo večkrat (na primer za vsakega člana gospodinjstva). Gre za dolg vprašalnik, katerega izpolnjevanje v izvirnem osebnem načinu anketiranja traja približno 45 minut.

V empirično študijo smo vključili 20 vprašanj s skupno 89 postavkami. Omejili smo se na vprašanja v obliki lestvic ter vprašanja z odgovoroma »da« ali »ne«. Izključili smo tudi vprašanja, ki se nanašajo samo na nekatere anketirance. S tem zagotavljamo, da so v analize vseh postavk vključeni isti posamezniki.

Vzorec anketirancev smo pridobili s komercialnega spletnega panela podjetja *Valicon*. K sodelovanju je bilo povabljenih 743 članov panela, za katere je podjetje razpolagalo z vsemi potrebnimi kontaktnimi podatki (e-poštnim naslovom, poštnim naslovom ter telefonsko številko). Posamezniki so bili naključno razvrščeni v enega izmed treh načinov anketiranja: spletno, telefonsko (CATI – *computer-assisted telephone interviewing*) ali osebno anketo (CAPI – *computer-assisted personal interviewing*).

Stopnja odgovora je bila 87 % v spletni anketi, 61 % v telefonski in 74 % v osebni anketi. Med eksperimentalnimi skupinami so se pojavile določene razlike v sociodemografski strukturi vzorca, vendar nobena ni bila statistično značilna ($p < 0.05$). Kljub temu smo v izračune vseh statističnih modelov vključili demografske kontrolne spremenljivke.

Za analizo upravljanja z vtisom so vsako postavko ocenili trije neodvisni strokovnjaki s področja anketne metodologije. Kot potencialno občutljivih je bilo označenih 68 (76 %) postavk. Odgovore, ki tipično kažejo na upravljanje z vtisom, je bilo mogoče določiti pri 49 postavkah. Pri ostalih, predvsem mnenjskih postavkah, je pričakovana smer družbeno zaželenega odgovarjanja manj jasna in precej odvisna od posameznikovih družbenih norm in vrednot.

Nediferenciacijo smo merili z dvema kazalcema. Prvi kazalec meri t. i. popolno nediferenciacijo, ki se pojavi, kadar anketiranec pri vseh postavkah znotraj lestvice izbere enak odgovor. Drugi kazalec omogoča merjenje različnih stopenj delne nediferenciacije. Izdelali smo ga s prilagoditvijo indeksa nediferenciacije, ki so ga razvili Linville in drugi (1989).

Za analizo podatkov smo uporabili dve metodi statističnega modeliranja: regresijsko analizo po metodi najmanjših kvadratov (OLS) in model parcialno sorazmernih obetov (GO-logit). S tem smo omogočili odkrivanje več vrst razlik med načini anketiranja, ki so ključni kazalci prisotnosti učinkov načina. Rezultate smo interpretirali z upoštevanjem statističnih značilnosti, prilagojenih večkratnim preizkusom statističnih predpostavk, ter ocen velikosti učinkov.

Rezultati

Rezultati analiz le delno podpirajo postavljene hipoteze, vendar so večinoma skladni s predhodnimi raziskavami in obetavni z vidika kakovosti podatkov v spletnih anketah.

Razlike med spletnim anketiranjem in primerjanima načinoma anketiranja so bile na večini postavk razmeroma majhne. Ocene aritmetičnih sredin so bile statistično značilno različne ($p < 0.01$) pri 22 % primerjav med spletno in telefonsko anketo ter pri 27 % primerjav med spletno in osebno anketo. Povprečne velikosti učinkov, merjene z uporabo Glassovega koeficienta Δ, so bile 0.171 oziroma 0.188.

Občutno večje število postavk z značilnimi učinki smo odkrili z uporabo modelov parcialno sorazmernih obetov, ki upoštevajo porazdelitev posameznih odgovorov na vprašanje. Ta razkorak med metodami analize kaže, da učinki načina anketiranja pri nekaterih postavkah vplivajo le na izbor določenih vrednosti lestvice, in sicer brez bistvene spremembe v merah srednjih vrednosti. Na obravnavanih postavkah je to največkrat posledica manj pogostega izbiranja ekstremnih odgovorov na spletu v primerjavi z drugima načinoma anketiranja. Ugotovitve tako opozarjajo na nezadostnost primerjav mer srednjih vrednosti kot edinega ali prevladujočega kazalca prisotnosti učinkov.

Rezultati analiz izrazito potrjujejo prvo hipotezo, da *spletni anketiranci manj težijo k upravljanju z vtisom*. Statistično značilne razlike na postavkah, ki so bile ocenjene kot potencialno podvržene upravljanju z vtisom, smo ugotovili pri 37 % primerjav med spletnim in telefonskim anketiranjem ter pri 51 % primerjav med spletnim in osebnim anketiranjem. Nadaljnje analize 49 postavk, za katere je bilo mogoče predvideti smer upravljanja z vtisom, so pokazale, da so spletni anketiranci v splošnem manj podvrženi izbiranju družbeno zaželenih odgovorov.

Rezultati močno podpirajo tudi drugo hipotezo, ki predpostavlja *manjšo težnjo spletnih anketirancev k izboru skrajnih odgovorov na zgornjem ali spodnjem delu lestvice*. Podporo tej hipotezi nakazuje že splošno opazovanje razlik v porazdelitvah odgovorov z uporabo modelov parcialno sorazmernih obetov. Statistično značilno razliko v verjetnosti izborov skrajnih odgovorov med spletno in telefonsko anketo smo odkrili pri 24 % analiziranih postavk, med spletno in osebno anketo pa pri 23 % odstotkih postavk. Spletni anketiranci so izkazovali manjšo težnjo k izbiranju skrajnih odgovorov pri vseh teh postavkah, z izjemo ene. Razlike med spletnim načinom anketiranja in primerjanima načinoma so bile posebno izrazite pri postavkah, občutljivih na upravljanje z vtisom.

Manj jasno sliko kaže preverjanje hipoteze, da *spletni anketiranci pogosteje izbirajo srednje vrednosti lestvic*. Verjetnost izbora srednjega odgovora je bila statistično značilno različna pri 9 % primerjav med spletno in telefonsko anketo ter pri 10 % primerjav med spletno in osebno anketo. Z izjemo ene postavke vse statistično značilne razlike kažejo na večjo verjetnost izbora srednje vrednosti na spletu. Pri tem pa ostaja nejasno, ali so ugotovljeni učinki posledica odklonov v procesu odgovarjanja med spletnimi anketiranci ali posledica manjše verjetnosti izbora srednjih odgovorov v primerjanih načinih zaradi večje težnje k izboru skrajnih odgovorov. Opazovanje razlik v porazdelitvah odgovorov ponuja le malo podpore prvi razlagi.

Nejasne rezultate kaže tudi proučevanje hipoteze o *višji stopnji nediferenciacije odgovorov med spletnimi anketiranci*. Ugotovljene razlike med načini so močno odvisne od posamezne lestvice. Anketiranci v telefonski in osebni anketi so izkazovali višjo stopnjo nediferenciacije pri postavkah, ki so bile ocenjene kot občutljive na upravljanje

z vtisom. Nasprotno pa je bila stopnja nediferenciacije pri lestvicah mnenjskih in vrednotnih vprašanj višja med spletnimi anketiranci, čeprav so razlike med načini relativno majhne. Postavljene hipoteze zato ne moremo potrditi.

Rezultati tako jasno potrjujejo hipotezi, ki se nanašata na odsotnost specifičnih učinkov spletnega načina anketiranja pri vprašanjih v obliki lestvic. Odgovori spletnih anketirancev so bili manj občutljivi na upravljanje z vtisom in manj skrajni, vendar brez potrjene izrazite težnje po večjem zadovoljevanju v obliki izbiranja srednjih vrednosti lestvic ali nediferenciaciji odgovorov.

Skladnost s predhodnimi raziskavami

Izsledki raziskave so v splošnem skladni s predhodnimi raziskavami, delno tudi z vidika nekonsistentnih rezultatov.

Ugotovitve o upravljanju z vtisom potrjujejo zaključke številnih primerjalnih študij o družbeni zaželenosti odgovorov in občutljivih vprašanjih v spletnih anketah (Lozar Manfreda in Vehovar 2002b; Jäckle in drugi 2006; Kreuter in drugi 2008; Chang in drugi 2009; Tourangeau in drugi 2013). Podobno kot drugi avtorji pri tem predpostavljamo, da višja stopnja poročanja o družbeno nezaželenih odgovorih pomeni točnejšo oceno (Bradburn et al. 1978). Rezultati tako utrjujejo prednost spletnih anket pred telefonskimi in osebnimi anketami pri občutljivih vprašanjih in vprašanjih o družbeno (ne)zaželenih temah.

S predhodnimi raziskavami so skladne tudi ugotovitve o manj skrajnih odgovorih v spletni anketi v primerjavi s telefonsko (Taylor 1999; Roster in drugi 2004; Christian in drugi 2007a; Dillman in drugi 2009; de Leeuw in drugi 2010b). Za razliko od študije Heerwegha in Loosveldta (2008) smo ugotovili statistično značilne razlike tudi med spletnim in osebnim anketiranjem. Ker je bila večina lestvic v osebni anketi predstavljena vizualno z uporabo kartic, razlike med načinoma najverjetneje niso posledica kanala za predstavitev vprašanj, temveč drugih dejavnikov. Z upoštevanjem poudarjenih razlik na ustreznih postavkah ima očitno najpomembnejšo vlogo pri povečevanju ekstremnosti odgovorov upravljanje z vtisom, kar izpostavljajo tudi Ye in drugi (2011).

Izsledki proučevanja izbiranja srednjih vrednosti lestvic in nediferenciacije odgovorov sledijo nekonsistentnim ugotovitvam obstoječih študij (Fricker in drugi 2005; Heerwegh in Loosveldt 2008; Chang in Krosnick 2009). Izrazitejše razlike v nediferenciaciji med načini anketiranja so bile večinoma v prid spletnemu načinu in najverjetneje posledica večjega upravljanja z vtisom med telefonskimi in osebnimi anketiranci. Rezultati so tako skladni z zaključki Klauscha in drugih (2012), ki so odkrili le malo dokazov o povečani prisotnosti te oblike zadovoljevanja v spletnih anketah.

Omejitve empirične študije

Ena ključnih omejitev empirične študije je njena nezmožnost, da bi razločili prepletene učinke. Čeprav smo analizirali veliko postavk, je variabilnost njihovih značilnosti premajhna za osamitev nekaterih učinkov. Tako je bila, na primer, večina obravnavnih postavk ocenjena kot občutljivih na upravljanje z vtisom. Na teh postavkah so razlike med načini najverjetneje hkratna posledica vplivov upravljanja z vtisom, izbiranja skrajnih odgovorov in nediferenciacije odgovorov.

Druga pomembna omejitev se nanaša na pomanjkanje eksperimentalnih manipulacij znotraj posameznega načina, ki bi omogočile neposrednejše pripisovanje odkritih razlik učinkom načina anketiranja. Z ustreznimi eksperimenti bi lahko ugotavljali tudi druge oblike zadovoljevanja v primerjanih načinih, predvsem težnjo k strinjanju ter učinke primarnosti in nedavnosti. Izvedene primerjave kljub temu močno nakazujejo na relativno nizko prisotnost zadovoljevanja v spletnih anketah.

V analizi smo se osredotočili na oblikovanje celostne slike prisotnosti ali odsotnosti izbranih odklonov v procesu odgovarjanja v spletnih anketah. Analiza velikega števila spremenljivk je omogočila vpogled v razširjenost teh učinkov, vendar s tem omejila možnosti njihovega podrobnega proučevanja na nivoju posameznih postavk in lestvic. V nadaljnje raziskovanje je zato smiselno vključiti predvsem kazalce veljavnosti in zanesljivosti merjenja. Čeprav bi bilo tehnično mogoče izvesti generično analizo vseh obravnavanih postavk, bi bila uporabna vrednost takšnega pristopa zelo vprašljiva. Bistveni element analize merske kakovosti je namreč natančno upoštevanje vsebine in latentne strukture postavk.

Majhno število odkritih negativnih učinkov spletnega načina anketiranja je lahko deloma posledica specifičnega vzorca posameznikov iz neverjetnostnega spletnega panela. Medtem ko takšni paneli omogočajo doseganje demografsko raznolike populacije anketirancev, so njihovi člani običajno vajeni izpolnjevanja (dolgih) spletnih anket in imajo relativno veliko izkušenj z uporabo računalnika in spleta.

IMPLIKACIJE UGOTOVITEV

Bistvena dodana vrednost disertacije izhaja iz sistematične evalvacije dejavnikov učinkov načina anketiranja v spletnih anketah. Celostna obravnava problema, ki v obstoječi literaturi večinoma ni predstavljena, izpostavlja kompleksne odnose med različnimi potencialnimi viri za nastanek učinkov. Konceptualni prikaz teh odnosov ponuja pomembno vodilo za pripravo in izvedbo spletnih anket ter omogoča natančnejše usmerjanje prihodnjih metodoloških raziskav. Empirične ugotovitve hkrati potrjujejo spletne ankete kot način, ki omogoča zbiranje podatkov primerljive ali celo višje kakovosti kot tradicionalni načini anketiranja.

Pomen za anketno prakso in kombinirane načine anketiranja

Upoštevanje kompleksnosti dejavnikov, ki lahko vodijo do nastanka učinkov načina anketiranja, je zlasti pomembno zaradi visoke fleksibilnosti spletnih anket. Spletni vprašalniki omogočajo uporabo številnih vizualnih in interaktivnih funkcionalnosti, katerih vključevanje je s pomočjo sodobnih programskih orodij za spletno anketiranje močno poenostavljeno. Fleksibilnost se kaže tudi na strani anketirancev, ki lahko do ankete dostopajo kadarkoli, kjerkoli in z uporabo zelo različnih naprav. Čeprav so to ključne prednosti spletnih anket, lahko njihovo neustrezno izkoriščanje vodi do resnih vplivov na kakovost zbranih podatkov. Konceptualni model omogoča uporabnikom spletnih anket tudi upoštevanje takšnih kritičnih dejavnikov med načrtovanjem in implementacijo ankete.

Z naraščajočim vključevanjem spletnih anket v kombinirane načine anketiranja, pri katerih je primerljivost rezultatov med posameznimi načini običajno bistvena, postaja razumevanje virov učinkov načina anketiranja zelo pomembno. Izvedeno proučevanje

učinkov uvaja konceptualno orodje za identifikacijo možnih virov, ki povzročajo razlike zaradi kombiniranja spletnega anketiranja z drugimi načini.

Smeri nadaljnjega raziskovanja

Elaboracija obstoječih raziskav je izpostavila vrsto odprtih vprašanj, ki zahtevajo nadaljnjo metodološko obravnavo. Kompleksnost in pogosta prikritost učinkov načina anketiranja pozivata raziskovalce k večsmernemu pristopu, ki združuje premišljeno oblikovane eksperimentalne načrte in uporabo različnih metod za analizo podatkov. Za boljše razumevanje virov učinkov je pomembno združevati ugotovitve o razlikah med načini anketiranja z ugotovitvami o vplivih različnih implementacij posameznega načina na dobljene ocene. Nove možnosti za empirično opazovanje učinkov omogoča tudi zbiranje parapodatkov (Couper 2005), ki opisujejo potek odgovarjanja na anketna vprašanja.

Zelo velik in premalo izkoriščen potencial za proučevanje problema ponujajo metaanalize. Z upoštevanjem rezultatov več raziskav je mogoče izdelati bolj posplošljive modele za pojasnjevanje okoliščin, ki vodijo do nastanka učinkov načina anketiranja. Predpogoj za uspešno uporabo metaanaliz pa je ustrezna dokumentiranost izvedenih raziskav in odprt dostop do potrebnih podatkov. Inovativni metodološki pristopi, podpiranje načel odprte znanosti ter strateško usmerjena sodelovanja med raziskovalci so zato ključnega pomena za razvoj novih znanj, ki bodo omogočila zagotavljanje najvišje kakovosti anketnih podatkov.