Question-Answer Sequences in Survey-Interviews[*]

WIL DIJKSTRA[*] and YFKE ONGENA
Department of Social Research Methodology, Free University, Amsterdam, The Netherlands

Abstract. Interaction analysis was used to analyze a total of 14,265 question–answer sequences of (Q-A Sequences) 80 questions that originated from two face-to-face and three telephone surveys. The analysis was directed towards the causes and effects of particular interactional problems. Our results showed that problematic respondent behavior is affected by the questionnaire design, whereas inadequate interviewer behavior is affected by respondent behavior, rather than directly by the questionnaire design. Two surveys used questions for which validating information was available. It appeared that the occurrence of such irregularities of interviewer and respondent behavior was related to the validity of the eventual responses. Explanations for the occurrence of problematic respondent behavior were proposed, concerning both cognitive and conversational factors, related to the wording of questions and response alternatives.

Key words: behavior coding, interaction analysis, question–answer sequence, question wording, response error, survey-interview

1. Introduction

A number of different methods have been developed to evaluate the quality of data gathered by means of standardized face-to-face or telephone interviews (e.g. Willis, 2005). For example, behavior coding can be used to investigate whether interviewers pose the question as worded in the questionnaire. Think aloud procedures may give insight into the cognitive processes involved in answering survey questions. The survey interview proper may be followed by an intensive interview, about how respondents interpreted the questions (e.g. to discover misconceptions). The eventual responses can be analyzed in order to obtain information about the

number of "Don't knows", unlikely frequency distributions, or possible inconsistencies (e.g., 18 years old and having three children).

Analysis of the interaction between interviewer and respondent is a method that yields information resembling the information from each of the above mentioned methods at once. Because interaction analysis means that all utterances of interviewer and respondent are coded, preserving the order of the interaction, it yields all information that can be obtained by behavior coding. Requests for clarification, or providing considerations, may yield information about cognitive processes and misconceptions. The eventual answer of the respondent can be compared with the score entered by the interviewer in case of seemingly inconsistent answers. Moreover, interaction analysis enables us to study causes of bad interviewer performance in more detail. For example, what kind of respondent behavior tends to precede suggestive probing?

A relevant distinction regarding the analysis of the interaction between interviewer and respondent, is between paradigmatic and non-paradigmatic question–answer sequences. A paradigmatic (Schaeffer and Maynard, 1996), or straightforward (Sykes and Collins, 1992) sequence consists of a question by the interviewer worded as scripted, an answer of the respondent that is directly scorable, and, possibly, some neutral utterance by the interviewer that the answer of the respondent is accepted, e.g. by repeating the answer or saying "OK". A sequence that deviates in one way or another from the paradigmatic sequence is called a non-paradigmatic sequence.

A further distinction is between 'problematic' and 'non-problematic' deviations. We define problematic respondent deviations as utterances by the respondent that require action by the interviewer, in order to explain the respondent's task. For example, a request for clarification, giving an inadequate answer, or even no answer at all, requires appropriate action by the interviewer. On the other hand, if the respondent 'thinks aloud', giving considerations in order to come upon an eventual answer, this is a deviation from the paradigmatic sequence, but not a problematic one.

Problematic interviewer deviations concern utterances by the interviewer that are expected to affect the validity of eventual response negatively. This, of course, is related to the controversial issue of the desirable degree of standardization; see for example Fowler and Mangione (1990); Schober and Conrad (1997). The degree of standardization can be put on a continuum (e.g. Viterna and Maynard, 2002). Our position here, as will appear from the more detailed description of problematic behaviors in Section 2.3, is adequately worded by Viterna and Maynard (2002: 368): "In the ideal type of survey center [...] interviewers are expected to follow proper standardization techniques, but are given permission [...] to relax the rules if they feel these techniques may lead to a wrong answer or to the

respondent's refusal to complete the interview". Interviewer behaviors like significant deviations from the scripted questions, or suggestive probes are definitely viewed as problematic behavior however.

Problematic deviations may also concern *absence* of behavior that is required, such as failure to provide a response, or failure to clarify a question, if the respondent asks for such clarification.

All other deviations are considered non-problematic. For example, the respondent may repeat the question, or may give considerations to explain the response. Or the interviewer may utter something like, "You're doing a good job", summarizes the considerations provided by the respondent, or provides clarifications.[1] Such deviations cannot be viewed as problematic in general, although particular instances of such deviations may very well be problematic. However, particular instances of such deviations may as well have a beneficial effect on the eventual quality of the data.

In this paper we will address the following research questions:
- What are the causes of problematic deviations?
- How do problematic deviations affect the validity of the eventual response?

2. The Data

2.1. CODING PROCEDURES

Our data came from a number of different surveys. All interviews were taped, digitized and stored as sound files on CD-ROMs. The interviews were transcribed and the transcriptions were divided into question–answer sequences (Q-A sequences,). A Q-A sequence includes all utterances belonging to a particular question from the questionnaire. Next, all separate utterances of the Q-A sequences were coded.

A number of different approaches exists to code Q-A sequences, generally called 'behavior coding'; for example, coding only particular utterances (e.g. the question as uttered by the interviewer), or coding the whole sequence (e.g. paradigmatic or not). A (nearly) full account of all studies that used behavior coding in one form or another and an evaluation of the different approaches can be found in Ongena and Dykstra (forthcoming).

The approach we followed is often referred to as interaction coding: successive utterances from a Q-A sequence are coded, preserving the order of the utterances. To this end each Q-A sequence is divided into meaningful utterances of interviewer and respondent and each utterance is coded. For our purposes, the coding scheme should be informative with respect to the course of the interaction and with respect to deviations from the paradigmatic sequence. Our coding scheme consists of a number of coding variables, each describing a particular aspect of the utterance.

The first two coding variables, called ACTOR and EXCHANGE provide information about the flow of the interaction: the 'actor' designates who is speaking, whereas 'exchange' distinguishes between questions, answers, requests for clarification, etcetera.

Two other coding variables provide information about deviations. One of them, DISTANCE, describes the 'distance' between an utterance and the question from a questionnaire. Posing a question from the questionnaire, or a direct answer to such a question, have a value of '0' on DISTANCE; providing considerations to an answer has a value of '1'. Further elucidations have value '2', whereas utterances that are completely irrelevant have value '3'.

The coding variable ADEQUACY describes deviations by making use of a number of more general concepts. First, *mismatches* concern the (mis)correspondence between an utterance and some reference. If the posed question deviates from the scripted one it is called a 'mismatch question'. If an answer does not equal one of the scripted response alternatives, it is a 'mismatch answer'. Still another example of a mismatch is when the interviewer repeats an answer of the respondent inadequately. Secondly, *validity* concerns the meaning of the utterance. An invalid question is one that deviates so much from the scripted question that the meaning is significantly altered (of course an invalid question is also a mismatch question; in this case the code for 'invalid' applies). If the respondent apparently misunderstood the question, the answer is invalid. The final concept, *suggestiveness*, only concerns questions. A question is suggestive, if the suggestion is not warranted by previous answers of the respondent.

The combination of codes on each code variable provides us with a complete description of the utterance. For example, a code like 'RA0M' means that the respondent (R) gives an answer (A) that is a direct answer to a question from the questionnaire (0), but a mismatch answer (M). On the basis of the codes assigned to the utterances, a number of problematic respondent deviations, problematic interviewer deviations, and unproblematic deviations can be discerned. An overview of the deviations used in this study can be found in Table I. These deviations are described in more detail in Appendix A. A full account of the coding scheme and information about the reliability of the codes can be found in Dijkstra (1999).

To transcribe, code and analyze the coded sequences, we used the computer program Sequence Viewer. A description of this program can be found in Dijkstra (2002).

2.2. THE SURVEYS

A total of 5 different surveys were transcribed and coded, amounting to 14,265 Q-A sequences. The surveys not only differ with respect to their

Table I. Types of deviations

Type of deviation	Description
Problematic R deviations	
Mismatch answer	Answer to question that does not meet required format
Invalid answer	Respondent answer to a question that is misunderstood
Request for clarification	Request to clarify question
Problematic I deviations	
Invalid question	Interviewer significantly changes meaning of question
Mismatch alternatives	Interviewer changes response alternatives
Suggestive probing	Interviewer suggests an answer to the respondent
Inference	Interviewer infers answer from inadequate responses
Choosing	Interviewer fills in score, based on inadequate responses
Other deviations	
Request for repetition	Request to repeat an utterance of the other actor
I omits alternatives	Interviewer does not read alternatives at all
DK answer	'Don't know' answer
Refusal	Respondent refuses to answer
Consideration	Explaining, motivating the response
Elucidation	Pertains to question topic, but does not clarify answer
Irrelevant talk	Irrelevant to question topic
Forward reference	Refers to next questions
Backward reference	Refers to previous questions or answers
Bystander answers	Answer given by a third person
Wrong registration by I	Interviewer fills in wrong score

topic, but also with respect to numbers of questions analyzed, population, mode of interviewing, etcetera. An overview of the surveys is given in Table II.

2.3. OCCURRENCE OF DEVIATIONS

Table III provides an overview of the percentage of occurrence of the different deviations. It should be noted that a particular sequence may contain (and often does contain) more than one deviation. In addition, the number of problematic sequences is shown, a problematic sequence being a sequence having one or more problematic deviations.

Despite the large variation in surveys (different questions, different modes, different types of respondents), there are some striking similarities. Most remarkable is that in all surveys, a mismatch answer by the respondent is the most frequently occurring problematic deviation. Choosing and

Table II. Overview of analyzed surveys

Topic	Questions	Respondents	Interviewers	Sequences	Mode
1. Satisfaction with living and social contacts (SAT)	7	384 Adults from new-built area	8 Female, 8 male, 3 days training	2670	Face-to-face
2. Partnership, work and living (PWL)	9	185 Age: 18–26	10 Female, 7 male, 3 days training	1227	Face-to-face
3. Environmental issues (ENVa)	13	146 Members environmental org	4 Female, 1 day training	1261	Telephone
4. Environmental issues (ENVb)	13	145 Members environmental org	5 Female, 1 day training	1468	Telephone
5. Television commercials (TV)	38	210, RDD sample	9 Female, 1 day training	7639	Telephone

suggestive probing by the interviewer, seem to be second 'best'. The high number of considerations, elucidations and irrelevant talk in survey 1, was caused by the questionnaire design: after the respondent answered the question from the questionnaire, he or she was asked to motivate the answer.

3. Causes of Problematic Deviations

A problematic deviation can be caused by: (1) the questionnaire design, (2) previous utterances in the interaction, (3) interviewer and respondent characteristics. Regarding interviewer and respondent characteristics, most deviations appeared to depend on both the interviewer and the respondent; e.g. some interviewers produced significantly more suggestive probing than others, and some respondents gave significantly more mismatch answers than other respondents. Because in most surveys the interviewers were quite homogeneous with respect to a number of characteristics (like gender or age), we were not able to find systematic effects of such background variables on the occurrence of problematic deviations. Regarding respondent characteristics we did find some significant relations between background variables and the occurrence of problematic deviations (e.g. female and older respondents produced slightly more mismatch answers

Table III. Percentage of sequences with a particular deviation (I = interviewer, R = respondent)

Type of deviation	1 (SAT)	2 (PWL)	3 (ENVa)	4 (ENVb)	5 (TV)
Problematic R deviations					
Mismatch answer	19.8	22.2	15.0	15.8	33.5
Invalid answer	2.2	2.3	1.4	1.4	2.3
Request for clarification	5.1	3.9	7.2	4.0	3.5
Problematic I deviations					
Invalid question/clarification	0.9	4.2	0.7	1.2	1.1
Mismatch alternatives	2.2	4.6	–	–	0.7
Suggestive probing	7.6	7.4	2.3	2.5	13.1
Inference	2.1	3.0	1.0	1.5	3.7
Choosing	16.6	4.0	6.4	8.7	14.2
Other deviations					
Request for repetition by R	1.5	2.3	1.6	3.3	3.7
I omits alternatives [a,b]	14.7	14.8	–	–	25.2
R gives DK answer	2.2	0.9	7.4	8.9	3.4
R refuses	0.1	0.0	0.0	0.0	0.0
R gives considerations	96.2	10.2	20.5	23.8	20.4
R elucidates on topic	63.0	1.7	1.2	2.6	3.7
Irrelevant talk	14.8	0.2	0.0	0.7	0.7
Forward reference	2.0	1.2	0.2	0.4	0.3
Backward reference	4.6	0.5	0.3	0.8	0.2
Bystander answers [c]	1.1	0.2	0.5	0.1	0.0
Wrong registration by I	3.9	1.6	3.0	3.6	3.1
Q-A Sequence					
Problematic	1011	423	286	322	3274
no problem	1659	804	974	1144	4365
Total number of sequences	2670	1227	1260	1466	7639
% problematic sequences	37.9	34.5	22.7	22.0	42.9

[a] Only if the interviewer is required to read alternatives; e.g. 'yes/no' questions are excluded.

[b] In survey 2, some questions had alternatives written on show cards. These show cards were bundled in a booklet that was handed to the respondent. The interviewer was required to indicate which card to use, e.g. "Please use card 11". If such a statement was missing from the interaction, *and* the interviewer did not read (part of) the alternatives aloud, the sequence was coded as 'omits alternatives'. In survey 5, there were several sets of items with the same alternatives; only the first question of such a set is taken into account.

[c] The presence of the bystander in the telephone surveys is undoubtedly underestimated. In surveys 3 and 4, it sometimes happened that the respondent said something like: "Wait a moment, I'll ask my wife". Only utterances from the bystander that appeared on tape could be transcribed and coded.

than male and younger respondents). However, these differences were quite small and may be partly due to capitalization of chance. In the next sections we will only discuss the questionnaire design and previous utterances as causes of problematic deviations.

3.1. THE QUESTIONNAIRE DESIGN

To determine the effect of the question or questionnaire design on the occurrence of problematic deviations, we have to look at the first problematic deviation in a sequence as these cannot be caused by previous utterances in the same sequence. Table IV shows the frequencies and percentages of the first problematic deviation in a sequence. Percentages are calculated as the percentages of the total number of problematic sequences in a survey.

In all surveys the first problematic deviation is usually produced by the respondent (see also Ongena, 2005). This discrepancy between interviewer and respondent is quite remarkable: after all, the interviewer has more opportunity to produce the first problematic deviation than the respondent, because a Q-A sequence invariably starts with an interviewer utterance. Moreover, we already saw in Table III that especially mismatch answers occur quite frequently. This seems to be even more true if we look at the first problematic deviation.

A striking difference between the surveys is the high percentage of choosing behavior in survey 1. Choosing behavior by the interviewer frequently occurs on the basis of mismatch answers, in which case choosing cannot be the first problematic deviation. In survey 1 however there was a clear design factor, that affected the occurrence of choosing behavior.

Table IV. The first problematic deviation

Problematic deviations	1 (SAT) Freq	%	2 (PWL) Freq	%	3 (ENVa) Freq	%	4 (ENVb) Freq	%	5 (TV) Freq	%
Mismatch answer	458	45	238	56	170	59	213	66	2359	72
Invalid answer	37	4	7	2	13	5	12	4	110	3
Request for clarification	114	11	41	10	82	29	54	17	226	7
Invalid question	15	1	47	11	3	1	10	3	42	1
Mismatch alternatives	17	2	42	10	–	–	–	–	20	1
Suggestive probing	116	11	31	7	12	4	17	5	353	11
Inference	30	3	14	3	4	1	9	3	53	2
Choosing	221	22	3	1	2	1	7	2	111	3
Total	1011	100	423	100	286	100	322	100	3274	100

In this survey, respondents were always asked to motivate their answer. Quite often, on the *next* questions respondents started with giving considerations, 'forgetting' to answer the scripted question. Interviewers often based the eventual score on these considerations, without even suggesting a response alternative. In a sense, this choosing behavior is equally due to both respondent (who should have answered the question adequately) and interviewer (who should have probed to obtain an adequate answer).

Another questionnaire design factor that may affect the occurrence of problematic deviations is the use of show cards (see also Prüfer and Rexroth, 1985). Only in survey 2 show cards were used. Of the nine questions analyzed, four questions were accompanied by a show card. As Table V shows, these four questions yielded the least problematic sequences.

Most kinds of problematic deviations occur less often if show cards are used, the remarkable exception being requests for clarification. If there are problems in deciding between two alternatives, a way out of this problem is giving a mismatch answer. However, if show cards are used, it is more clear to the respondent that a mismatch answer is not adequate. Instead of giving a mismatch answer in case of problems, the respondent may ask the interviewer to assist him or her in the answering process, by a request for clarification.

Of the five questions without show cards in survey 2, four questions suffered from serious flaws: two questions were posed as 'yes/no' questions, but nevertheless were accompanied by a set of five response alternatives. Such questions tend to evoke premature responses from the respondent (i.e. "yes" or "no"), after the question has been posed, but before the alternatives are read. Two questions (including one of these 'yes/no' questions)

Table V. The use of show cards and problematic deviations

Problematic deviations	Show card		No show card	
	Freq	%	Freq	%
Mismatch answer	53	10	185	26
Invalid answer	5	1	2	0
Request for clarification	27	5	14	2
Invalid question/clarification	3	1	44	6
Mismatch alternatives	1	0	41	6
Suggestive probing	6	1	25	4
Inference	7	1	7	1
Choosing	0	0	3	0
(No problematic deviation)	416	80	388	54
Total	518	100	709	100

contained response alternatives that were both unevenly distributed and difficult to distinguish; e.g. 'hardly ever, only a few times, now and then, often, very often'. An alternative like 'never' is missing, whereas the first three alternatives do not differ much; the distance with alternative four is much larger. Hence, not only lack of show cards, but also problems with the wording of the question and the response alternatives may account for the high number of problematic deviations of these questions.

From Table III it appears that surveys 3 and 4 have the lowest number of problematic sequences. All questions in these surveys were 'yes/no' questions. Survey 5 on the other hand has the largest number of problematic sequences: only two of the 38 questions were 'yes/no' questions (yielding 'only' 11 and 12% problematic sequences). On the other hand, all 'yes/no' questions from surveys 3, 4 and 5 were factual questions, which may also account for the low number of problematic sequences. In survey 1 however, there were four 'yes/no' questions, and three questions with response alternatives, none of them factual. The percentages of problematic sequences in this survey ranged from 21 to 37 for the 'yes/no' questions and from 41 to 55 for the other questions. Although the occurrence of problematic sequences will certainly depend on the content of the question too, the difference between 'yes/no' questions and other closed questions is quite striking. It seems quite plausible to conclude that in general, 'yes/no' questions yield less problematic sequences than other types of questions (i.e. questions with response alternatives other than 'yes' and 'no').

Survey 5 contained the largest number of different questions. We already discussed that the two 'yes/no' questions yielded relatively few problematic sequences. Quite a large number of questions (17) were worded as an assertion, and the respondent was asked which alternative from a set of alternatives applied to the assertion. These questions yielded the largest amount of problematic sequences (between 48 and 65%). Further analysis showed that the reactions of the respondent to these questions were very comparable to inappropriate 'yes/no' questions that are accompanied with a set of response alternatives. For example:

  I: TV commercials are funny to look at
  R: Yes

Respondents often immediately reacted in such a way (with "yes" or "no") and did not give the interviewer the time to read the response alternatives (strongly disagree, disagree, etc.). These assertions concerned two sets of items. For both sets, the first question was *preceded* by presenting the response alternatives; hence the respondent knew the preferred response format. Nevertheless, these two questions too yielded large amounts of problematic sequences (54 and 61 percent, respectively).

Of the remaining questions in survey 5, three questions concerned factual information, readily available to the respondent (the last day the respondent watched television, the respondent′s age and the number of persons in the household). The percentage of problematic sequences was relatively low (14, 10 and 6%, respectively). There was one more factual question, about the respondent′s highest level of education, that yielded quite a lot problematic sequences (37%). Respondents mentioned names of educational institutes, or gave otherwise responses that did not fit one of the response alternatives.

All other questions required the respondent to make some kind of judgment or estimation (e.g. "How many days a week do you watch television on average?") or concerned attitudinal questions. The percentage of problematic sequences ranged from 27 to 53. The differences in percentages seemed to be mainly due to the amount of cognitive processing required to answer the question (in general, the number of requests for clarifications was somewhat higher for these questions, than for the questions discussed above: on average 5 vs. 2%). The question with 53% problematic sequences was "What percentage of the time did you watch television attentively?" and referred to the last time the respondent watched television. This question requires the respondent to distinguish between 'attentively' and 'not attentively', to estimate the time 'watched attentively', and to calculate a percentage (the total time was already estimated in the previous question). Such a question requires a lot of cognitive effort, apparently from the interviewer too:

I: And what percentage of the time did you look attentively?
R: Attentively.
I: Yes
R: Fifteen minutes
I: Fifteen minutes from one hour and a half is about a quarter
R: Yes
I: Hence 25%
R: Something about like that, yes.

In summary, the questionnaire design can definitely affect the occurrence of problematic deviations. Moreover, the questionnaire design primarily causes problematic deviations by the respondent, particularly mismatch answers. Problematic deviations by the interviewer seem to be much less affected by questionnaire design.

Characteristics of the question design that enhance the occurrence of problematic deviations, seem to be (a '+' is a positive effect and a '−' is a negative effect on the occurrence of such deviations):

- The use of show cards (−).
- The use of 'yes/no' questions, with 'yes' and 'no' as response alternatives (−).
- The use of 'yes/no' questions, with response alternatives other than 'yes' and 'no' (+).
- The use of assertions, with response alternatives other than 'yes' and 'no' (+).
- Unevenly distributed response alternatives (+).
- Asking respondents to motivate their answers in a number of successive questions (+).

3.2. PREVIOUS UTTERANCES

Problematic deviations may not only be caused by questionnaire or question design, but by previous utterances (from the other actor) too.

Tables VI and VII show the frequencies of the last respondent utterance preceding the first suggestive probe, respectively the first inference, and whether this frequency is above (a positive z-value) or below chance (a negative z-value). Adequate answers, mismatch answers and considerations have by far the highest frequencies. 'Other R utterances' concern a large number of different kinds of utterances, like requests for repetition, elucidations, irrelevant talk, repeating (part of) the question, and so on. All these utterances occur very infrequently, whereas the z-values are usually insignificant or negative (an exception is 'don't know' responses in survey 1, which occurs 5 times before a suggestive probe, but has a z-value of 3.9). In a number of cases a suggestion or inference is not preceded by an utterance of the respondent at all. The total rows refer to the number of sequences with a suggestive probe, respectively an inference.

The pattern is remarkably similar in all surveys: suggestive probes are above chance preceded by mismatch answers and considerations. Inferences seem to be caused by mismatch answers only, not by considerations. Only

Table VI. Utterances preceding the first suggestive probe (R = respondent)

Preceding utterance	1 (SAT)		2 (PWL)		3 (ENVa)		4 (ENVb)		5 (TV)	
	Freq	z	Freq	z	Freq	z	Freq	z	Freq	z
Adequate answer	7	−6.0	4	−2.6	2	−1.1	1	−2.4	84	−6.9
Mismatch answer	43	10.1	49	6.4	4	1.0	10	3.9	492	17.7
Consideration	90	4.1	8	−0.4	13	5.8	12	3.7	200	2.6
Other R utterance	54		14		6		11		141	
No R utterance	10		16		4		3		87	
Total	204		91		29		37		1004	

Table VII. Utterances preceding the first inference (R = respondent)

	1 (SAT)		2 (PWL)		3 (ENVa)		4 (ENVb)		5 (TV)	
	Freq	z	Freq	z	Freq	z	Freq	z	Freq	z
Adequate answer	22	3.2	11	2.4	1	−0.7	6	1.1	24	−3.6
Mismatch answer	15	6.8	17	2.1	5	3.6	10	5.6	191	14.7
Consideration	13	−1.6	4	−0.2	5	2.9	5	1.3	36	−2.6
Other R utterance	7		4		1		1		31	
No R utterance	0		1		0		0		0	
Total	57		37		12		22		282	

survey 3 shows a somewhat different pattern. This is a bit remarkable, because surveys 3 and 4 are quite similar with respect to types of questions, respondents and even interviewers.

Mismatch alternatives and invalid questions occur too infrequently to yield significant (both from a statistical and a practical point of view) results. Choosing behavior cannot be analyzed in the same way, because choosing concerns the absence of behavior. Sequences that end up in 'choosing' do not contain adequate answers by definition but do contain 'significant' more mismatch answers in all surveys, which is partly an artefact of the definition of choosing.

The conclusion seems to be clearcut: problematic interviewer behavior is invoked by deviations by the respondent, particularly mismatch answers and considerations.

Problematic respondent behavior on the other hand does not appear to have such clear 'causes' in terms of previous utterances; except that they are far above chance preceded by adequately posed questions by the interviewer, which is hardly surprising: answers, including mismatch answers are a common reaction to a question. Only suggestive probes appear to occur somewhat more often before a mismatch answer. Further analysis shows that these suggestive probes are usually also preceded by a mismatch answer: the respondent persists in giving a mismatch answer.

In summary, problematic respondent behavior seems to be particularly caused by the questionnaire design, whereas problematic interviewer behavior is primarily affected by respondent deviations, especially mismatch answers and considerations.

4. Effects of Problematic Deviations

4.1. SUBSEQUENT UTTERANCES

Mismatch answers are by far the most frequently occurring problematic respondent behavior. Table VIII shows what happens immediately after the

first time a mismatch answer occurs in a sequence. It should be noted that an analysis in terms of previous utterances does not necessarily reflect the results of an analysis in terms of subsequent utterances. For example, if a suggestive probe is above chance preceded by a mismatch answer, this does not necessarily imply that a mismatch answer is followed above chance by a suggestion.

Repeating the question, presenting adequate response alternatives, suggesting an answer and inference occur most often after the first mismatch answer. Of the other interviewer utterances, repeating the answer, or neutral acknowledgements like "uhmm" have by far the highest frequencies and occur above chance (if one neglects such utterances, the results are essentially similar to Table VIII, except that the frequencies are somewhat higher). In a number of cases a mismatch answer is not followed by an interviewer utterance at all.

Again, the different surveys yield quite similar patterns, and again with the exception of survey 3. Adequate reactions after a mismatch answer are repeating the question, or presenting the response alternatives (once again). Only presenting adequate alternatives appears to occur above chance after a mismatch answer. Further analysis shows that presenting response alternatives yields an adequate answer much more often than just repeating the question. Suggestive probes and inferences are inadequate reactions, but occur clearly above chance after a mismatch answer. Like presenting adequate response alternatives, these reactions appear to be highly effective, in that they yield adequate answers.

Choosing behavior is of course not suitable for the analysis of subsequent behavior; it is the absence of behavior that 'occurs' after the sequence. Requests for clarification nearly invariably yield an adequate clarification by the interviewer. The other problematic deviations occur too infrequently to warrant analyses.

Table VIII. Utterances following the first mismatch answer

	1 (SAT)		2 (PWL)		3 (ENVa)		4 (ENVb)		5 (TV)	
	Freq	z	Freq	z	Freq	z	Freq	z	Freq	z
Repeats question	16	0.6	8	−3.8	23	2.7	16	0.3	157	−1.5
Adequate altern.	109	9.9	148	5.5	0	−	0	−	571	5.9
Suggestive probe	42	11.1	41	10.9	4	0.8	9	2.5	440	21.2
Inference	17	9.7	11	3.9	2	0.6	8	3.5	156	16.1
Other utterances	303		42		138		170		1101	
No I utterance	42		21		22		28		134	
Total	529		271		189		231		2559	

4.2. PROBLEMATIC DEVIATIONS AND THE VALIDITY OF THE ANSWERS

Does interaction analysis yield information about the quality of the data? Of course one may observe that respondents give mismatches; that interviewers in response to this, suggest an answer, or just fill in a score, based on their inference of the respondent's answers; etc. But does this affect the eventual quality of the data? Maybe the interviewer's suggestions or inferences are perfectly right. Or maybe the interviewer's judgments yield even better data!

Our surveys 3 and 4 shed some light on the usefulness of interaction analysis. These surveys were held among members of an environmental organization and concerned questions about membership duration, payments to the organization, knowledge about campaigns of the organization, etc. In both surveys for 11 of the 13 analyzed questions validating information was available, e.g. from the official records from the organization.

Tables IX and X show the number and percentages of correct, incorrect and don't know responses for problematic and non-problematic sequences. To determine the correctness of a response, the eventual response as

Table IX. Response validity for problematic and non-problematic sequences in survey 3

	Non-problematic sequence		Problematic sequence		Total
Incorrect response	94	13%	63	25%	157
Correct response	579	81%	166	67%	745
Don't know response	43	6%	20	8%	63
	716	100%	249	100%	965

$\chi^2 = 22.813$; $df = 2$; $p < 0.001$; $G^2 = 21.402$; $df = 2$; $p < 0.001$.

Table X. Response validity for problematic and non-problematic sequences in survey 4

	Non-problematic sequence		Problematic sequence		Total
Incorrect response	105	12%	90	30%	195
Correct response	683	81%	166	56%	849
Don't know response	58	7%	40	14%	98
	846	100%	296	100%	1142

$\chi^2 = 70.831$; $df = 2$; $p < 0.001$; $G^2 = 66.189$; $df = 2$; $p < 0.001$.

reported by the respondent was used (available from the coded sequences), not the score filled in by the interviewer, in order to prevent that recording errors interfere with our results.

Both tables clearly show that problematic sequences yield considerably more incorrect than correct answers. Further analysis shows that in both surveys, the occurrence of mismatch answers, requests for explanation and choosing behavior is significantly related to response error ('don't know' answers were not taken into account in these analyses). The other problematic deviations were only significantly related to response errors in one of both surveys or in neither survey. This may be partly due to the very low frequencies of occurrence of these other problematic deviations. Nevertheless, the occurrence of problematic deviations seems to be related with response error. This does not necessarily imply that problematic deviations cause response errors, as will be discussed in the next section.

5. Discussion

To summarize our results thus far:

- The questionnaire design affects problematic deviations of the respondent, rather than of the interviewer
- The occurrence of mismatch answers is especially affected by the questionnaire design
- Problematic interviewer deviations are primarily caused by problematic respondent behavior, in particular mismatch answers
- Problematic deviations are related to response error.

The occurrence of mismatch answers seems to be a main factor in understanding the question answer process, both in the mind of the respondent and in the interaction with the interviewer.

Why do respondents give mismatch answers? Saying that mismatch answers are caused by the questionnaire design, is no explanation yet. We suggest two different processes that may account for the occurrence of mismatch answers, related to cognitive and to conversational factors in answering questions, respectively.

5.1. COGNITIVE PROCESSES

A response to a survey question is assumed to involve a number of steps (e.g. Sudman et al., 1996; Tourangeau et al., 2000) like understanding the question, retrieval of relevant information from memory, forming a judgment from the retrieved information and formatting the response. Mismatch answers may indicate problems with one or more of these steps. It is not very likely that mismatch answers occur because of problems

with understanding. Such problems will more likely yield a request for clarification, or an invalid answer. Retrieval of relevant information from memory, may be a point in case however. Such retrieval is often accompanied by deliberately giving considerations to a not yet given answer: a kind of 'think aloud'. Also 'don't know' answers may indicate problems with retrieval.

If we neglect all interviewer utterances, and just look at the above mentioned type of deviations, as well as adequate responses, the order in which such utterances occur may throw some light upon the cognitive process. In Table XI, the probability of the transition between these types of utterances is expressed as z-values (data are only from survey 5). Please note that a transition between two respondent utterances may often be intermediated by an utterance of the interviewer, hence z-values should be interpreted with care: a high z-value may also mean that a particular respondent utterance is followed above chance by a particular interviewer utterance, which is in turn followed above chance by a particular respondent utterance. Z-values higher than 5.00 are in bold face; z-values lower than −5.00 are italic.

First, adequate answers are typically followed by considerations. Apparently, the respondent first provides an adequate answer and then gives considerations for the answer. On the other hand, the occurrence of an adequate answer after such a consideration is far below chance.

This pattern is completely the reverse for mismatch answers. The occurrence of considerations after a mismatch answer is far below chance, whereas after a consideration, the occurrence of a mismatch answer is far above chance.

These results strongly suggest that the occurrence of mismatch answers does reflect cognitive problems. If respondents have cognitive problems in retrieving relevant information, they may accompany this process by stating aloud the relevant pieces of information. If the answer is readily available

Table XI. Transitions between a number of respondent utterances

Respondente utterance	Consideration	Adequate answer	Mismatch answer	Don't know	Invalid answer	Request for clarification
Consideration	–	−10.32	8.78	10.13	3.04	0.74
Adequate	12.30	–	−12.74	−4.51	−3.01	−0.20
Mismatch	−9.75	11.49	–	−5.97	−1.22	−1.61
DK	7.52	−5.85	−1.66	–	−1.11	1.48
Invalid	−1.34	−2.69	5.27	−1.26	–	1.89
Req. clar.	−9.57	2.68	6.40	2.29	3.48	–

(the respondent just 'knows' the answer), the respondent provides an ade-
quate answer and provides the consideration afterwards, for the benefit of
the interviewer, or to justify the response.

Don't know responses seem to take an intermediate position. A 'don't
know' is above chance followed by considerations, but don't know's also
occur above chance after considerations. Apparently 'don't know' answers
are in between mismatch and adequate answers: they may point at cogni-
tive problems, but are also legitimate responses.

Not surprising of course, mismatch answers are followed above chance
by adequate answers: this may partly reflect interference by the inter-
viewer. Adequate answers are followed below chance by mismatch answers.
This succession only occurs if the mismatch answer definitely points to a
different score than the adequate answer. For example, in response to the
question "How many days a week do you watch television on average?",
the respondent might first say "three days" (adequate answer), but sub-
sequently corrects this response with "No, most days" (mismatch answer,
because it points to a score different from "three days"). If the utter-
ance incorporates the adequate answer (like "six days", followed by "Most
days"), it is not coded as a mismatch answer.

Finally, invalid answers and requests for clarification are followed above
chance by mismatch answers, but do not seem to be related to consider-
ations. This suggests that these kinds of problematic respondent deviations
involve cognitive problems that are different from the type of cogni-
tive problems that occur if mismatch answers are preceded by consider-
ations: invalid answers and requests for clarification indicate problems with
understanding, whereas considerations preceding mismatch answers indicate
problems with retrieval.

Mismatch answers are related to response errors (Section 4.2). If mis-
match answers express cognitive problems with retrieval, this relation need
not be a causal one. Rather, such problems themselves may be related
to both the occurrence of mismatch answers and response errors, mak-
ing the relation between mismatch answers and response errors a spurious
one.

5.2. THE INTERVIEW AS A CONVERSATION

Cognitive problems are not the whole story in explaining the occurrence of
problematic sequences, and mismatch answers in particular. In many cases
there is no reason to expect that the answer to a survey question is not
readily available, but nevertheless mismatches occur. Consider these exam-
ples of a question after the respondent's age:

    I: And if I may ask, what is your age?
    R: I'm nearly seventy.
    I: Nearly seventy. Then I fill in 70, OK?
    R: Sure, it's only a few months.
    I: OK.

and

    I: And what is your age, if I may ask?
    R: Well uhm, I'm retired.
    I: Uhm OK.

In the first example, the response "I'm nearly seventy" does not necessarily indicate that the respondent is 69; he may very well be 68, hence the answer is a mismatch. What matters here, however, is that in both cases there is no reason to expect that the required information is not readily available. Nevertheless, these respondents give a mismatch answer (and a mismatch answer was the first response to the question after the respondent's age in 14 out of the 197 sequences in survey 5).

As another case in point, in Section 3.1 we discussed that in survey 5, assertions were quite often answered with "yes" or "no" instead of one of the five response alternatives. It is quite unlikely that the respondents encountered problems with retrieval that often.

According to Schaeffer (2002: 98): "... respondents may apply rules learned in other speech events and reject to impose interview rules". Suchman and Jordan (1990) argue that the way a survey interview is standardized, and is not like a conversation, may hamper the quality of the data. Tourangeau (1990, in a comment on Suchman and Jordan, 1990) raises the problem that conversation could have "its own characteristic problems, so making face-to-face interviews more like conversations would introduce a new set of errors into the survey process at the same time as it reduced others".

An interview is often viewed as a conversation with a purpose (Cannell and Kahn, 1968), but the purpose of the respondent may be quite different from the purpose of the researcher. Respondents may derive their motivation to take part in a survey interview from the very fact of becoming engaged in some kind of conversation. After all, what is the benefit for the respondent to just provide the exact and precise information required by the purpose of the standardized survey-interview? In a usual conversation, responses like "most days" or "retired" are quite common. If such 'conversational' responses are not among the response alternatives, they are by definition mismatch answers. Moreover, such 'conversational' answers may be stimulated by the way questions themselves are worded. A question like "What's your age" is quite common in a conversation; a question like "In which year were you born" is not. Hence, the former question may 'trigger'

the respondent to answer in a more conversational style too, whereas "In which year were you born" will not (see also Ongena, 2005).

The difference between the researcher's and the respondent's purpose may also suggest potential effects of elucidations and irrelevant talk. According to Suchman and Jordan (1990), "the interview, rigidly constrained by an externally imposed, often repetitious script, becomes observably boring to respondents". Elucidations and irrelevant talk may prevent such boredom and may keep the respondent motivated and thus may affect the validity of the information provided in a positive way. Dijkstra (1987) found that a personal style of interviewing (invoking relatively many elucidations and irrelevant utterances from the respondent) yielded more valid responses than a formal style of interviewing. However, although uttering elucidations and irrelevant talk may satisfy the respondent's purpose, it may also be at variance with the proper task of the respondent: giving adequate answers to questions from the questionnaire.

In survey 5, in 293 (3.8%) sequences, the respondent gave elucidations or irrelevant utterances. Of these 293 sequences, 41% of the first problematic deviations, were mismatch answers, whereas this percentage was 31 for the other sequences (without elucidations and irrelevant utterances). It is even more interesting to look at sequences during which such elucidations and irrelevancies did *not* occur, but *did* occur in the immediately preceding sequence. Mismatch answers occurred in 44% of the sequences having themselves no elucidations or irrelevant utterances but were immediately preceded by sequences with elucidations and irrelevancies. If neither the sequence itself, nor the immediately preceding sequence contained elucidations and irrelevancies occurred, this percentage was only 31. Apparently, elucidations and irrelevancies invoke a conversational style of responding to a survey question, which may explain the higher occurrence of mismatch answers.

Here is a clear contradiction. Allowing a more 'conversational' interview may motivate the respondent to give more *valid* answers on the one hand, but may also evoke less *precise* answers, i.e. mismatch answers. This in turn may lead the interviewer to attack the problem with inadequate solutions, like suggestive probes, inferences or even choosing, probably leading to less valid data.

Here the difference between cognitive and conversational causes becomes important. In Section 5.1 we suggested that mismatch answers need not cause response errors, if mismatch answers reflect cognitive problems: both mismatch answers and response errors may be caused by cognitive problems. The situation may be different however in case a mismatch answer results from a conversational way of responding. Although respondents may be better motivated to provide information and to think about their responses, the need to formulate exact responses, that fit one of the

response alternatives, may also be less important to them. Stated otherwise, in case of cognitive problems, respondents may be well aware that their mismatch answer is not an adequate response. For example, a response like "between agree and strongly agree", is apparently not a conversational response, but rather indicates that the respondent does not have enough information to decide between both alternatives, and is well aware of this. On the other hand, the respondent may be convinced that a "yes" answer to the assertion "TV commercials are funny to look at", is quite an adequate response. More important, a suggestion like "Strongly agree?" from the interviewer is more likely to be affirmed by the respondent in case of a 'conversational mismatch' than in case of a 'cognitive mismatch'. A suggestion in case of a 'cognitive mismatch', may even stimulate the respondent to restart cognitive processing and to come up with an adequate and more valid response, whether or not in accordance with the suggestion.

5.3. TOWARDS A MODEL OF THE SURVEY INTERVIEW

The relations we discussed in this paper, both the relations that we empirically found and the relations we hypothesized in the previous sections, can be incorporated into a process model of the survey interview (see Figure 1).

The model only concerns errors; it does not contain (causes of) adequate responses; nor does it contain adequate (interviewer) solutions to problems created by the respondent. As a way of summary, we will briefly discuss this model here.

Important questionnaire factors are the difficulty of the question, the use of show cards and the wording of the question. The difficulty concerns the amount of cognitive processing that is required by the respondent. Difficult questions may be retrospective questions or questions with unclear concepts. It is assumed that the difficulty enhances the probability of providing considerations ("think aloud") before an answer to the question itself is given, of mismatch answers because of cognitive problems and of requests for clarification.

The use of show cards was shown to enhance the probability of requests for clarification (see Table V). Moreover we hypothesize that it will decrease the probability of 'conversational' mismatch answers. Important factors that enhance the probability of 'conversational' mismatches are questions worded as assertions or 'yes/no' questions, whereas 'yes' and 'no' are not among the response alternatives.

Mismatch answers affect the probability of inadequate interviewer behavior (suggestions, inferences and choosing). In case of cognitive mismatches, suggestive probes and inferences need not necessarily lead to more response errors; it is equally possible that such reactions just stimulate further
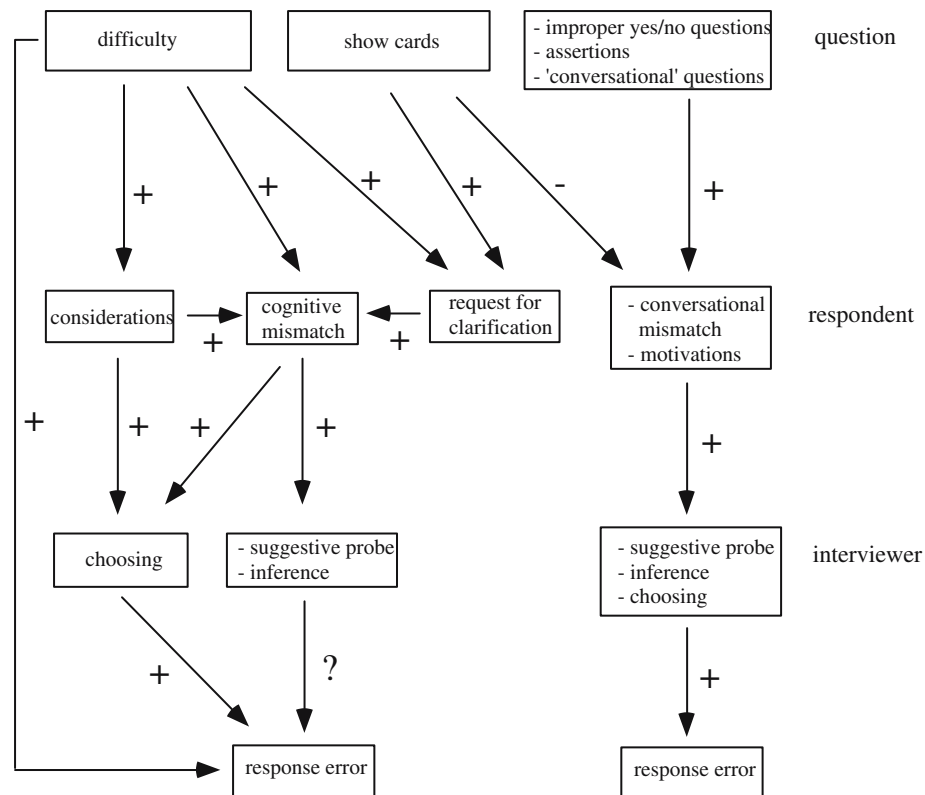
Figure 1. A model of errors in the survey interview.

cognitive processing by the respondent, and may even lead to more valid responses. Choosing behavior does not have this function, and is assumed to yield more response error.

Inadequate interviewer behavior after a conversational mismatch however, will generally lead to more response error. Typically, respondents agree with the interviewer.

Future research will be directed to means of distinguishing 'conversational' mismatch answers from 'cognitive' ones, e.g. on the basis of question characteristics, reaction times (we hypothesize that cognitive mismatch answers have longer response latencies than conversational mismatch answers) and the kind of mismatch answer. In addition, it is quite important to have validating information available to test the model.

Acknowledgements

commercials (Smit, 1999). The authors are also indebted to the members of the department of Social Research Methodology for their valuable comments on an earlier draft of this paper and to Nora Cate Schaeffer for her comments on the QDET version of the paper and who inspired us to develop the model presented in this paper.

Appendix A: Description of some Deviations

PROBLEMATIC RESPONDENT DEVIATIONS

A *mismatch answer* is an answer to a question from the questionnaire, but does not unequivocally point to a score that can be filled in by the interviewer. Suppose the response alternatives are 'never', 'seldom', 'sometimes', 'often' and 'always'. In that case a response like 'not always', is a mismatch answer. As another example, suppose the question reads: "How many days a week do you watch television, on average?" A response like "most days", although an answer to the question, is a mismatch answer, because a number between 0 and 7 is required. If the response unequivocally points to a particular score answer (e.g. "I look every day") it is not a mismatch answer, although the response is not in accordance with the required format. Decisive is whether the interviewer is required to probe to obtain a scorable answer, not whether the response is according to the required response format. For example a question from survey 3: "Did you pay more than 25 guilders subscription to 'milieudefensie'?" A response like "I paid 30 guilders" is not a mismatch, although the alternatives are 'yes' and 'no'. A probe like "Is that yes or no" does not make sense, or may even disturb the relationship between interviewer and respondent. The concept of mismatch answer is quite similar to what is labeled by Houtkoop-Steenstra (2000) as 'unformatted answer'. The term 'inadequate answer' has a broader and less strict meaning.

An invalid answer is an answer to a question that is misunderstood by the respondent, as far as can be determined by the interaction. For example, among questions about education in survey 2, one question was "How satisfied are you with the course you followed?" An answer like "I'm very satisfied that I followed it" is an invalid one.

A *request for clarification* is a request by the respondent to explain the respondent's task or the meaning of the question. It is viewed as a problematic deviation because it requires action from the interviewer.

PROBLEMATIC INTERVIEWER DEVIATIONS

An *invalid question* significantly alters the meaning of the question. For example, if the question reads "How satisfied are you with the course you

followed?", but the interviewer asks "How satisfied are you that you followed the course?", this is a significant alteration of the meaning of the question, although the words are nearly the same. On the other hand, if the question reads "What is your age?", but the interviewer asks "How old are you?", this is not viewed as an invalid question. Although the wording is completely different there is no reason to expect that this would affect the validity of the response. Such an alteration without changing the meaning is called a *mismatch question*; mismatch questions are not viewed as problematic because they are not expected to affect the validity of the response. Also a clarification of a question can be invalid, if the clarification changes the meaning of the question as intended by the researcher.

Mismatch alternatives are significant changes in the response alternatives, that may affect the response, but without changing the meaning of the alternatives themselves. In most cases this code was applied when not all response alternatives were read: usually because the interviewer skipped one of the alternatives. If skipping alternatives implied an unwarranted suggestion, it is viewed as suggestive probing (see below). If only part of the response alternatives were read because of an interruption by the respondent, this is not viewed as a mismatch (the availability of sound files appeared to be very important to decide upon whether or not only part of the alternatives were read because of an interruption; this does not appear from transcripts only). A somewhat different category is *invalid alternatives*, a code that is applied when the meaning of the alternatives is altered. This hardly ever occurred in our datafiles, and for the sake of simplicity, both categories are taken together. As an example, in survey 5, about TV commercials, the assertion "TV commercials are funny to look at", was accompanied by five alternatives ranging from strongly disagree to strongly agree. In one sequence the interviewer presented alternatives ranging from "Not at all funny" to "Very funny".

Suggestive probing means that the interviewer presents a particular response, or a limited set of response alternatives to the respondent, in a questioning voice. The presented response or set of responses are not warranted by previous answers of the respondent. An example of suggestive probing is:

I: How many days a week do you watch television, on average?
R: Most days.
I: Is that six days?

"Is that six days?" is a suggestive probe. However, "Is that 4, 5, 6 or 7 days" is not suggestive. Although a limited set of alternatives is presented, this limitation is warranted by the previous answer of the respondent.

An *inference* is an inadequate repetition of the respondent's answer (a *mismatch* between a repetition and the utterance that is repeated). Like

suggestive probing, an inference is not warranted by a previous response, but is posed without a questioning voice. For example:

I: How many days a week do you watch television, on average?
R: Most days.
I: Seven days. OK.

Choosing concerns the absence of behavior: the interviewer apparently bases the eventual score on insufficient information; for example only mismatch answers, or even the absence of a response to the question at all, whereas the respondent remains uninformed about the score filled in by the interviewer. This is not a code applied to a particular behavior, but inferred from the (order of) codes of the whole sequence. For example, if the respondent answers "Most days" in response to the question "How many days a week do you watch television, on average?", but, unlike an inference, the interviewer does not mention the score that she fills in (but just says "OK", or immediately proceeds with the next question).

NON-PROBLEMATIC DEVIATIONS

Question–answer sequences can deviate from the paradigmatic sequence in quite a lot of not necessarily problematic ways. These deviations may not only have a detrimental, but also a beneficial effect, or no effect at all on the quality of the response. We will discuss some of them here, because they are informative for the course of the interaction and do affect the eventual response in a number of cases.

An utterance is called a *request for repetition* if the respondent (or interviewer) asks to repeat (part of) an utterance of the other actor. A request for repetition by the respondent is not viewed as a problematic deviation. Although frequent requests to repeat a particular question, *may* point to problems with such a question, such a request usually does not indicate problems with the task. Rather, a request for repetition is usually affected by variables like the clearness of the interviewer's voice, hearing difficulties of the respondent, background noise, bad telephone lines, long questions, or a large number of response alternatives.

Identifying an utterance as a request for repetition is not always easy. It is quite common that, after a question is posed by the interviewer, the respondent repeats (part of) the question in a more or less questioning voice. Such an utterance should usually not be coded as a request for repetition. The very fact that the respondent does repeat (part of) the question, implies that there are no problems with intelligibility.

A request for repetition should be carefully distinguished from a request for clarification. The difference becomes quite clear if one realizes that a question that yields many requests for repetition, should preferably be

posed using a self-administered questionnaire, which will solve all problems mentioned above. A question that yields many requests for clarification, should preferably be posed by an interviewer, if reformulation does not solve the problem.

Omitting alternatives is a deviation indicated by the absence of behavior: the interviewer does not read the alternatives at all. If the interviewer reads only part of the alternatives, this is not viewed as omitting alternatives. In many cases, omitting alternatives can be viewed as a serious problem. This is not always the case however. It often occurs that the respondent already answers the question, before the interviewer started to read the alternatives. If this answer equals one of the response alternatives, it may be a bit awkward, if the interviewer nevertheless presents the alternatives. Such behavior can be viewed by the respondent that the interviewer is not really listening, or, even worse, that the interviewer does not believe the respondent. For example (alternatives are 'never', 'sometimes' and 'often'):

> I: How often do you watch erotic movies on television? Is that...
> R: No, never!
> I: Is that never or sometimes?

DK answers is not viewed as problematic because a 'don't know' does not necessarily indicate that the respondent does not understand the task. Moreover, insisting that a respondent gives a substantial response in case of a 'don't know' may also worsen the quality of the eventual response. Please note that a DK answer may be different from the eventual response. A DK answer refers to an utterance in the course of the interaction, which may or may not be the same as the eventual response.

A *refusal to give an answer* is not viewed as a problematic deviation. In our opinion a respondent has a perfect right to refuse, and a refusal is a legitimate response and does not mean that the respondent has problems with the task and does not necessarily require action from the interviewer. In all our datasets, such refusals hardly ever occur (four times in survey 1, all in response to the question "Does there live anybody in this neighbourhood whom you dislike?", two times in survey 5 and never in the other surveys). Thus, refusals are hardly of any importance here, but refusals may occur more often if sensitive topics are involved.

Considerations concern reasons to explain the answer given by the respondent ("I'm very satisfied, because … "). Providing a consideration to an answer by the respondent is not viewed as problematic. There is no reason at all for the interviewer to explain the respondent's task if the respondent motivates the answer. This is not to say that observing that particular answers yield more considerations than others, is uninformative. First, such considerations can provide insight in the respondent's cognitive processes. Second, questions that yield such respondent behavior, may

be more difficult to answer than questions that do not yield this behavior. Third, providing considerations may distract the respondent from the task to *also* provide a scorable answer: after having mentioned a number of considerations, the respondent may be convinced that the question is sufficiently answered.

However, in our opinion, it would be a bad strategy to design a questionnaire in such a way, that it yields less considerations provided by the respondent. After all, providing considerations indicates that the respondent *does* some cognitive processing. A question that stimulates this, may be a better question than a comparable one that does not stimulate this. Moreover, considerations may inform the interviewer about misunderstanding the question proper.

The difference between *elucidations* and *irrelevant* talk is that elucidations pertain in some way to the question topic, whereas irrelevant talk does not.

Forward and *backward* references concern utterances of interviewer and respondent that refer to subsequent, respectively previous questions: "We will talk about that in a few minutes" or "As I said before...". There is no reason to view such behavior as problematic. However, if a particular question yields many forward and backward references, this may indicate that questions are too similar, or that the order of the questions may affect the responses.

The presence of a *bystander*, during the interview may undoubtedly affect the response, but it is unclear whether this is in a positive or negative sense. For example, the sheer presence may prevent socially desirable answers on particular questions, but may stimulate socially desirable answers on different questions. In surveys 3 and 4, although these were telephone interviews, it was clear that in some cases the respondent asked information from his or her partner (e.g. with respect to the amount of contribution paid to the environmental organization). If anything, this may have had a positive effect on the responses, because the questions concerned factual information. In survey 1, the effect of the partner was most prominent: in thirty sequences the partner provided considerations, in order to 'help' the respondent in answering a question and in fourteen sequences the partner even answered the question instead of the respondent. All questions in these fourteen sequences concerned some form of a judgment, and, quite remarkably, in most cases the respondent disagreed with his or her partner. Although we do not view interference by a partner as desirable, it is unclear what the effect is on the quality of the responses.

A final error that can be made by the interviewer is wrong *registration of the response*. This does, by definition, not affect the answer given by the respondent and is thus, according to our definition of problematic interviewer behavior, no problematic behavior.

Note

1. These are the types of behavior that are most controversial from the standpoint of strict standardization. For example, if the respondent asks for clarification, the interviewer should respond with something like "Whatever it means to you". But consider the following (real) example:

   I: How many persons live in your household, including yourself?
   R: Household?
   I: Yes
   R: What do you mean, to clean up the house?
   Responding with "Whatever it means to you" instead of explaining the concept will quite probably yield less valid answers.

References

Cannell, C. F. and Kahn, R. L. (1968). Interviewing. In: G. Lindzey and E. Aronson (eds.), *The Handbook of Social Psychology*, Vol. 2. Reading, Mass.: Addison-Wesley, pp. 526–595.

Dijkstra, W. (1987). Interviewing style and respondent behavior: an experimental study of the survey-interview. *Sociological Methods and Research* 16(2): 309–334.

Dijkstra, W. (1999). A new method for studying verbal interactions in survey-interviews. *Journal of Official Statistics* 15: 67–85.

Dijkstra, W. (2002). Transcribing, Coding, and analyzing verbal interactions in survey interviews. In: D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer and J. van der Zouwen (eds.), *Standardization and Tacit Knowledge; Interaction and Practice in the Survey Interview*. New York: Wiley, pp. 401–425.

Fowler, F. J. and Mangione, T. W. (1990). *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. Newsbury Park, CA: Sage.

Houtkoop-Steenstra, H. (2000). *Interaction in the Standardized Survey Interview: The Living Questionnaire*. Cambridge: Cambridge University Press.

Ongena, Y. P. (2005). Interviewer and Respondent Interaction in Survey Interviews. Ph D dissertation, Vrije Universiteit, Amsterdam.

Ongena, Y. P. and Dijkstra, W. (forthcoming). Methods of Behavior Coding in Survey Interviews. Journal of Official Statistics.

Prüfer, P. and Rexroth, M. (1985). Zur Anwendung Der Interaction-Coding-Technik. *ZUMA-Nachrichten* 17:2–49.

Schaeffer, N. C. (2002). Conversation with a purpose – or conversation? Interaction in the standardized survey-interview. In: D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer and J. van der Zouwen (eds.), *Standardization and Tacit Knowledge*. New York: Wiley, pp. 95–123.

Schaeffer, N. C. and Maynard, D. W. (1996). From paradigm to prototype and back again: interactive aspects of cognitive processing in standardized survey interviews. In: N. Schwartz and S. Sudman (eds.), *Answering Questions. Methodology for Determining Cognitive and Communicative Processes in Survey Research*. San Francisco: Jossey-Bass, pp. 65–88.

Schober, M. F. and Conrad, F. G. (1997). Does conversational interviewing reduce survey measuremnet error? *Public Opinion Quarterly* 61: 576–602.

Smit, E. G. (1999). Mass media advertising: information or wallpaper? Unpublished Ph. D. dissertation, University of Amsterdam, Amsterdam.

Suchman, L. and Jordan, B. (1990). Interactional troubles in face-to-face survey interviews. *Journal of the American Statistical Association* 85: 232–241.

Sudman, S., Bradburn, N. M. and Schwarz, N. (1996). *Thinking about Answers*. San Francisco: Jossey-Bass.

Sykes, W. and Collins, M. (1992). Anatomy of the survey interview. *Journal of Official Statistics* 8: 277–291.

Tourangeau, R., Rips, L. C. and Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

Tourangeau, R. (1990). Comment [on Suchman and Jordan, 1990]. *Journal of the American Statistical Association* 85: 250–251.

Viterna, J. S. and Maynard, D. W. (2002). How uniform is standardization? Variation within and across survey research centers regarding protocols for interviewing. In: D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer and J. van der Zouwen (eds.), *Standardization and Tacit Knowledge; Interaction and Practice in the Survey Interview*. New York: Wiley, pp. 365–397.

Willis, G. B. (2005). *Cognitive Interviews. A Tool for Improving Questionnaire Design*. Thousand Oaks: Sage.