

Dissertation Faculty of Social Sciences

Probing Behaviour in Open Interviews

**A Field Experiment on the Effects of Probing Tactics
on Quality and Content of the Received Information**

Gerben Moerman



Probing Behaviour in Open Interviews
A Field Experiment on the Effects of
Probing tactics on Quality and Content of the
Received Information

ISBN: 9789086594436

© 2010 by Gerben Moerman, Amsterdam, The Netherlands

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without prior permission from the author.

Cover: Gerben Moerman

Copy editing: Gail M. Zuckerwise

VRIJE UNIVERSITEIT

Probing behaviour in open interviews
A field experiment on the effects of
Probing Tactics on Quality and Content of the
Received Information

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L.M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Sociale Wetenschappen
op vrijdag 16 april 2010 om 11.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Gerben Albert Moerman

geboren te Schipluiden

promotor: prof. dr. J. van der Zouwen
copromotor: dr. H. van den Berg

Is peace the answer,
then what's the question?
Fight the fight, Living Colour, 1990

Thesis committee:

dr. H.R. Boeije (University of Utrecht)

prof. dr. W. Dijkstra (Vrije Universiteit Amsterdam)

prof. dr. G. Loosveldt (K.U.Leuven)

prof. dr. F.P.J. Wester (Radboud University Nijmegen)

Contents

ACKNOWLEDGMENTS	I
1 PROBING BEHAVIOUR IN 'OPEN' INTERVIEWS	1
1.1 INTRODUCTION	1
1.2 A TYPOLOGY OF INTERVIEWS	2
1.3 INTERVIEWER BEHAVIOUR	5
1.3.1 TASK-ORIENTED INTERVIEWER BEHAVIOUR	5
1.3.1.1 Task-oriented interviewer behaviour in closed interviews	5
1.3.1.2 Task-oriented interviewer behaviour in open interviews	6
1.3.2 PERSON-ORIENTED INTERVIEWER BEHAVIOUR	8
1.3.2.1 Person-oriented interviewer behaviour in closed interviews	8
1.3.2.2 Person-oriented interviewer behaviour in open interviews	10
1.3.3 THE RELATION BETWEEN THE TWO DIFFERENT ORIENTATIONS	13
1.4 PROBING TECHNIQUES FOR OPEN INTERVIEWS	14
1.4.1 TYPES OF PROBING TECHNIQUES	15
1.5 PROBING TACTICS	19
1.5.1 WHAT ARE PROBING TACTICS?	19
1.5.2 PROBING TACTICS VERSUS INTERVIEWING STYLE	21
1.5.3 THE DIFFERENT PROBING TACTICS AND THEIR RESPECTIVE LOGICS	22
1.5.3.1 Accommodating probing tactic: implicitly (not-)taking for granted	22
1.5.3.2 Encouraging probing tactic: explicitly taking for granted	23
1.5.3.3 Challenging probing tactic: explicitly not-taking for granted	24
1.6 RESEARCH QUESTION	26
2 COMPARING INTERVIEWS ON QUALITY	31
2.1 ISSUES OF QUALITY IN OPEN INTERVIEWS	31
2.1.1 ASSESSMENT OF INTERVIEWER BEHAVIOUR	31
2.1.2 ASSESSMENT OF THE ROLE OF INTERACTION IN ACQUIRING INFORMATION	32
2.1.3 ASSESSMENT OF THE QUALITY OF THE RECEIVED INFORMATION	33
2.1.3.1 The criteria <i>not</i> chosen for the assessment of the quality of the received information	34
2.2 INTERVIEW TYPE, TOPIC AND ANALYTIC FOCUS IN THIS RESEARCH	36
2.2.1 THE INTERVIEW TYPE: CONCEPT CLARIFICATION INTERVIEW	36
2.2.2 THE INTERVIEW TOPIC: SOCIAL CATEGORISATION	37
2.2.3 THE ANALYTICAL FOCUS: MEMBERSHIP CATEGORISATION ANALYSIS	37

2.2.3.1	Membership categorisation coding	39
2.3	CRITERIA FOR THE QUALITY OF INFORMATION RECEIVED IN INTERVIEWS	40
2.4	QUALITY OF RECEIVED INFORMATION: A SUMMARY	45
3	DESIGN AND METHODOLOGY OF THE DATA COLLECTION	47
3.1	INTRODUCTION	47
3.2	GENERAL DESIGN	47
3.2.1	EXPERIMENTAL DESIGN	48
3.2.1.1	Experimental variables: probing tactics	49
3.2.1.2	Training	50
3.2.2	DESIGN OF THE INTERVIEW STUDY	50
3.2.2.1	Interview topics	50
3.2.2.2	Questionnaire	51
3.2.2.3	The research population	51
3.2.2.4	Interviewers	52
3.2.2.5	Respondents	53
3.3	PILOT STUDY	53
3.3.1	SELECTION OF INTERVIEWERS	54
3.3.2	SELECTION OF RESPONDENTS	54
3.3.3	TRAINING	54
3.3.4	THE INTERVIEWS	55
3.3.5	QUESTION DEVELOPMENT	55
3.3.6	ANALYSIS OF FEASIBILITY	55
3.4	METHODOLOGY OF THE MAIN EXPERIMENT	57
3.4.1	ORGANISATION OF THE EXPERIMENT	58
3.4.2	TREATMENT	58
3.4.2.1	Interviewer instruction	58
3.4.2.2	Training	59
3.4.3	QUESTIONNAIRE	61
3.4.4	SELECTION AND ASSIGNMENT	65
3.4.5	SELECTION OF INTERVIEWERS	65
3.4.6	SELECTION OF RESPONDENTS	66
3.5	THE INTERVIEWS	70
4	DESIGN AND METHODOLOGY OF THE DATA ANALYSIS	73
4.1	DATA? WHAT ARE THE DATA?	73
4.2	TRANSCRIPTION	74

4.3	WHAT ARE THE UNITS OF OBSERVATION AND ANALYSIS?	76
4.4	CODING	77
4.4.1	GENERAL CODES	79
4.4.2	INTERVIEWER BEHAVIOUR CODES	80
4.4.3	RESPONDENT ANSWER CODES	82
4.4.4	SELECTION OF CODERS	84
4.4.5	TRAINING OF THE CODERS	84
4.4.6	CODING PROCESS	86
4.4.6.1	Coding process, phase 1: interviewer behaviour codes	86
4.4.6.2	Coding process, phase 2: respondent answer codes	88
4.5	RELIABILITY	90
4.5.1	RELIABILITY TESTS	90
4.5.1.1	Reliability of interviewer behaviour coding	90
4.5.1.2	Reliability of respondent answer codes	91
4.5.2	RELIABILITY IMPROVEMENT BEYOND POST-HOC TESTS OF THE INSTRUCTION	92
4.5.2.1	Improvement by group work	92
4.5.2.2	Improvement by checks and comparison	93
4.6	FROM CODES TO STATISTICS	94
5	QUALITY OF INTERVIEWER BEHAVIOUR	97
5.1	QUALITY IN SCRIPTED INTERVIEWER BEHAVIOUR	98
5.1.1	ILLEGITIMATELY SKIPPING QUESTIONS	98
5.1.2	REFORMULATION OF QUESTIONS	99
5.1.3	QUALITY OF THE QUESTIONING PER QUESTION	100
5.2	QUALITY IN NON-SCRIPTED INTERVIEWER BEHAVIOUR	101
5.2.1	ASSESSMENT OF THE INTERVIEW	102
5.2.1.1	Interview assessment by the interviewer	102
5.2.1.2	Interview assessment by the respondent	103
5.2.2	INTERVIEWER MISTAKES IN NON-SCRIPTED INTERVIEWER BEHAVIOUR	104
5.3	PROBING	105
5.3.1	THE AMOUNT OF PROBE TURNS	106
5.3.2	PROBING TECHNIQUES	107
5.3.3	PROBING TACTICS	110
5.3.3.1	Sets of preferred techniques	110
5.3.3.2	Quality of the probing tactics using multivariate analysis of variance	111
5.3.3.3	Quality of probing tactics using a classification technique	113
5.4	CONCLUDING REMARKS ON INTERVIEWER BEHAVIOUR	115

6	PROBING TACTIC AND QUALITY OF INFORMATION	119
6.1	RELEVANCY OF THE INFORMATION	119
6.2	DEPTH OF THE INFORMATION	119
6.3	AMOUNT OF THE INFORMATION	120
6.4	ELABORATENESS OF THE INFORMATION	122
6.5	SPECIFICITY OF THE INFORMATION	123
6.6	CONCLUDING REMARKS ON THE EFFECTS PROBING TACTICS HAVE ON THE QUALITY OF THE INFORMATION	124
7	PROBING TACTIC AND CONTENT OF INFORMATION	127
7.1	THE COMPARISON OF THE CONTENT OF THE INTERVIEWS	127
7.2	THE AMSTERDAMMER PART OF THE INTERVIEW	128
7.2.1	THE AMSTERDAMMER QUESTIONS AND ANSWERS	128
7.2.2	THE PREDICATE TYPES FOR AMSTERDAMMER CATEGORIES	129
7.3	THE FRIEND PART OF THE INTERVIEW	130
7.3.1	THE FRIENDS QUESTIONS AND ANSWERS	130
7.3.2	THE PREDICATE TYPES FOR FRIEND CATEGORIES	132
7.4	THE ALLOCHTHON PART OF THE INTERVIEW	133
7.4.1	THE ALLOCHTHONS QUESTIONS AND ANSWERS	133
7.4.2	THE PREDICATE TYPES FOR ALLOCHTHON CATEGORIES	135
7.5	CONCLUDING REMARKS ON THE EFFECTS PROBING TACTICS HAVE ON THE CONTENT OF THE INTERVIEWS	138
8	CONCLUSION	141
8.1	SHORT SUMMARY OF THE RESEARCH AND ITS CONCLUSIONS	141
8.2	WIDER POTENTIAL	143
8.3	PROBING TACTICS AND FEEDBACK OPTIONS FOR RESPONDENTS	145
8.3.1	HOW TO ENSURE ROBUSTNESS	146
8.4	FINAL WARNING	148
	BIBLIOGRAPHY	151
	SUMMARY IN DUTCH / SAMENVATTING	161

Acknowledgments

In his admirably written book *Outliers: the story of success* Malcolm Gladwell (2008) shows what everyone who has finished a PhD-thesis already knew: you need more than just your own talent and dedication to be successful. Having a good background, being in the right spot at the right time, having others supporting you and receiving opportunities is at least as important. For me these aspects -and one more- have been important in finishing this PhD.

In his book Gladwell does not really delve into the most important aspect I had to rely on while writing this thesis: patience of the people surrounding me in private and academia. The patience that my promotor Hans van der Zouwen and my co-promoter Harry van den Berg have shown has been incredible. While both of them have at least partly retired, they gave me the support I obviously needed, even when normal human beings in terms of patience would have told anyone to drop dead. I abused their time with long and winding pages of unfinished work and I probably frustrated them with my stubbornness fighting windmills. Hans, I am very thankful for your enthusiasm about the research; your kind words in the meticulous memos were more encouraging than you probably knew. Harry, thank you for the inspiring discussions we had on this research or anything else in the world, and the effort you've shown in all things that come to 'daily supervising' and beyond.

Since I knew beforehand that I was not going to be able to do a PhD in solitude, I wanted to be part of a larger project. I could not have been luckier than to have Christine Carabain as my fellow. I was in the right spot at the right time, since she started a few months before me, and I could simply copy the steps she was taking for the first few years. Unfortunately for me, she was much quicker, so I lost the chase. Christine, thank you for the support, discussion and your friendship.

I was also in the right spot at the right time when starting my PhD at the VU. The atmosphere among the group of people working at the Research Methodology Department (M&T) was great. Besides this collegiality, the colloquia were instructive and inspiring. I would like to thank the colleagues at the VU for their advices and kindness in helping me out in so many ways (as a sidekick during the interviewer trainings, as an expert looking at my questionnaire, as a test-respondent, as a pilot study interviewer, providing me with 8 computers for coding, teaching me about entropy, to name just a few).

As part of this PhD-project, I joined the Interuniversity Graduate School of Psychometrics and Sociometrics (IOPS). Being a qualitatively trained 'non-western sociologists', the courses that I took in this research school were intriguing and more than helpful. I would like to thank all members of IOPS who gave any comments or advice at the IOPS conferences.

This research would not have been possible without the support of many people during the fieldwork and the coding. I want to thank O+S Amsterdam and especially Willem Bosveld for providing me with a sample of Amsterdammers. I am appreciative of the effort Thomas Muhr of ATLAS.ti has put into fixing bugs and issues and even generating special functions so I could deal with my enormous files. I am indebted to the following groups of people: all interviewers for their enthusiasm and their good work, all transcribers for their hard work and all coders for the accurate coding (and the 'gezelligheid' of course). I also want to make a case sensitive thank you to Gail Zuckerwise for copy editing my text. I am appreciative to Huib van Nimmerdor for his support in printing this book.

Besides the support of others, Gladwell also mentions receiving opportunities as necessary causes for success. I am indebted to many colleagues (and sometimes even students) at the University of Amsterdam who gave me the opportunity to sometimes prioritise my thesis over other work. I am grateful for the support of my roommate Carolien Bouw whom I can always count on.

I'm indebted to two of my best friends, Richard van Alphen en Vincent Wareman. They have supported me practically, in organising, transcribing and even coding, but even more socially, by not blaming me when I was always too busy for them. I owe you quite some pints and friendship.

Gladwell shows that a good family background is the root of success. Well, I have quite some roots. I thank my parents, who have supported and motivated me while working on this PhD as they have done wholeheartedly throughout my life.

The largest THANK YOU goes out to my love, Lies. You suffered sometimes under my stress, sleepless nights and mental absence, but you still gave me everything to help me follow my dream: you offered so many opportunities, support, care and advices (you know I wish that I had listen better sometimes). But above all, you gave me the success that counts mostly: love and our two daughters, Anne and Jette. Heel veel dank voor alle geluk meisjes. Dit boek is voor jullie!

Gerben

Maarssen, February 2010

1 Probing Behaviour in ‘Open’ Interviews

1.1 Introduction

This study is about the behaviour of interviewers in open interviews and the possible effects it has on the information received in these interviews. Usually, when explaining my research topic to lay people, they start nodding and come up with examples from their own lives or from what they have seen on TV. This is no surprise since, according to Atkinson and Silverman, we live in an ‘Interview society’ (Atkinson & Silverman, 1997) in which everybody runs into interviews in one way or another. We are interviewed by educators, clinicians, opinion pollsters, customs officers and employers. It is likely that we have even conducted an interview, ourselves, during our own education. And, on television, on the radio and throughout the Internet, we run into many entertaining interviews held by journalists.

Not only is the personal interview very common in daily life, but also, it remains widely used in social research. It is probably not used in 90% of all research, as was estimated by Brenner in 1981 (Briggs, 1986), since document analysis is increasingly used. But, in contemporary qualitative research, the interview is still the prime method (Cassell, 2005).

Since the interview is used so widely, one would expect a large range of methodological research on interviewing methods. This expectation is fulfilled largely of interviewer behaviour in survey interviews with closed questions. A large amount of books and articles have been published on the effects of interviewer behaviour (see for example Dijkstra, 1983). Additionally, a large amount of books and articles have been published about research interviewing using interviews with open-ended questions. However, most of these articles and books are based on the experiences gained by the author(s) while conducting the interviews. Using autobiographical anecdotes, these authors explain how to interview to novice interviewers. Reading those books and articles, professional interviewers will probably recognise experiences and will possibly reconsider their own interviewing strategies. These reflexively written books are excellent for these goals. However, there is a methodological problem with the autobiographical experience they are based on: Would the same strategy work in different situations, with different interviewers, with different respondents and with different topics? Or, is there any strategy that is best in most cases?

Usually, when explaining my research topic to interviewers and researchers, they start nodding and come up with examples from their own research experiences. Typically, they also start talking about their uncertainties, as researchers and interviewers, of the effects that their interviewing behaviour could have on the quality and content of the

received information. “Could the information have been different, if I would had interviewed differently?” or “How could I have done it better, in order to acquire better information?” The goal of my research is to empirically answer these questions. Formulated in a proper statement, the focus of this research is: to assess the effects of different interviewing strategies for open interviews on the quality and content of the acquired information.

This topic is discussed by researchers from all epistemological and ontological walks of life. For instance, for so-called ‘positivists,’ the question is whether the interviewer behaviour has led to bias in the deliverance of the information from the mind of the respondent onto the audio recording of the interviewer. In other words, the question is whether the interviewer influences the respondent too much, leading to a problematic validity of the received information. For ‘constructionists’, such as Holstein and Gubrium (1995), the question is slightly different, since according to them there is no such thing as pre-existing information in the mind; all information brought forward in the interview is “constructed in relation to the ongoing communicative contingencies of the interview process” (Holstein & Gubrium, 1995, p. 14). So, the constructionist question is not whether co-construction *has* occurred, but *how* this co-construction has occurred. Notwithstanding these differences, from both ontological perspectives, the question on how the interviewer behaviour influences the received information is urgent and salient. Variation in the received information due to dissimilarities in interviewer behaviour means either that bias is sneaking in, or that different co-constructions between the interviewer and the respondent are taking place.

1.2 A typology of interviews

Since this research concerns the effect of specified interviewer behaviour in open research interviews, it is important to define the interview type I am focussing on: *face-to-face open research* interviews. Since categorisation and definition of any type is always done in contrast to other types, I have to define several types of interviews and show the exact differences between the interviews under study here and other types of interviews.

The first aspect used to distinguish the type of interview in this study from other interview types is the *research* aspect. Research interviews differ from other types of interviews in their main objective. Open research interviews do not need to be entertaining or to contain a scoop as do media interviews, nor is their goal to receive a confession; and, neither is the one being interviewed in the position to acquire a job or to receive therapy. The objective of a research interview lies within only one aspect of the process, which is receiving the highest quality information possible in order to answer a specified research question.

The second aspect that distinguishes the face-to-face open research interview from other types of interviews is the *face-to-face* component. There have been many studies addressing the effects of interview modes, comparing face-to-face interviews with telephone interviews. In a meta-analysis done on the quality of the data by De Leeuw and Van der Zouwen (De Leeuw & Van der Zouwen, 1988), it was shown that there are hardly any differences between the different interview modes used in survey research. For open questions however, there are minor indications in favour of the face-to-face interview, relating to the amount of information received in answer to these questions. This is consistent with the description of Sturges and Hanrahan (2004), who also see no effects of the mode in which qualitative interviews are held.

The third aspect to distinguish the face-to-face open research interview from other interview types is the *open* part. Although the goals of most research interviews are comparable - receiving the highest quality information possible to answer a specified research question- dependent on this research question, different types of interviews are used. Open interviews are often distinguished from other types of interviews by using three dimensions (compare for instance the typology used in Van den Berg, 1996).

The first dimension that distinguishes the interviews concerns the predominant question type used in the interview. In general, differences are made between open questions and closed questions. As Carabain (2007) shows, this difference is not as black and white as is often assumed. Van Den Berg, for instance, uses a dimension based on the information that the question entails for the respondent. He sees three different types of questions: closed questions, semi-open questions and open questions. "*Closed* questions entail information on a) the topic or theme of the question, b) the relevant dimension to be used in answering the question, c) the relevant values on the dimension to be used in choosing an answer" (Van den Berg, 1994, p. 3 (my italics)). *Semi-open* questions involve information on both the topic and the dimension to answer, whereas *open* questions only contain information on the topic to talk about.

Dohrenwend (1965) already showed that looking at syntactical features alone is not enough to differentiate between open and closed questions since seemingly open questions could imply relevant dimensions and relevant values on these dimensions as well.

The second dimension to distinguish open interviews from other interviews is by looking at the amount of structure in the interview. This often results in a distinction among the types: *Structured*, *Semi-structured* and *Unstructured* interviews (Dicicco-Bloom & Crabtree, 2006). These three interview types could be scaled on the dimension of topic control; the more the interviewer is in control of the topics, the more the interview is structured. In structured interviews with a questionnaire controlling for the sequence of the questions on a topic, as well as the approach to the topic, the control is high. On the

other hand, in an unstructured interview, only a topic guide is used, so both the question order as well as the approach to the topic is often not completely decided on by the interviewer alone. Semi-structured is the middle category and is consequentially often used to lump together all forms of interviews between very structured and very unstructured interviews. For instance, when reading the following citation from the lemma 'Semi-structured interviews' in the 'Sage Encyclopedia of Qualitative Research Methods', it is clear that Ayres sees an extreme variety in semi-structured interviews as well.

Researchers who use semi-structured interviewing develop a written interview guide in advance. The interview guide may be very specific, with carefully worded questions, or it may be a list of topics to be covered. The interviewer may follow the guide to the letter, asking the questions in the order they are given, or the researcher may move back and forth through the topic list based on the informant's responses. In either case, the topics of the interview guide are based on the research question and the tentative conceptual model of the phenomenon that underlies the research. (Ayres, 2008)

The third dimension that is often used to differentiate between interview types is the amount of standardisation. Completely standardised interviews are held mainly to be able to compare the results of different interviews, so primarily in cross-sectional research designs. For completely unstandardised interviews, the focus is not comparison, but mostly the uniqueness of a case or story. Therefore, this type of interview is held predominantly in single case studies.

Using this third dimension to distinguish different types of interviews is not very informative, when *not* taking into consideration the different aspects of the interview in which the standardisation can take place. For instance, standardisation could be used for the questionnaire, and for the question wording, but also for general interviewer behaviour or very specific interviewer behaviour, such as standardised reactions to respondent questions.

Very often, these three dimensions are combined, resulting in a simple dichotomy with completely standardised interviews, meaning structured interviews with closed-ended questions and standardised interviewing behaviour, on the one hand, and unstandardised interviews, which is to say unstructured interviews with open ended questions and completely free interviewing behaviour, on the other hand. This dichotomy neglects the huge range of interviews in between.

The last two dimensions (level of structure and amount of standardisation) are very important for distinguishing between different interview types, but as terms themselves they are only precise enough when talking about the extremes.

The interviews under study here are face-to-face open research interviews. In this case, this means that the goal of the interview is to obtain information that is relevant to

the research question, that the interviews contain mostly open questions, are held face-to-face, are partly standardised and are mostly structured.¹

1.3 Interviewer behaviour

The interviewer’s task is to obtain information while listening and encouraging another person to speak (Dicicco-Bloom & Crabtree, 2006, p. 319).

In research concentrating on interviewer behaviour, the behaviour is generally divided in two types of behaviour orientations: task-oriented interviewer behaviour and person-oriented interviewer behaviour. Both interviewer behaviour orientations are part of the interviewer’s presiding role over the conversation.

The two types of interviewer behaviour are distinguishable in both open interviewing and closed interviewing. For instance, for ethnographic interviewing, Spradley phrases it as follows: “Ethnographic interviewing involves two distinct, but complementary processes *developing rapport* and *eliciting information*. Rapport encourages informants to talk about their culture. Eliciting information fosters the development of rapport” (Spradley, 1979, p. 78). On the other side of the interviewing spectrum, for standardised survey interviewing using closed questions, there is a rather large body of literature in which both orientations are distinguished (a.o. Dijkstra, 1983).

1.3.1 Task-oriented interviewer behaviour

1.3.1.1 Task-oriented interviewer behaviour in closed interviews

In an interview the prime goal of an interviewer is usually described as the retrieval of relevant information. This information naturally needs to be relevant, thus leading towards an answer to a prior stated research question (Baarda, De Goede, & Van der Meer-Middelburg, 1996). Therefore task-oriented interviewer behaviour includes all behaviour that is directly aimed at pursuing this information retrieval goal.

The researcher’s goal for holding interviews with closed-ended questions is usually to collect measurements of a sample of respondents in order to estimate characteristics of a population. Since measurement is the main goal of interviews with closed-ended questions, the most important issue at stake for task-oriented interviewer behaviour is to retrieve relevant information that is as reliable, comparable (using standardisation) and as valid as possible. All task-oriented interviewer behaviour in closed-ended interviews, therefore, is aimed at these three objectives.

Some authors believe that all task-oriented behaviour in closed-ended interviews is aimed at retrieving information in a comparable way. Fowler and Mangione display four

principles of standardised interviewing, in order to pursue this goal of comparable information retrieval:

1. Read the questions as written
2. Probe inadequate answers non-directively.
3. Record answers without discretion
4. Be interpersonally non-judgmental regarding the substance of answers (Fowler & Mangione, 1990, p. 35).

However, according to the principles of Fowler and Mangione, a complete task-orientation/standardisation is impossible. Schaeffer and Maynard (2003) show how Fowler and Mangione's strict principles cannot be completely followed by practitioners, since the last three principles are too stringent for a *conversation* with a purpose. Houtkoop-Steenstra (1995) showed that interviewers do not even follow the first principle. In other words, this standardisation cannot be completely fulfilled because sometimes task-orientation aimed at standardisation conflicts with the person-orientation in survey interviewer behaviour.

Van der Zouwen and Smit's work on repair probes of interviewers provides an example study on task-oriented interviewer behaviour in closed interviews. Repair probes are probes used by an interviewer when a respondent's answer does not fit in with one of the answer categories given by the interviewer (as is the prime goal in closed-ended questions). Interviewers usually have to intervene in order to obtain an answer that fits within one of the answer categories. The conclusion they draw in their study is that it is most important that the interviewer takes an initiative to repair inadequate responses; whether the repair is done successfully depends on so many factors that there is no single best way to repair an inadequate answer. (Van der Zouwen & Smit, 2006). This means that standardisation of interviewer behaviour may interfere with task-orientation; sometimes unstandardised interviewer behaviour results in more valid responses.

1.3.1.2 Task-oriented interviewer behaviour in open interviews

The goal for open interviews is often not hypothesis testing and measurement, but exploration, description, discovery and theory generation (Glaser & Strauss, 1967) or theory construction (Charmaz, 2006). Task-oriented behaviour in open interviews is therefore often aimed at receiving *as much, as much specific, as much elaborate and as little ambiguous* information as possible from the respondent. Rather, interviewers' task-orientation is a sine-qua-non for quality information. Rubin and Rubin put it this way: "The depth, details, and richness sought in interviews, what Clifford Geertz (1973) called 'thick description', are rooted in the interviewers' first-hand experiences and form the material that researchers gather and synthesize. To get to this level of detail, depth, and

1. Probing Behaviour in ‘Open’ Interviews

focus, researchers work out *main questions, probes, and follow-ups.*” (Rubin & Rubin, 2005, p. 13).

This sine-qua-non is equally important for positivists and constructionists. Both yearn for thick descriptions: detailed and elaborate information from the interview. So interviewers from all walks of life will need a task-orientation to gather information, let alone gather (or construct) high quality information.

Naturally, in this task-orientation, a large amount of *procedural problems* (Briggs, 1986) can occur and lead to what Briggs calls “communicative blunders”. His study on the effects of certain task-oriented interviewer behaviour on respondent behaviour and the context dependency of these effects is one of the few in-depth analyses of task-oriented interviewer behaviour in research with open interviews.

Bearing in mind the necessity and importance of task-oriented interviewer behaviour in open interviews, it is incredible that there is a lack of methodological studies on the effects of different uses of task-oriented behaviour for open interviews such as probing techniques, initial questions and topic order. This lack of research is probably due to two reasons: a general lack of studies of open interviews anyhow and a focus on the other type of interviewer behaviour, person-oriented interviewer behaviour.

One of the explanations for the first reason, the general lack of studies on interviewer behaviour, is that standardisation of interviewer behaviour is very hard to establish in open interviewing. All respondents differ in their responses to initial questions, so interviewers always have to improvise while fulfilling their tasks. This makes it difficult to prescribe task-oriented behaviour; it often depends on the situation and context which type of interviewer behaviour will lead to more, and more specific, more elaborate, more in-depth and more relevant information. Besides, due to the ideographic nature of most qualitative research, there is a huge variety of interview types.

Box 1.1 Example of the variety of open interviews using the dimensions scope and focus of the interview (Rubin & Rubin, 2005)

	Narrowly focused scope	In-between	Broadly focused scope
Focused mainly on meanings and frameworks	Concept clarification interview	Theory elaboration interview	Ethnographic interview
In-between	Exit interview	A. Oral histories, B. Organisational culture	Life history interview
Focused mainly on events and processes	Investigative interview	A. Action research, B. Evaluation research	Elaborate case studies

Rubin and Rubin (2005), for instance, use two dimensions to describe nine types² of open or qualitative interviews (see Box 1.1). The first dimension is the *scope of the interview*, and the second dimension is the *focus of the interview*. This typology is probably not even near complete, but at least it shows the vast differences between different types of interviews,

making it perfectly comprehensible why it is understudied: where do researchers have to start?³

So, the first reason why task-oriented interviewer behaviour is understudied with respect to open interviews is because interviewer behaviour, in general, is understudied due to the large number of different variables to tackle in such a study.

The second reason why task-oriented interviewer behaviour is understudied is that in most interviewing literature on open interviews, the focus lies on the other type of interviewer behaviour: person-oriented interviewer behaviour.

1.3.2 Person-oriented interviewer behaviour

1.3.2.1 Person-oriented interviewer behaviour in closed interviews

In closed interviews, interviewer behaviour is seen as a potential source of distortions that leads to a reduction of the comparability of answers from different respondents. The discussion about the importance or desirability of person-orientated interviewer behaviour during interviews partly parallels the discussion on standardisation (Beatty, 1995). Although it is possible to standardise the task-oriented interviewer behaviour, person-oriented interviewer behaviour is much harder to standardise. Therefore, person-oriented interviewer behaviour is often seen as a threat to standardisation. A strict task-oriented interviewer stance combined with a strict standardisation is often seen as the core of survey research. Jocelyn Viterna and Douglas Maynard explain this in their opening sentence: “The quality of survey data relies heavily upon standardization in the survey interview” (2002, p. 365).

In other words, in this view, the personal role of the interviewer needs to be minimised. Interviewer behaviour that is aimed at creating the *functional personal relationship*, called rapport, is seen as potentially distortive. Therefore, interviewers should not try too desperately to create a relationship with the respondent. Rather, they would do better to focus on the minimisation of the effects that their behaviour has on the data (Fowler & Mangione, 1990).

In their handbook, Hyman and Cobb, therefore, warn for the possible dangers of rapport. The danger exists in the possibility that one interviewer has greater ability to establish a personal relationship of trust than another interviewer, due to personal features. Therefore, the inter-interviewer variation will strongly increase, which in turn leads to different measurement and, thus, a strongly decreased reliability of the acquired information (Hyman & Cobb, 1975 [1954], p. 257).

Due to the poor univocality of the definition of the concept and the difficult operationalisation, Weiss (1968, 1970) and Goudy and Potter (1975), therefore, warn for the use of the term rapport. These authors suggest that researchers should not focus on

the relation but on more specific concepts and more specific behaviour, such as types of person-oriented interviewer behaviour. Twenty years later, Beatty (1995) sees this mid-seventies demise of the concept as leading towards an over-standardisation of survey interviewing. This is crowned by the work of Fowler and Mangione, who see rapport in closed interviews as inevitably necessary, but only within the function of reciprocity towards the respondent, and within the strict set of principles presented in section 1.3.1.1. If rapport is used more than as strictly inevitable, standardisation is at risk, and as a consequence, the influence of the interviewer is disproportionately high (Fowler & Mangione, 1990).

Methodological research on interviewing style in closed interviews

One of the researchers who focused on more specific and thus more easily measurable concepts than rapport is Dijkstra. (Dijkstra, 1983, 1987; Van der Zouwen, Dijkstra, & Smit, 1991). Dijkstra did an experiment in which he compared the effects that different interviewing styles, namely a formal and a personal interviewing style, had on data. The interviewing styles can be seen as determinants for rapport, as one can argue that by using the personal interviewing style, an interviewer can more easily create rapport than by using a formal style. The operationalisation and instruction of these interviewing styles is much less complicated than rapport because it relates to the behaviour of a single person instead of a conversational relationship between two people.

In both interviewing styles, the interview was standardised. In the personal, socio-emotional style, eight interviewers were told that respondents are only prepared to give accurate and sufficient information if a personal tie exists between the interviewer and the respondent. The interviewers were expected to create a personal relation: rapport. Dijkstra instructed the interviewers using three rules. The first rule was that the interviewer had to react empathically, especially when a respondent uttered feelings or spoke about personal experiences. The second was that at appropriate times the interviewers had to share some personal stories or experiences. And the third rule was that the interviewer should chat with the respondents over coffee on topics like hobbies or the weather (Dijkstra, 1983, pp. 44, 54).

The eight interviewers that used the formal interviewing style were instructed not to deviate from the primary task of data collection. The explanation the interviewers received was that the level of rapport with the respondent is directly and negatively related to the quality of the data because the high level of rapport over-influences respondents and leads to inaccurate answers. The interviewers only had to work on the relational aspects, when doing otherwise would have been rude and unfriendly towards the respondent.

The results from this study were that the personal interviewing style had a positive effect on the accuracy and the amount of information given by the respondent in the interview. In addition, more personal information was shared and less social-desirable answers were given (Dijkstra, 1983, pp. 95, 123). The results of this study did not only contradict the expectations of Dijkstra himself, but also the basic assumption of the daily practice of many survey-research institutes and the general opinions of many survey researchers⁴.

However, later reanalysis of the data by Dijkstra and Van der Zouwen led to the conclusion that interviewers using a personal interviewing style show more inadequate behaviour than interviewers using a formal interviewing style. The inadequate behaviour mainly consists of deviating from the interview topic, posing leading questions and suggesting answer options. (Dijkstra & Van der Zouwen, 1988).

1.3.2.2 Person-oriented interviewer behaviour in open interviews

As described in section 1.3.1.2, in general, interviewer behaviour in open interviews is understudied, and the studies that have been published are mainly on person-oriented interviewer behaviour. However, the focal point differs tremendously from the discussion in closed interviews. The discussion in closed interviews parallels discussions on standardisation, whereas the discussion in open interviews is dominated by ethics.

One of the most influential sources on interviewer behaviour in open interviews is the publication by Ann Oakley (1981), 'Interviewing Women: a contradiction in Terms', especially when using it to distinguish open interviews from closed interviews. In this publication, Oakley poses the view that the hierarchical, objectifying and falsely 'objective' stance of the neutral, impersonal interviewer is impossible, as well as unacceptable. When, as feminist researcher, you would want to interview women meaningfully, you should depend on empathy and mutuality. To accomplish this mutuality, the interviewer ought to approach the respondent with genuine sympathy and share personal information, answer questions and even discuss opinions with the respondent.

This view of person-oriented interviewer behaviour in open interviews has had a large influence outside feminist interviewing as well. Almost all authors that discuss interviewer behaviour in their work do this in relation to power issues. Mishler (1986) on the one hand, and Briggs (2003) and Kvale (2006), on the other hand, stress the inequality and the dominance in interviews as well. According to Mishler (1986), the only solution for this inequality is that the respondents are actively involved in the interview as well as in the analysis afterwards through forms of member validation. He therefore follows the solution of Oakley, a genuine interviewing relation and a true dialogue.

According to Kvale (2006), this idea of a true dialogue in interviews is a myth. Following other critiques of Oakley, he rejects the possibility of true dialogues based on

some form of similarity in power relations in open interviews. However, this does not mean that knowledge gained through interviews is invalid: "The use of power in interviews to produce knowledge is a valuable and legitimate way of conducting research. With interview knowledge jointly constructed by interviewer and interviewee, overlooking the complex power dynamics of the social construction process may, however, seriously impair the validity of the knowledge constructed." (Kvale, 2006, pp. 485-486).

Following Kvale, I would also say that no interview is without power differences. However, Kvale nuances the absoluteness of interviewer power by describing a few possibilities for respondents to counter this power, yet I believe that interviewers can be rather powerless in interviews as well; for some good examples, see the research of Bravo-Moreno (2003).

How to deal with interviews with 'repugnant others' is another issue with the ethical stringent view of person-oriented interviewer behaviour as only being possible through true mutuality (Springwood & King, 2001). How does one interview a serial killer, when this is only truly possible through some form of friendship? Do we, therefore, only have to interview people we like and who are to be emancipated? Or, do we leave this quasi-universalistic ethics and simply revert to our roles and impression management as we do in everyday life (Goffman, 1959)?

Besides the ethical argumentation used by Oakley, the assumption in her 1981 publication is that it is only through this equal relation that a respondent will open up. The radical form of person-oriented interviewer behaviour is therefore partly task-oriented because it will lead to more in-depth and truthful information. So, while the explicit goal is the moral and personal genuineness of interviewer behaviour, the implicit goal is more instrumental.

Abell *et al.* (2006) analyse one of the prescribed behaviours for building a genuine relation in open interviews, which is self-disclosure. Through self-disclosure, interviewers, rather instrumentally, intend to create similarity and mutuality. Abell *et al.* show that interviewers not only can fail to create similarity through self-disclosure, but also can end up with a substantiation of the differences between the interviewer and the respondent. This recommended form of behaviour that intends to create a genuine relation can sometimes work counterproductively.

So, the instrumentality of person-oriented interviewing behaviour is not only unavoidable, it is also traceable in research with friendship-based interviewing behaviour.

Methodological research on interviewing style in open interviews

In comparison to the interviewer behaviour described in the section above, researchers acknowledging the aspect of role-playing in interviewing could have studied a subset of person-oriented interviewer behaviour: interviewing style. This is more productive than

Oakley's view of interviewers as the respondents' true friends because by varying the behaviour, a researcher can analyse the effects of person-oriented interviewer behaviour on the quality of the information. Although most interviewers realise that part of what they do is role-playing, it is intriguing how little is known about effects of different interviewing styles in open interviews.

Insufficient research has been done on interviewing styles in open interviews, and the research that has been undertaken remains flawed in its design (e.g. Van der Drift & Derksen, 1985). However, since theirs is one of the few studies done, I discuss it here briefly. In their article, Van der Drift and Derksen distinguish between three types of interviewing styles:

1. Formal interviewing style: (2 interviewers) in this style, the interviewer literally asks the questions as formulated by the researchers. The interviewer sticks to the question sequence. The interviewer minimises encouragement of the respondents by withholding empathic remarks. Finally the interviewer does not deviate from the research topics.
2. Empathic interviewing style: (3 interviewers) in this style, the interviewer formulates the questions, depending on the understanding of the respondent. The question sequence is altered to the 'needs' of respondents. Interviewers working with this style are deviating from the research topics. In this style, the respondents are encouraged by empathic remarks.
3. Intimate interviewing style: (1 interviewer) in this style, the interviewer literally asks the questions as formulated by the researchers. The interviewer sticks to the question sequence. The interviewer encourages the respondents by empathic remarks and showing his or her own opinions (Van der Drift & Derksen, 1985).

The third style presented here is rather awkward and incomparable, since self-disclosure is also used as a method. We know from the above-mentioned research done by Abell *et al.* (2006) that self-disclosure can be very counterproductive for interviewing. The other two styles seem sensible at first, but upon taking a closer look, the styles differ in regards to question sequence structure, question formulation, off-topic deviation and empathic remarks. This combined with the low N of interviewers makes the results of the research less valuable.

Remarkably, in handbooks on open interviewing, there is a rather high level of agreement on the most successful interviewing style. In general, it is established that it is only possible for respondents to 'open' up their 'true' feelings, experiences, meanings and opinions, when they are interviewed in a friendly and personal interviewing style (Baarda, et al., 1996; Emans, 1990; Evers & De Boer, 2007b; Gorden, 1992; Kvale, 1996; Rubin & Rubin, 2005; Seidman, 2006; Weiss, 1994). Or, as Fontana and Frey state it: "Because the

goal of unstructured interviewing is *understanding*, it becomes paramount for the researcher to establish rapport” (Fontana & Frey, 1998, p. 60). In general, researchers using open interviews agree on the notion that a high level of rapport leads to a better quality of information that is received.

Interviewers can develop and sustain rapport by treating the respondents with understanding, showing their interest and attention, smiling and flattering respondents. All these interviewer behaviours are directed towards the respondent as a person and not towards the content or the quality of the content of the information. In line with this argumentation, most interviewers doing open interviews will make use of a more personal interviewing style. I define a personal interviewing style as an interviewer stance -and the behaviour following from this stance-, which is sustained throughout the interview and is aimed at creating and sustaining a *personal* relation between the respondent and the interviewer. In the interviews held in this research, the line of reasoning of the general literature on interviewing is followed, and hence, a *personal interviewing style* will be used.

1.3.3 The relation between the two different orientations

After the above descriptions of task-oriented and person-oriented interviewer behaviour, one could wonder whether it is always exclusively one of the two possible interviewer orientations. The answer is: “probably not”. The distinction between task-oriented and person-oriented interviewer behaviour is a useful and effective analytical distinction that is used to train interviewers and to develop awareness on the possible effects of their behaviour.

During the interviews, respondents never know why the interviewer behaves in the way he or she does. Most respondents would probably not wonder whether the interviewer behaved person-oriented or task-oriented. Interviewers are naturally much more aware of the purpose of their own behaviour. However, after an interview, many interviewers cannot explain why they interviewed in a certain manner. Most often, they just did.

This is generally not bad interviewer behaviour; it is probably due to the fact that most interviewer behaviour during an interview is task-oriented and person-oriented *at the same time*. Interviewers *can* strive for both goals while posing a question. In his discussion on the establishment of rapport in ethnographic interviews, Spradley (see citation in section 1.3) shows not only person-oriented techniques to create rapport, but mostly task-oriented techniques instead, such as asking descriptive questions, making repeated explanations, restatements (in the sense of paraphrases) and questions on use of terms, rather than on meaning. Full rapport is reached when informants not only *share the definition of the interview* but also start participating actively in the research. In other words, when rapport has been successful, indirectly the goal of task-oriented interviewer

behaviour, to receive quality information, is also reached (Spradley, 1979). However, in 1969, Gorden already warned us of the flaws in reasoning that rapport automatically means task-orientation:

“Often the neophyte thinks he has conducted an excellent interview because “rapport was perfect” and the respondent was “completely at ease, talked spontaneously, and commented that she had enjoyed the interview”. Yet when the interview is analyzed for the amount and clarity of relevant data, it is found to be incomplete, superficial and ambiguous.” (Gorden, 1969, p. 69)

In subsection 1.3.1.1 on task-oriented interviewer behaviour in closed interviews, probing was presented as task-oriented interviewer behaviour. Probing in survey interviews is mostly for repairing answers that did not fit within one of the stated answer categories. By pointing at failures of the respondent to meet the survey criteria, this repair is, at times, less person-oriented.

Since response alternatives in open interviews are not scripted in advance, all respondents will answer quite differently. So, after the initial question, the interviewer is dependent on the answer of the respondent for the course of the interaction. Only if the respondent answers adequately, full and candid at once, the interviewer can acknowledge and pose the next question. However, this rarely occurs, so interviewers will have to react on incomplete, ambiguous, superficial, irrelevant or very verbose answers. The goal of these reactions is to probe the respondent for more, more precise, more elaborate, more in-depth and/or more unambiguous information. Consequently, one could argue that this is task-oriented interviewer behaviour. In itself, those probes probably are task-oriented, but in the context and the form in which they are uttered they can be equally person-oriented. Therefore, whereas the distinction is very simple for interview styles and the technical management of the conversation, for probing techniques and tactics, it is still useful, but more complicated.

1.4 Probing techniques for open interviews

When reading the literature on probing in open interviews, good probing is usually seen as decisive for good interviewing. However, what exactly is included in *good* probing and what *good* probing entails could be rather difficult to establish. Since interviewers need to improvise on unpredictable initial answers from respondents, “the successful interviewer must have mastered a wide range of specific techniques so that their use is automatic before he attempts the more complex task of adjusting the pattern of questioning to the ongoing context of the interviews” (Gorden, 1980, p. 275). Seidman addresses the issue a little differently in his chapter *Technique isn't everything, but it is a lot*, but the conclusion is identical: “It is tempting to say that interviewing is an art, a reflection of the personality of the interviewer and cannot be taught. This line of thinking implies that you are either

good at it or you are not. But that is only half true. Researchers can learn techniques and skills of interviewing” (2006, p. 78). In other words, an interviewer needs to be prepared to interview in a range of available techniques.

Gorden and Seidman both use a rather broad idea of techniques: as specific forms of verbal and nonverbal behaviour used during the *entire* interview. This means questions, follow-up questions, prompts and probes. Emans (1990) uses the more specific term probing techniques for all behaviour, such as prompts, probes and follow-up questions, after the initial scripted questions. Other authors, such as Rubin and Rubin (2005) and Evers and De Boer (2007b) use a distinction between main questions, probes and follow-up questions. The main questions are prepared prior to the interview. The probes are used to get more, more specific or clearer answers. Follow-up questions are completely new questions induced by the respondent’s previous answer. The difference between probing and following up, however, is not always clear. In unstructured interviews, this difference is less relevant, while in semi-structured or structured interviews, it depends on how much deviation from the original topic is allowed. In this research, I stick to Emans’ use of probing as all interviewer behaviour after the initial question, with the purpose to get more, more specific, more elaborate, more in-depth, more relevant and less ambiguous information from the respondent⁵.

The *probing techniques* themselves are also defined and categorised differently. Wengraf (2001), Rubin and Rubin (2005), Kvale (1996), (Kvale & Brinkmann, 2008) and in some cases Gorden (1980) use the *content* or the *function* of the probe to distinguish between different types of probing techniques. In this research I distinguish different probing techniques based on the *format* of the probe, rather than the function or the purpose. In the subsection below, I present the different probing techniques.

1.4.1 Types of probing techniques

The first probing technique is what Gorden (1992) calls ‘active silence’. Active silence is the most non-directive technique existing, since it allows the respondent “to proceed in whatever direction is most interesting or meaningful” (Gorden, 1992, p. 149). The technique is generally considered to be very effective for obviously incomplete answers. By remaining silent after the respondent finishes an answer, the interviewer shows that the respondent has not yet finished. Emans also suggests that verbal probing techniques draw the attention towards the interviewer, while active silence offers the possibility for a respondent to think and respond without distraction (Emans, 1990).

Box 1.2 Probing techniques distinguished in this study

Active silence	Active silence
Minimal probes /continuers	2.1 Humming 2.2 Echoing 2.3 Comment
Unfinished question / sentence	Unfinished question / sentence
Question Repetition or Reformulation	4.1 Question Repetition 4.2 Question Reformulation
Request for Elaboration	5.1 Request for Elaboration a. Directive b. Non-directive
Request for Specification	6.1 Request for Specification 6.2 Request for Specification by Example 6.3 Request for Specification by Contrast 6.4 Request for Specification of Own Opinion a. Directive b. Non-directive
Follow-up question	7.1 Follow-up question 7.2 Request for Reasoning 7.3 Request for Experiences 7.4 Request for Feelings a. Directive b. Non-directive
Expressions	8.1 Expression of Doubt 8.2 Expression of (Lack of) Comprehension 8.3 Expression of Perceived Inconsistency
Paraphrasing or summarising	Paraphrasing or summarising
Reflection	Reflection

The first audible minimal probing technique or continuer is humming. This humming is reacting using with what Gorden calls “non-verbal noises” (Gorden, 1980, p. 372). Saying hm-hm or mh-mh is a little more directive but exhibits more attention to the respondent’s story than active silence. The purpose of this technique is therefore not only to get more, more specific, more elaborate and more in-depth information from the respondent but also to show attention. Additionally, the intonation and melody of the hum can lead to a range of different functions for this probing technique.

Echoing, the repetition of one or a few words from the respondent, is the second audible minimal probing technique. As a probing technique, it is less direct than a full request. However, by selecting words for the respondent to repeat, the interviewer conveys more meaning and direction than by using ‘active silence’ or humming. As is the

case with humming, the intonation and melody is very important for the exact meaning of this repetition.

As a third minimal probing technique, commenting is often overlooked in interviewing literature. This is strange since in 1982, Snow *et al.* (1982) fiercely defended using comments instead of direct questions. A comment can work tremendously well to either encourage or challenge a respondent to continue, reconsider or explicate a previous answer. Examples of the commenting probing technique are “oh yeah?” “interesting” or “really?”. A comment is comparable to an echo in the sense that it derives its meaning from the context in which the comment is used and how it is uttered. Gorden (1969, 1980, 1992) calls the three minimal probing techniques mentioned above *encouragement probes*.

Probing technique number three, the unfinished sentence/question, is not a technique that is mentioned in probing literature. In its formulation, it lies between a minimal probe and a full request. The interviewer slowly starts uttering a question such as “but, what exactly...” and then simply stops talking. Another example of half a sentence is “so you mean...”. As a probing technique, the unfinished sentence/question is still rather non-directive, or in Gorden’s words, the topic-control is rather low (1980, 1992).

The fourth technique is a repetition or reformulation of the initial question. This probing technique can function as an elicitation technique, but it is often used to control the topic and keep the respondent on track or as a ‘subtle’ clarification of the intended question. Some authors, such as Evers and De Boer (2007b), distinguish this technique from a clarification of the initial question. I see them simply as two versions of this kind: a repeated version or a reformulated version. Whether it functions as a clarification depends on the respondent’s interpretation of the question (Emans, 1990).

A request for elaboration is a probing technique, in which the respondent is *asked* to elaborate on an earlier given response. Oksenberg, Cannell, and Kalton, call these probes simply “tell me more” probes (Oksenberg, Cannell, & Kalton, 1991). Gorden makes a distinction between immediate elaborations and retrospective elaborations, in which the first is a request over the preceding respondent turns, whereas the latter is a request to the respondent to elaborate on something said earlier in the interview (Gorden, 1992). A more important differentiation is the one proposed by Emans (1990, p. 79), which is whether the respondent uses the probe in a directive or non-directive way. Hence, by naming *what exactly* to elaborate on, or simply asking “Could you tell me more?”, the interviewer either controls the topic, or lets the respondent ponder by him- or herself.

The request for specification and its differentiation in a directive or non-directive way, is comparable to the request for elaboration. The difference is that instead of asking respondents to elaborate the interviewer asks the respondent to specify. Request for specification by example is basically requesting the respondent to come up with an

example. In the request for specification by contrast the interviewer uses a technique suggested by Spradley (1979), in which the interviewer requests for a comparison of two terms previously used by the respondent. In the third request for specification, the request for specification of own opinion, the respondent is asked to specify whether an opinion previously presented is his or her *own* opinion, and if not what the own opinion is.

The follow-up question can be used in four different versions. The first version is a 'normal' follow-up question, a probe in which a new (sub-)topic is introduced. By definition this probe is directive. The other versions of follow-up questions are more specific requests: a request for reasoning or explanation, (sometimes in the form of a simple 'why?'), a request for experiences and a request for feelings. Price (2002) uses three comparable follow-up requests in a combined set he calls: 'the ladder question technique'. He starts probing with requests for experiences and goes on to probe on knowledge, beliefs and feelings. In this research the three request follow-up questions are not used in a separate combined set. Like the other requests, these requests could be formulated in a directive or non-directive way.

The eighth set of probing techniques consists of three types of expressions: expression of doubt, expression of (lack of) comprehension and expression of perceived inconsistency. As probing techniques, they are somewhat comparable to comments, although they are more directive and more explicitly praising or critical. The expression of comprehension could also be phrased as a question, often introducing some form of a summary.

The difference between paraphrasing and summarising is largely dependent on the length and the verbosity of the response from the respondent. Therefore, both probing techniques are taken together. Baarda *et al.* (1996) see five different functions of a summary. The summary a) structures the responses, b) functions as a probe, when information is lacking or unclear, c) can keep the respondent on topic, d) can increase rapport, since the interviewer shows attention and e) can function as a topic terminator or a bridge to a next question or topic.

Reflection is the final probing technique discussed here. Kvale calls this probing technique interpreting questions (1996). In the way I use the term reflection, it is not necessarily a question; some combination of a summary and a request or a positive or negative implication of a response are equally possible. In a reflection, the interviewer could point the respondent towards similarities, differences and consequences from a response with something previously mentioned by the respondent. The difference with the expressions is that the reflections often include some element of a summary and are more complicated than the expressions.

The reflection is the most difficult probing technique. The difficulty in reflecting as a probing technique, rather than suggesting an answer, lies in three aspects: the interpretation of the response, the possible relation to earlier responses (in the sense of consequences, similarities or inconsistencies) and the formulation of this interpretation and this possible relation. If one of these aspects is wrong, the reflection quickly turns into a suggestion.

1.5 Probing tactics

Since probing is so important in open interviews, one would expect that a large amount of information on the effects of different probing techniques and probing behaviours would be available. However, due to the lack of studies on interviewer behaviour in open interviews, this is unfortunately not the case. In most literature on open research interviews, authors fail to notice the possibility that besides the interviewing style, interviewers have a huge range of types of behaviours at their disposal while probing. This is strange, especially in the light of the remarks made in section 1.4, that probing is *the* most important interviewer behaviour in open interviews. So if one needs to choose a characteristic element of the open interview from among all sorts of interviewer behaviour types, probing is the most salient. Therefore, in this research, I chose not to analyse interviewing styles, but to focus on the interviewer behaviour after posing the initial questions, thus the probing behaviour.

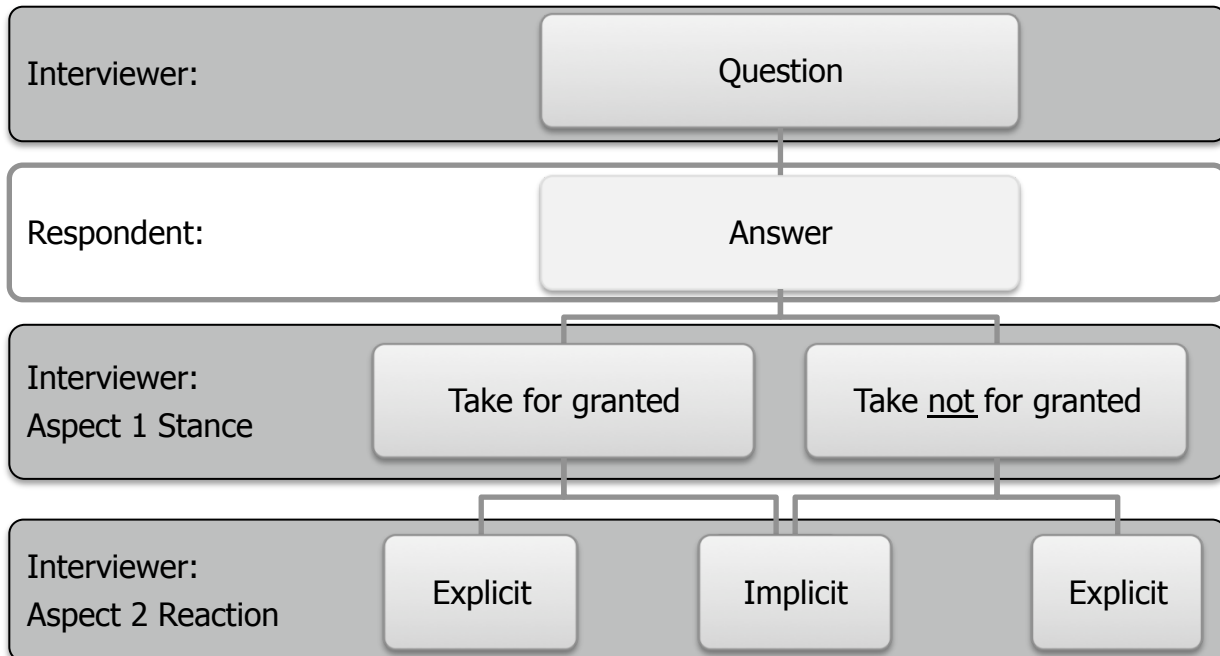
While using probing techniques, interviewers often base their reactions not only on the content of the responses, but also from their stance of initial acceptance of the answer. Often this stance leads to the use of a specific set of related probing techniques, or at least to ways in which these probing techniques are used. So, the choice of a certain probing technique is based on both the evaluation of an answer in terms of the interview goals, and a stance towards the answer of the respondent with a reaction that follows from that stance. That is what I call a *probing tactic*.

1.5.1 What are probing tactics?

A probing tactic is an interviewer *stance towards the responses* of the respondent **and** *the reaction following* from this stance. When a respondent answers an initial question, an interviewer generally has two possibilities: either to take the answer for granted or not. In itself, taking an answer for granted is not directly a probing tactic, since it is only the first aspect: a *stance* towards the response. The second aspect is *the reaction following* from this stance. An interviewer can choose either to show his or her stance by using an *explicit* role (but subtly), or to not give the respondent any idea about his or her stance.⁶ When the role is *implicit*, it is not important what the interviewer stance is, since this lies all in the

interpretation of the respondent. In Figure 1.1, the interviewer's and respondent's behaviour are schematically shown, followed by the two aspects of the interviewer's probing tactics.

Figure 1.1 Schematic Representation of the Two Aspects of the Probing Tactics: Stance and Reaction



The probing tactics are called probing tactics and not probing *stance* to show both the explicit or implicit role of the interviewer (aspect 2) and the instrumental nature of the probing tactics. As was argued in subsection 1.3.2.2, just as in everyday life, interviewing is full of tactical role-playing. For instance, when an interviewer does not accept the answer at all, he or she can still act as if the answer is perfect and stimulate the respondent to continue talking. Naturally, the opposite is possible as well. As tactics are generally a part of a strategy, within this study, the interview strategy contains a personal interviewing style combined with a probing tactic.

In itself, a probing technique is not an aspect of a probing tactic. An interviewer could use a probe for more specific information and combine it with the two aspects of the probing tactic. Just as actors can put very different Hamlets on stage, even though all actors are using the same lines, interviewers can use the same probing techniques and sometimes even the same words, while conveying different messages through the probing tactic. An echo can be used to channel acceptance or doubtfulness of the answer to the respondent. In other words, a technique can function as a sign vehicle for any probing tactic. However, as the lines and verses of Shakespeare mark some boundary towards the actor's performance of Hamlet, so do some probing techniques work better to communicate or not to communicate a certain stance.

In most interviews, it is the interviewer who decides whether or not to take answers for granted and whether or not to communicate this to the respondent. After every answer of from respondent, it is typical that an interviewer already has to make too many decisions: on the meaning of the answer, on the quality of the answer, on further probing or not, on what probing technique to choose, on the formulation of the probe, on the person-orientation and the task-orientation in his or her behaviour and probably on many more aspects. A choice of a probing tactic is one of many decisions an interviewer needs to make. Therefore, I believe that most interviewers keep this decision rather stable within one interview and only when an answer strikes them, they act differently and use another stance and reaction. Margaret Wetherell, for instance, proposes to challenge respondents (only) when they express views that the researcher finds offensive (2003).

Since probing tactics are often neglected in interviewing literature, there is hardly any evidence whether reacting differently on answers that seem remarkable to the researcher is beneficial or not. We do not even know what the effects of keeping a probing tactic stable are, nor do we know anything about the effects of these different tactics.

1.5.2 Probing tactics versus interviewing style

Since the distinction between probing tactics and interviewing style is new in this research, I need to discuss the differences between both types of interviewer behaviours. Probing tactics can be distinguished from interviewing styles in two general ways.

The first difference between probing tactics and interviewing styles concerns the different purposes. The purpose of interviewing styles is to create a positive atmosphere, to create a good conversational ground with the respondent as a person and a good task-orientation of the respondent. On the other hand, the purpose of the probing tactic is more specific: to elicit better, to get more, more specific, more elaborate and more in-depth information from the respondent. Consequentially, a probing tactic, in contrast to interviewing style, is explicitly aimed at the *quality of the content*. However, it would be rather naïve to believe that respondents can distinguish between feedback on them as individuals, on the content in itself or on the quality of the content. And as was already shown in section 1.3.3, a probing technique in itself could be both person-oriented and task-oriented. Nevertheless, the prime purposes and the orientation of probing tactics and interviewing styles are different.

The second difference between interviewing styles and probing tactics is their place in the interview. A probing tactic is a stance towards the quality of the content of a given response and a reaction that follows from that stance, *while probing, and using specified probing techniques*. In contrast, the interviewing style is used throughout the entire interview, so

from the introduction until the interview is completed, throughout questioning, probing and chatting.

1.5.3 The different probing tactics and their respective logics

In this research I distinguish three different probing tactics: accommodating, encouraging and challenging. Naturally, these three tactics are identical to the three outcomes from aspect 1 and 2 presented in Figure 1.1: the stance and the reaction following from that stance.

Besides being the logical outcome of the two stances and the explicit or implicit reactions following these stances, these three different probing tactics are logically comparable to suggestions and reflections from other authors on interviewing. In this literature on how to interview or sometimes even how to probe, one can find different logics for why these probing tactics would result in a higher quality of the information received by means of the interview.

1.5.3.1 Accommodating probing tactic: implicitly (not-)taking for granted

The logic behind the accommodating probing tactic could be found in two very opposite fields of interviewing: in narrative interviewing and in closed interviewing.

In the debate on standardisation of survey interviewing, the most important issue is comparability. Therefore interviewer behaviour is standardised and minimised if possible. Fowler and Mangione prescribe the fourth principle of standardisation: “Be interpersonally non-judgmental regarding the substance of answers”(Fowler & Mangione, 1990, p. 35). In other words, in standardised interviewing the answer of the respondent is taken for granted, or even if it is not, the reaction is implicit. When translating this to probing in open interviewing, it means that the interviewer should refrain from disturbing the respondent in answering the questions as much as possible. All probes, therefore, should be uttered non-judgementally and neutrally.

In some forms of narrative interviewing, the interviewer should also refrain from being judgmental. The logic behind this is to reduce the influence on the respondent and to give the individual the freedom to continue- for instance, a life story- without being disturbed by interviewer questions. In narrative analysis, Riessman (2003) for example, explains that the best storytelling occurs when the interviewer can withdraw from the interview and thus give space to the storyteller. Kvale (1996) also seems to follow this line of reasoning, since he thinks that the less the interviewer says and the more the respondent speaks, the better. When doing multiple narrative interviews, Wengraf advises not to ask any questions during the first narrative session. “You are a Story-facilitator” (Wengraf, 2001, p. 122); “You are helping the informant uncover the life-history that is

relevant to him or her, helping the interviewee to follow their own 'systems of relevancy'" (Wengraf, 2001, p. 124).

The accommodating probing tactic can be defined as a tactic in which interviewers withhold evaluative reactions while probing, in order to give the respondent freedom to answer along the individual's chosen path.

Probing techniques that seem to be typical for an accommodating tactic are the three minimal techniques: active silence, humming and echoes and, by definition, the rather neutral question repetition. It is also typical for summaries to be used relatively often by accommodating interviewers, since in a good summary no positive or negative evaluation (or explicit stance) is shown. Comments and expressions are rather rare within the accommodating tactic; although a question of comprehension such as "Did I get it correctly?" is possible. Almost all other techniques presented in 1.4.1 can be used to accommodate the respondent. An example of a request for elaboration probing technique to be uttered when using the accommodating probing tactic is: "Could you tell me more about that?"

The accommodating probing tactic is successfully performed if the interviewer is able to convey the message of accommodation: "You can tell me anything, I will never judge and I give you full freedom to deliver your story, as relevant, complete, detailed, elaborate and in-depth as possible".

1.5.3.2 Encouraging probing tactic: explicitly taking for granted

The second probing tactic used in this research is the encouraging probing tactic. The logic behind the encouraging probing tactic parallels the logic of what is often called empathic interviewing in qualitative interviewing literature. In empathic interviewing, the interviewer does show emotions and uses these emotions to support the respondent in telling his or her story. This empathic interviewing shares some characteristics with the interviewing prescribed by Ann Oakley (1981) and discussed in section 1.3.2.2. Oakley's prescription emphasizes that the interviewer should not only be friendly to the respondent, but also sympathetic, equal, reciprocal and truly interested in the respondents and their stories.

Although the description of this empathic interviewing is most often rather vague and usually more moral than methodological, many scholars interpreted it in a Rogerian way: the respondent should always be approached with an unconditional positive attitude (Rogers, 1951). An interesting example of the notion of empathic interviewing is presented in Chirban's book on the interactive-relational approach. The central thesis of this book is that the interview is based on a relation, created in the interaction through the posture of both the interviewer and the respondent. Since the focus lies solely on the relation, it lacks any links to more task-oriented interviewer behaviour such as probing

techniques (Chirban, 1996). So, it seems that an interviewer only needs to be empathic to successfully conduct an in-depth interview.

Although most authors of 'How-to-interview' books are silent or not strongly in favour of this way of interviewing, and focus on techniques as well, I believe that in practice, most interviewers follow the encouraging logic. They probe encouragingly in order to try to elicit information, while creating rapport at the same time. So they use it as some sort of natural extension to the personal interviewing style. Moreover, many interviewers believe in the reciprocity principle that Dijkstra (1983) used as an explanation of the success of the personal interviewing style: in a personal interviewing style, the interviewer motivates the respondent by being friendly and interested and gets higher quality information in exchange.

Translated to my research this behaviour is operationalised in an encouraging probing tactic. This means that the interviewer takes the answer of the respondent for granted and by uttering positive expressions or comments on the quality of the answer, encourages the respondent to continue answering. In the encouraging probing tactic, in opposition to many empathic interviews, suggestive interviewing and self-disclosure of the interviewer are not allowed. As is the case with all probing tactics, the probing behaviour is solely focussed on the content of the answers and as a part of the interviewers' probing behaviour.

Typical probing techniques for the encouraging probing tactic include positive comments, the expression of comprehension and a reflection on the respondent's answer following the path taken. Naturally, humming and other probing techniques could be intonated such that the conveyed meaning is encouragement: "You can tell me anything, I am unconditionally interested and encourage you to deliver your story, as relevant, complete, detailed, elaborate and in-depth as possible".

1.5.3.3 Challenging probing tactic: explicitly not-taking for granted

The third tactic derived from Figure 1.1 is a probing tactic in which the answer is not automatically taken for granted, and this is subtly shown to the respondent. This interviewer behaviour is rarely applied in most research interviews. Interviewers are commonly afraid to subtly counter a respondent for fear of suggesting answers or falling of the high rope of rapport.

The logic behind this tactic, however, is not at all uncommon to us; in movies we watch the bad cop taking turns with the good cop, and we are keen on journalists critically interviewing politicians. In this way of interviewing the interviewer critically probes the suspect or the politician to tell the complete truth. By being critical, the interviewer not only shows his or her interest and processing of the information that is

brought forward, but also critically asks for clarification, detail and elaboration and even points out possible inconsistencies within the answers.

In research interviewing literature, some authors (Weiss, 1994; Kvale, 2006; Wengraf, 2001) do discuss what they call the antagonistic interview. The argument is that in addition to the warm personal and consensus-seeking research interviews, there are alternative conceptions of the interview that are possible. In these cases, the antagonism is not only shown during probing, but throughout the entire interview, as part of a more general interviewer behaviour or role. Both Weiss (1994) and Wengraf (2001) are quick to admit that they prefer to work from a personal interviewing style. "It depends on a very considerable amount of rapport or, more usually, a very considerable power-over. It will feel potentially threatening and controlling. It is unlikely to leave the interviewee feeling good." (Wengraf, 2001, p. 155).

Kvale sees more possibilities. Among the six different conceptions of what Kvale calls "agonistic interview alternatives" (2006: 486), there is one conception that comes close to the challenging probing tactic as I use it. Kvale calls it "actively confronting interviews". "Actively confronting interviews do not necessarily aim for agreement between interviewer and interviewee, as the interviewer critically questions what the interviewee says, for example, if he contradicts himself." (Kvale, 2006: 487). As we will see in the challenging probing tactic, what is important in this type of interview is that the interviewer does not impose any ideas on the respondent. The other alternatives Kvale mentions are rather indifferent to interviewer influence or are in favour of purposely suggesting ideas to invoke discourses crossing swords (Holstein & Gubrium, 1995; e.g. Brinkmann, 2007; Tanggaard, 2007).

Quite some authors have published on the possibility of using more challenging ways of interviewing although these are seldom employed. In his PhD, study Howard Becker interviewed Chicago public school teachers. As part of his challenging tactic he used two sets of behaviours. After some time passed in the course of the interview, Becker assumed a "skeptical air" (Becker, 1954, p. 31) and probed previous statements for evidence from the respondents' own experiences. This challenged the respondent to "put up or shut up" (Becker, 1954, p. 31). Another challenge Becker used was playing dumb and pretending not to understand implicit attitudes. According to Becker, he "was able to coerce many interviewees into being considerably more frank than they had originally intended" (Becker, 1954, p. 32). However, Becker has suggested that among other respondents this approach might lead to different results. Becker warns for a possible threat to rapport in long-term studies in which the interviewer can have multiple interviews with a respondent, for instance in an ethnographic field study.

Much later, Hathaway and Atkinson (2003) used a 'two Cop-personas' tactic, to both ensure rapport and pose challenging questions. Following Becker's (Becker, 1954)

research note and a reference to Briggs's meta-communicative exploration (1986), Hathaway and Atkinson first put the respondents at ease by being friendly to gain rapport. At first, they use accommodating and encouraging probing, thus "leaving more critical and pressing or sceptical questions until the more active interview stages. The two personas are thus complementary, ranging on a continuum from accommodation to more challenging lines of inquiry that are intended to confront the claims constructed by informants." (Hathaway & Atkinson, 2003, pp. 164-165).

The translation of the more challenging interviewer behaviour as is described in the literature discussed above is not a giant leap; many of the specific characteristics and techniques can be used flawlessly. In their fear of losing rapport, however⁷, most authors fail to notice the subtle difference between interviewing style and probing tactic, and so they miss the point that an interviewer can be warm in his or her person-orientation, but critical as a part of the task-orientation. So in his or her stance towards the respondent's answer, an interviewer does not take every answer for granted and shows this doubt. This is done without suggesting other ideas or answers, just as in the other two probing tactics. And again, all probing techniques can be used to convey a challenging tactic, but some are better than others. For instance, negative comments, an expression of doubt, an expression of a lack of comprehension or an expression of perceived inconsistency or a request for explanation, especially in the form of a simple why?-probe, is rather typical for the challenging probing tactic.

The message the challenging interviewer has to convey is: "You can tell me anything, but I am carefully and critically listening, in order to challenge you to deliver your story, as relevant, complete, detailed, elaborate and in-depth as possible".

1.6 Research question

The above discussion of interviewer behaviour, showing the distinction between the interviewer's person-orientation and task-orientation, teaches us that in general there is a serious lack of research on the effect of general interviewer behaviour in open research interviews. Although many articles have been published on person-oriented interviewer behaviour in open interviews, most of these articles are either literature reviews or based on autobiographical experiences rather than more systematic empirical research. For task-oriented interviewer behaviour within open interviews there is also a strong lack of methodological research on the effects of different techniques and behaviours. However, in 'How-to-Interview' books, one can find much information on the probing techniques that are available to open interviewers.

It is argued that in open interviewing probing has the most direct influence on the quality and the content of the information that is obtainable through the interview. It is also argued that while probing an interviewer usually has to think about two aspects

simultaneously: which probing technique to use, and also which probing tactic to use. The latter is so unknown that most interviewers probably do not think about it and will either always probe using the same predetermined probing tactic, often in line with their interviewing style, or use different probing tactics depending on their own opinions with respect to certain topics. Nonetheless, nobody really knows the consequences in these cases.⁸

Consequently, the research question of this study is **“What effects do the three different probing tactics have on the *quality* and *content* of the information received in the interviews?”** The different probing tactics are the probing tactics as they are described in section 1.5.3: the *accommodating*, *encouraging* and *challenging* probing tactics.

Thus far, there is one more hiatus in interviewing literature that has yet to be discussed. It should be addressed and established before any answer to the research question is possible. This hiatus is the dependent concept of the research question. Therefore, before discussing the methodology of the experiment, I address how *quality of received information* can actually be measured.

Notes

¹ As the reader will notice I have not used the term qualitative interview yet. This is due to the wide variety of meanings for which the term is used. Some researchers use it for any interview using predominantly open questions (e.g. Kvale & Brinkmann, 2008; Rubin & Rubin, 2005), while many others use it only for unstructured interviews with open-ended questions and completely free interviewing behaviour. Weiss (1994), for instance, differentiates between qualitative interviewing and ‘fixed-question-open-response interviewing’. Nor did I use any of the terms that are sometimes used for various interview types, such as ““permissive,” “nondirective,” “focused,” “understanding-listening,” “supportive,” “depth,” ... “free wheeling,” “subjective,” “expressive,” “spontaneous,” “projective,” “phenomenological,” “indirect,” “transactional” or “psychiatric”” (Gorden, 1969, p. 30) because as Gorden remarks, these are mostly techniques that are too specific to be used “for divergent interviewing situations and at all points within a particular interview” (Gorden, 1969, p. 30).

² Within two of these types they make a more subtle division, leading to the eleven types presented in the nine cells of Box 1.1.

³ In this study, I have decided to start at the upper left of Box 1.1, using *concept clarification interviews*. More on the type of interview within this study is discussed in section 2.2.1.

⁴ Compare for instance (Fowler & Mangione, 1990) and (Viterna & Maynard, 2002).

⁵ Seidman even opposes the use of the term probing, since he associates probing both with “a sharp instrument pressing on soft flesh” and with a “powerful interviewer treating the participant as an object” (2006, p. 83). He prefers to use ‘exploring with the participant’, but for me, that idiom conveys a sense of long walks into unknown territory. Therefore in this research I will use the word probing, and simply associate it with trying to get relevant, more, more specific, more elaborate and more in-depth information from respondents answers.

⁶ Naturally, it is very important when explicitly taking an answer for granted or not that the answer is appreciated for the quality of the content and not for the content itself. In the interviewer instruction extensive attention is given to this issue.

⁷ Brinkmann (2007), speaking in opposition to possible (also constructionist) critics to challenging interviewer behaviour from an ethical point of view, shows how in what he calls epistemic interviews “the interviewers do not try to suck as much private information out of the respondents as possible, without themselves engaging in the conversation with all the risks that are involved in this.” (Brinkmann, 2007, p. 1134). Besides, he criticises the ethical stance that the respondent is always right from a constructionist Socratic standpoint, stating that: “In epistemic interviews, the client is not necessarily right (nor wrong, for that matter), for opinions and beliefs are debated, tried, tested, and challenged in an open conversation, where the validity of the respondent’s statements does not depend on how he or she “really feels” but rather on public and intersubjective criteria— perhaps even ethical ones.” (Brinkmann, 2007, p. 1134).

⁸ Naturally the Jakobson-Hymes model as Briggs (1986) uses is a more fine-grained model for unravelling the components of the interview situation than the model I use, with the focus on

1. Probing Behaviour in 'Open' Interviews

interviewer behaviour and specified in distinctions between task or person-orientation, interviewing style, Probing Technique and Probing Tactics. However, since the focus in this research does not only lie on a few interactions, my description of interviewer behaviour in the interview situation is sufficient and parsimonious.

2 Comparing Interviews on Quality

As was argued in the previous chapter, there is a serious lack of studies on interviewer behaviour in open interviews and there are hardly any studies that address probing behaviour. An even more problematic lack concerns the instruments that are available for assessing the quality of the information received in open interviews. What makes a better interview in terms of the information that is received, and what makes an interview worse? Naturally, in literature on interviewing, many examples of good interviews are shown; unfortunately, what is missing are an assessment of *why* the information is good and the *instruments* for deciding this for yourself.¹

The lack of fine-grained instruments for assessing the quality of the information received by using open questions can be due to the fact that there have not been many methodological studies concerned with the effect of interviewer behaviour on the received information in open interviewing, so few researchers have been motivated to answer this question. This is a rather dreadful justification nonetheless since all ‘How-to-Interview’ books should contain instruments or at least hints that enable any researcher to assess the results of an interview. Furthermore, all interviewers should want to know how good the interview was in terms of received information.

In this chapter, I will first look briefly at three different issues of interview quality. Afterwards, I will show the boundaries set by the interviews as they are held for this study. Subsequently, I will come to a conceptualisation and operationalisation of the quality indicators used in this study. I will also briefly discuss alternative criteria and the reasons for not including them in this analysis.

2.1 Issues of quality in open interviews

When authors publish on open interview studies and discuss interview quality, the focus lies on the process of the interview, a post-interview self-assessment of interviewer behaviour and/or the interaction. As was shown in section 1.3, this is completely in line with the general focus in open interviewing with issues of person-oriented interviewing behaviour. In publications and even in textbooks researchers typically stress that if rapport was good, the interview was good. However, these statements on rapport do not tell the reader anything about the quality of the information, except that the author assumes that through good rapport, one receives better information. But how do we know that this is really the case?

2.1.1 Assessment of interviewer behaviour

There are some authors of ‘How-to Interview’ textbooks that developed instruments to assess some issues of quality in interviews. An example comes from Jeanine Evers and

Fijgie de Boer's (2007b, 2007a) chapters on designing and doing an individual interview. Here, they mention the importance of self-assessment of the interviewer. In an appendix they even present a checklist for the self-reflection on the interviewer quality. With this checklist, the interviewer is prompted to reflect on the introduction, the interviewing techniques, the summaries, probes and prompts and some person-orientation. Gorden's checklist (1992) is comparable to the Evers and De Boer's checklist, although it is slightly more detailed and checks for the establishment of a communicative atmosphere, pacing when posing questions, listening, observing and evaluating the answers. In other words, both checklists are comprehensive self-assessments on the quality of interviewer behaviour that are useful for a post-interview (self-)reflection of the interviewer.

Contrary to closed interview literature, in the open interview handbooks as discussed in Chapter 1, there is no discussion on the coding of interviewer behaviour as a tool for the assessment of the quality of the interview.

2.1.2 Assessment of the role of interaction in acquiring information

The last few decennia there has been an increasing focus on the importance of the interaction within the open interview. Since conversation analysts and ethnomethodologists have turned towards interaction in open interviews, many new insights have come up (Roulston, 2006). Although these researchers do not aim at establishing tools to assess the interaction, their approach of the interview as an interactive process is fundamental to the choices that have been made in this research.

From the early eighties onwards, Carolyn Baker has established a line of research on the interaction that takes place in the interview. By applying insights from ethnomethodology and membership categorisation analysis to open research interviews, she distinguishes the interview as a place where 'culture in action' is taking place. The questions and probes an interviewer poses are not neutral invitations for answers at all. On the contrary, questions and probes actively construct the respondent as a member of a particular category (someone who presumably knows something relevant that is unknown to the interviewer) and this invokes how the respondent should answer. Therefore, the information received through the interview should not be treated as a factual report but as an account constructed in the interaction. In several analyses, Baker shows how interviewers and respondents present themselves, or are 'forced' to present themselves as members of certain categories (e.g. Baker, 2004).

It follows that from a more constructionist perspective the interview is seen as the locus of construction. However, this does not necessarily mean that an interview should be analysed only on the level of interaction. In their influential book 'The Active Interview', Holstein and Gubrium suggest that an interview should be analysed for its data as well as for the interaction. They explain this using *hows* and *whats*:

“.. we think understanding *how* the meaning-making process unfolds in the interview is as critical as apprehending *what* is substantively asked and covered. The *hows*, of course, refer to the interactional, narrative procedures of knowledge production, not merely to interview techniques. The *whats* pertain to the issues guiding the interview, the content of questions, and the substantive information communicated by the respondent.” (Holstein & Gubrium, 1995, p. 4)

In his ‘Art(fulness) of open-ended interviewing’, Timothy Rapley (2001) illustrates how interviewers handle the tension of interacting in a conversation while they must simultaneously collect data on a topic. In his data set, interviewers handled this tension by producing themselves locally as facilitative and neutral speakers, through asking very open-ended questions and not commenting on the answers. Nevertheless, in the context of the interaction, Rapley shows that interviewers’ decisions about *what* to follow-up in the answer and *how* to follow-up can result in the *form* of an interview largely affecting the *content* of the information received through the interview. Hence, Rapley (2001) insists that when analysing an interview, every researcher should analyse the received information in conjunction with the interaction in which it was produced. According to Potter and Hepburn (2005), this is seldom done sufficiently, with only a few exceptions: (Rapley, 2001; Van den Berg, Wetherell, & Houtkoop-Steenstra, 2003; Lee & Roth, 2004).

The awareness that the role of the interviewer and thus the interaction leads to the co-construction of answers from the respondent is one of the underlying motivations of this research. Following this line of reasoning, I account for the interaction by assessing the behaviour of the interviewer *and* the quality and content of the information received. The interaction itself is not taken into consideration for the quantitative analysis presented in this thesis.

2.1.3 Assessment of the quality of the received information

As was hinted at in the previous section, many researchers account for the quality of the information received in an open interview by pointing to the factors that are assumed to lead to higher quality information. For instance, “the questions were posed as scripted”, “the rapport was good”, “the respondent felt comfortable” and “the interviewer probed a lot”. As was argued before, we simply do not know whether this really leads to better information. We also do not know how accurate the self-assessment instruments are.

Because the focus of this research is on the *effects* of probing tactics, it is necessary to establish good criteria for determining the quality of the information received through the interviews. The five criteria I established are: *relevancy*, *depth*, *amount*, *elaborateness* and *specificity* of the information received. These criteria are selected on the suitability for the interview goal and rooted in interviewing literature, when possible. The criteria will be

conceptualised and operationalised in section 2.3. Before discussing them in-depth, two issues have to be dealt with.

The first issue concerns some often mentioned criteria that are not chosen (see subsection 2.1.3.1). The second issue concerns the type of interview used in this study, the substantive interview topics selected for this study and the analytical focus to be used. Although the five criteria mentioned above can be used for very different types of open interviews and substantive interview topics, it goes without saying that the operationalisation of those criteria depends on the specific characteristics of the interview type and the goals of the interview that is being considered (section 2.2).

2.1.3.1 The criteria *not* chosen for the assessment of the quality of the received information

Standard criteria such as reliability and replicability, or its qualitative alterations (LeCompte & Goetz, 1982) and its qualitative alternatives (Guba & Lincoln, 1985) are not useful indicators for the quality of the received information, since these criteria assess the quality of the research as a process instead of the information produced through the research. When considering validity, we see that more realist or ‘positivist’ books focus on validity as an issue of truth, or *accuracy* of the information: “How do we know whether a respondents tells the truth, the whole truth and nothing but the truth?” (Weiss, 1994; Gorden, 1992; Baarda, et al., 1996; Emans, 1990).

As an indicator for quality of information, accuracy is not very useful for two reasons. The first is that other data are required to compare with the answers of the respondents to establish the accuracy of those answers. This triangulation (Webb, Campbell, Schwartz, & Sechrest, 1966; Denzin, 1970) implies a comparison with data from another source that has been gathered with another method. This will result in the classic triangulation problem: how can we assume that the data obtained elsewhere are more accurate than the data received through the open interviews?

The second reason is more philosophically grounded. Because we know how important the role of interaction is in open interviews, it is rather naïve to believe that no co-construction of meaning is taking place. Since this interpersonal co-construction is taking place, the information acquired during the interview is at least partly an accomplishment.

Due to these reasons, some authors suggest transforming the classical question of accuracy from ‘How do you know if your informant is telling the truth?’ (Dean & Whyte, 1958) into ‘How does the informant try to persuade me of the truthfulness of my account?’ (Atkinson, Coffey, & Delamont, 2003, p. 122). Accuracy as a criterion, therefore, would not fit with this new question. Instead, they suggest *credibility* as a part of the trustworthiness set developed by Guba and Lincoln (1985) as a better alternative.

However, the trustworthiness alternatives are not essentially different from the classic criteria or, as Long and Johnson put it, for validity and its alternative credibility: “The only difference between the terms is the presumed objective reality of positivism and the constructed realities of constructivism. The underlying concept appears to be identical: to match what is reported by the researcher to the phenomenon under investigation.” (Long & Johnson, 2000, p. 32). Thus, what is reported is at least partly co-constructed in the interaction and interpretation, and therefore, it cannot be judged in terms of accuracy or credibility. Constructionists would argue that this accounts for all kinds of information (Edwards, Ashmore, & Potter, 1995). However, when taking a milder point of view, one can argue that this differs for different kinds of information. Some factual information is less prone to this co-construction, whereas respondents’ social categorisation in concept clarification interviews is potentially more subject to co-construction.

In addition to accuracy, one could also suggest *incongruency* or inconsistency as an indicator for the quality of information received. Gorden recognises four types of incongruencies:

1. Incongruencies between one generalization and another
2. Incongruencies between one specific fact and another
3. Incongruencies between facts and generalizations
4. Incongruencies between respondent’s statements and information already known (1992, p. 137).

These incongruencies are not always a problem of the quality of information. Van den Berg has shown, using interviews held by Wetherell, that respondents can be very incongruent and even contradicting. Respondents are sometimes self-contradicting, not because they deliver low quality information due to social desirability, or forgetting, or any other reason, but basically because respondents can apply different frames and meanings to the same terms or categories (Van den Berg, et al., 2003). With this in mind, it does not make sense to analyse a large amount of information on incongruencies, when these incongruencies after scrutinised analysis are not effects of a lack of quality but result from a shift in frames.

A last possible criterion that is not chosen in this research is the *level of ambiguity*. From a conceptual point of view, it goes without saying that information that is not ambiguous is better than information that is ambiguous. However, to indicate the level of ambiguity of the answers of the respondent is rather difficult. The criterion is difficult to operationalise and assess reliably by different coders.

Therefore, the previously mentioned five criteria that are chosen are most relevant and useful for assessing the quality of the information received in open interviews. In section 2.3 each of these criteria will first be discussed from a general point of view, and then, the operationalisation of those criteria will be exposed. However, in developing

usable empirical indicators, it is necessary to take into account the interview type, the interview topic and the goal of the interviews being studied. This is the topic of section 2.2.

2.2 Interview type, topic and analytic focus in this research

As was shown in the previous section, the goal of this section is to briefly introduce three contextual elements, in order to understand the exact operationalisation of the quality criteria into indicators in section 2.3. First the interview type for this research will be discussed briefly, and then the topic of the interviews and the focus of analysis will be addressed.

2.2.1 The interview type: concept clarification interview

In Box 1.1 on page 7, an overview of nine types of open interviews was presented. In order to compare different probing tactics, a specific type of open interviews had to be selected. The interviews held in this study focussed on frameworks used for talking about social categories and the meanings that are attached to these categories. In the typology of Rubin and Rubin (2005), this type of an interview is called a concept clarification interview. In their discussion of this type of interview, it is focussed on emic terms or vocabulary and thus the emic meaning of folk terms. In this sense, it is comparable to the domain analysis that is done on interviews that were held in the first stages of an ethnographic study (Spradley, 1979). However, in this study, the terms are not necessarily folk terms, but are common categories that are used in everyday conversations. That does not exclude the possibility that the meanings and frameworks attached to these concepts will vary noticeably, especially in cases that include a diverse group of respondents.

Concept clarification interviews comprise a rather common type of interview. For instance, in a case study, a researcher will often start with interviews in which meanings and frameworks on a limited range of concepts are articulated. In open interviews that are held to inform the construction of a closed question survey questionnaire, or in cognitive interviews that are held in the testing phase of a questionnaire, one of the main goals is to establish the diversity of meanings that exist for the relevant concepts. The concept clarification interview type is also the type of interview to be used when the answers to closed survey questions raise questions. So, the concept clarification interview is used as an aide in a large spectrum of research fields, from ethnography to large-scale survey research.

2.2.2 The interview topic: social categorisation

Social categorisation has become more and more valued as an important topic for consideration in research within the social sciences. In 1965, Michael Moerman had already published an article in the *American Anthropologist* that convincingly revealed that ethnic categories are not as strictly defined as was commonly accepted. Rather, the demarcation and thus the meaning of the social category differs among different people and even in different situations. This accounts for both the self-ascription to a category and for one's ascription to a category by others (Moerman, 1965). From this point forward, the focus in social science has shifted from studying the category boundary and "the cultural stuff it encloses" (Barth, 1969, p. 15) as a stable objectivist entity, towards the analysis of *prototypes* of categories (following the work of Rosch & Lloyd, 1978; Lakoff, 1987) or *perspectives* on categories (Brubaker, Loveman, & Stamatov, 2004).

In several articles on qualitative interviews, Baker (1997, 2004) suggests to use membership categorisation analysis as a tool for analysing the content and interaction of interviews. By analysing the categories and the predicates that are used by interviewees and respondents, who are seen as competent members of their cultures, one can understand how people constitute and understand their life-worlds.

As described in section 1.6, the goal of my research is to come to a conclusion on the effects of probing tactics on the quality of the information received through the interview. Therefore, since the interviews are focussed on concept clarification of social categories, it is feasible to follow the suggestion of Baker and to use membership categorisation analysis. Nevertheless, I will not use it to qualitatively reconstruct the categories and predicates that are used to describe the world as Baker and many others do². Instead, I use it quantitatively, to compare the content of different interviews and as an instrument for comparing the amount, the elaborateness and the specificity of the information received through the interview.

2.2.3 The analytical focus: membership categorisation analysis

When people talk, they implicitly or explicitly categorise. Without categorisation, we simply cannot talk. By using nouns, we classify objects, people, feelings or thoughts; and by using verbs, we attach extra meaning and often action to something. Or, as Lakoff has put it: "Categorization is not a matter to be taken lightly. There is nothing more basic than categorization to our thought, perception, action, and speech" (Lakoff, 1987, p. 5).

How necessary categorisation is and how we cognitively categorise is very interesting for cognitive linguists and cognitive psychologists. However, what interests social linguists and sociologists is how this categorisation functions in talk-in-interaction. In other words, sociologists are interested in what people *do* by using certain categories.

In the sixties, Harvey Sacks developed a toolbox for the analysis of these categories in interaction. Sacks was strongly influenced by the ethnomethodology of Harold Garfinkel and fully agreed with the notion that is central to the work of Garfinkel: people (re)produce social reality through their behaviour (Garfinkel, 1967). Since talking is one of the social behaviours that people do throughout the day, it is possible to assess how people deal with the world outside of them by analysing their naturally occurring conversations. Since talking about others is impossible without categorisations, categories are a very important feature of talk-in-interaction.

What is intriguing about this way of talking about others is that conversational partners need very little clues in order to understand what categories mean and how categories are being used. It is precisely this aspect that intrigued Sacks and led him to develop a set of concepts, rules and maxims called membership categorisation analysis. This ‘machinery’ or ‘inference making machine’, as he called it himself, has been a fruitful ground for later elaborations by many others such as Watson (1978), Jayyusi (1984) and Housley and Fitzgerald (2002), to name a few. In this research, a reconsidered form of Harvey Sacks’s membership categorisation analysis is used.

The first analytical concept to be dealt with in membership categorisation analysis is the central term *membership categorisation Device* (MCD). A membership categorisation device could be defined as a collection of social categories and some rules of application for these categories (Sacks, 1995, p. 246). For instance, for the categorisation device of Amsterdammers, one can come up with some collection of categories such as ‘real’ Amsterdammers and ‘import’ Amsterdammers. The rules of application make clear when these categories are relevant, for example, when one wants to differentiate Amsterdammers along place of origin or even social and (sub)cultural lines.

One of the general rules of application that is developed by Sacks is *the economy rule*. This rule states that “a single category from any membership categorization can be referentially adequate” (Sacks, 1995, p. 246). This means that to talk about a membership categorisation device, people only need to mention a single category and do not necessarily need to explain or explicitly contrast this category to another category. For instance, when assigning someone to the category ‘best friend’, it is clear that the speaker talks about the device of friends. Besides, it is clear that there are other category labels that are possible for describing other types of friends.

The second rule of application for membership categorisation devices is *the consistency rule*: When people talk about someone using a category from a device, they are often consistent and may use categories from the same device for speaking about others. (Sacks, 1995, p. 246). This consistency rule leads to a corollary that is often called the *bearer’s maxim*. This means that if two or more categories are used to categorise two or more persons and if these categories can be *understood* as categories from the same device,

they should be heard as *being* from the same device (Sacks, 1995, p. 247). This is to say that if someone is called a Turkish immigrant and someone else is a Moroccan, 'hear' the second category as Moroccan *immigrant*, since both categories are part of an immigrant device.

Not only the analyst, but also the respondent uses these rules. This is because when an interviewer starts asking questions that use a certain category or set of categories, the respondent will also be applying these rules unconsciously and will try to answer the interviewer within the same device.

Simply naming the *categories* used and possibly assigning them to a device is not enough for gaining insight into how membership categorisation devices are used. Sacks developed the concept of category bound activities to account for the powerful use of activities in combination with categories, or even instead of categories. With category bound activities, Sacks intended to notice that it is common sense to assume that many activities are done by certain categories. For instance, a waitress *serves* food, a taxi driver in Amsterdam *cheats* on you, a policeman *arrests* a suspect and a mommy *picks up* a *crying* baby.³

Watson (1978) argues, following Sharrock (1974), that not just activities are bound to categories, but other predicates are as well, such as knowledge, beliefs, features and rights and obligations. These predicates can be used in the same way as Harvey Sacks used activities (Hester & Eglin, 1997).

These category bound predicates are so powerful that sometimes just naming the predicate and using the economy and/or the consistency rule is enough for understanding what category people are talking about. For instance, when someone is talking about friends and says, "I can call him up at night and he will come over and help me", it is clear that this person is referring to a good friend and not a nurse, general practitioner or a policeman.

2.2.3.1 Membership categorisation coding

As was shown, membership categorisation analysis is more than simply describing how categories are used. Through the machinery, as Sacks calls it, an analyst can rather quickly interpret how the categorisation is done because the rules formally formulate what we do unconsciously in everyday conversation.

Many conversation analysts are opposed to coding or quantification, since they believe that superficiality would sneak in (Schegloff, 1993). However, it is ridiculous to spend years developing new insights into what rules people use for talking, and then denying people use of these same rules when it comes to analysing a larger number of conversations or interviews. In addition to coding, quantification could lead to comparison, which naturally comes at the cost of contextualised knowledge and nuance,

but is very valuable when looking at more than just a few lines of data. Beyond that, the membership categorisation rules are very helpful for coding the information, since the rules can be applied as reliable interpretation rules for coders. Recently, King (2010) also showed how he successfully used membership categorisation rules for coding categories and predicates in interviews with young people on their ‘gap year’.

The membership categorisation coding system consists of two main sets of codes: codes concerning the categories and codes concerning the category bound predicates.

The categories are coded per topic. Once a respondent mentions a category, it is possible that he or she will refer to the category later without explicitly mentioning it again, for example, with categories such as “They” or “Those people”. These category indexes are independently coded as well. Respondents could use many different category labels or apply many implicit indexes to one category. Both result in a high number of category instances.

As described in section 2.2.3, since Sacks, most developments in membership categorisation analysis have been on the extension of category bound activities to category bound predicates such as category bound features and category bound beliefs (e.g. Sharrock, 1974; Watson, 1978; Jayyusi, 1984).

In this research, I extend these reconsiderations by classifying 16 different category bound predicate types (from now on, these are often abbreviated as *predicate types*). These predicate types have been defined in advance in order to deal with the large amount of different specific predicates as they are spontaneously used in the interviews. These types are then used in the coding of the respondent turns in the interviews.

2.3 Criteria for the quality of information received in interviews

Relevancy of the information

This criterion for evaluating the quality of the information is only discussed in a few interviewing books. Emans (1990) and Gorden (1992) suggest evaluating the relevance of the information to the research question. At first, this seems a good indicator of quality. However, relevancy is sometimes rather difficult to judge. Gorden suggests that “the interviewer should sometimes probe responses that seem irrelevant, but have the potential to be relevant” (1992: 137). In open interviews, whether information turns out relevant depends on the probing of the interviewer. A seemingly irrelevant statement could turn out to be relevant after probing, but without probing, the same statement would remain irrelevant. Therefore, relevancy is an important criterion in this study.

What counts as relevant or irrelevant talk depends heavily on the broadness of the goal for the interview. In this study, the goal for the interviews is rather broad: interviewers need to get as much relevant, in-depth, elaborate and specific information as

possible. Consequentially, the operationalisation of relevancy has been inverted; it is focussed on *irrelevant* interview talk. Moreover, the concept is slightly broadened and less dependent on individual judgements: The interview transcripts had to be coded on *off-topic talk*, which could be talk over the coffee, but also diverted answers from the respondents that were at least two logical steps away from the original question. Off-topic talk is measured by counting the number of turns that contain exclusively Off-topic talk. The advantage of this approach is that all other talk could be seen as on-topic and relevant.

Depth of the information

'In-depth interview' is a term first-year sociology students are quick to use for any introductory interview assignment. This is not as peculiar as it seems because using the term in-depth invokes all kinds of notions of respondents opening up, uncovering and disclosing very personal information that is normally deeply hidden. In other words, by claiming depth, these students claim quality. In interview reports in journals, the term is a little less abused, but it is still often used as an empty claim for quality.

Often, a direct relation is drawn between rapport and depth of the interview (Evers & De Boer, 2007a, p. 59). In Chirban's (1996) 'Interviewing in Depth', rapport is not just a relation, but it simply *means* depth. This is not necessarily true. Again, this direct relation is typical for the assumption that person-orientation almost automatically leads to high quality data.

In an early article 'Dimensions of the Depth Interview', Gorden describes depth as context dependent:

The 'depth' of any item of information depends upon its meaning for the respondent, which, in turn, depends upon how he perceives the relationship between the information and the total social context in which it is given. What is in one social situation a mere 'objective fact,' as, for example, the respondent's age, may be a devastating threat in another (Gorden, 1956, p. 158).

Since depth is context dependent, he further suggests focusing on the obstructions to depth and varying interviewing techniques and tactics to deal with these obstructions, instead of answering the question: "Which dimensions of depth could be recognised?"

In recent literature, depth is predominantly associated with uncovering personal information such as information relating to personal events or personal feelings. Rubin and Rubin, however, also relate depth to "an answer that goes beyond the superficial, beyond the first response, to a second and third level, and maybe more." (Rubin & Rubin, 2005: 130). In other words, they also see depth in alternative explanations and more complex data. Nevertheless, all of the additional examples they present on depth contain personal information.

When taking an approach that is more in line with Goffman (1959) and Bangerter (2000), the respondents' self-disclosure of personal information could be interpreted in a different way. A respondent that seems to be 'opening up' is, then, merely working on self-presentation. Using a discursive psychology approach, Antaki, Barnes and Leudar show that disclosure of personal information is highly dependent upon and largely structured by the conversational context; they demonstrate that in psychology what is considered a self-disclosure is mostly a social performance in an interaction. Furthermore, they forcefully attack the psychologists' approach to disclosure of personal information as a dependent variable for not accounting for the strategic aspect of these personal revelations. For instance, respondents often use extreme-case formulations especially in relation to self-disclosures in order to make them significant to the speaker (Antaki, Barnes, & Leudar, 2005).

Although the strategic production of disclosures could be seen as a major critique of the use of personal information as an indicator of depth, it will still be used in this study. Atkinson, Coffey and Delamont (2003: 123) show that precisely this strategic production could be seen as a major quality indicator. So, as a quality indicator for depth, it does not matter so much if respondents are revealing *true* personal information because it is interesting enough that they use it in itself, albeit as a rhetorical device.

In this research, depth is operationalised by counting the number of respondent speech turns that contain new empirical information about the personal life or feelings of the respondent.

Amount of the information

In their work on relevance, Gorden (1992) and Emans (1990) suggest that interviewers should also consider completeness when evaluating the answers of respondents. Naturally, this also depends on the scope of the interview goal and the scripted questions.

It could be assumed that the information is complete when the respondent does not give any new information and new probes just leads to redundant answers. When interviewers follow up this 'probing for saturation' with summaries of the relevant answers and probes for possible additions before continuing to the next topic, this assumption becomes more grounded. The use of this stop-criterion has the great advantage that most checking on completeness is done by the interviewer in conjunction with the respondent. As is the case with personal information, we can never be sure that an answer is really complete. Therefore, the amount of information can function as a proxy for completeness.

For researchers from all kinds of epistemological and ontological backgrounds the amount of information is important. Moreover, regarding the amount of information, almost all authors on interviewing would agree that the more relevant information that is

brought forward in the interview, the better. This notion is so much part and parcel of the open interview that authors fairly often do not mention amount of information as an evaluation criterion, but rather offer comments such as “listen more, talk less” (Seidman, 2006, p. 78). Although they present a list of what good answers are, Rubin and Rubin do not even mention the amount of information, while discussing other criteria (Rubin & Rubin, 2005). Weiss only briefly touches upon the amount of information in his discussion on events: “If an event is of critical importance for your study, you should try to get as much information about what happened as your respondent can supply, up to the point where the respondent becomes restive”(Weiss, 1994, p. 80). I would claim that this is at least as important for other topics as it is for event reconstruction for instance, for concept clarification interviews but for other interview types as well.

Thus, the difficulty with this criterion thus does not lie in the criterion itself but much more within the operationalisation and the measurement of this concept. Everyone seems to agree that in open interviews the amount of relevant information is important, but the question remains how to account for that. What counts as information depends partly on the interview goal. To answer this question, membership categorisation analysis, the analytical focus for this research, offers a useful base.

To assess the amount of relevant information in the interviews on social categories, as they have been conducted in this research, it is sensible to count the number of times category bound predicates are used. Since categories are seldom used without predication, the number of category bound predicates used serves as a good indicator for the amount of relevant information.

Since repetition of exactly the same predicate-category combination could be either repetition of the same information or extra rhetorical stress being put on that aspect of a certain category, it does not add up to *new* information. Therefore, repeated predicate-category combinations are not included in the indicator. So, the number of times respondents mention new category bound predicates is a sensible measure for the amount of information.

Elaborateness of the information

The elaborateness of the information is the fourth quality indicator that is important for this study. Again, only Rubin and Rubin (2005) note the importance of elaborateness. They call it richness. “Richness means that your interviews contain many ideas and different themes, often including those that you did not anticipate when you began your study” (Rubin & Rubin, 2005). The attention to the more serendipitous information is rather free of charge, since in all open interviews serendipitous information could come forward, even in interviews with a rather narrow focus, such as concept clarification interviews. The attention to the number of ideas and different themes is naturally

dependent on the interview goal. If the interview goal is concept clarification, the number of dimensions a respondent recognises is more important than many ideas on other concepts.

The question of operationalising the elaborateness of information is comparable to the previous quality criterion: What information should be counted to establish the broadness? To answer this question, the analytical focus of membership categorisation analysis is helpful. When a respondent elaborates on a topic, it is very likely that he or she will be broadening either the membership categorisation device with new category labels or the predication of the category, by naming many different predicate types. Therefore, two indicators are used for the elaborateness of information, one of which is for category labels and the other being for predicate types. The more different category labels or predicate types a respondent uses, the more the respondent is elaborating on the topic

Specificity of the information

In their extensive lists of what good answers are, Rubin and Rubin underscore the importance of detail (Rubin & Rubin, 2005)⁴. A detailed account of an event does not just include a large amount of information on that event, but specific information as well. Just as is the case with events, in concept clarification interviews, such as the ones held in this research, it is important to get as much specific information for the concept as possible. The more specific and precise the information is the better.

As with the relevancy, depth, amount and the elaborateness of the information, positivist, constructionist and emotionalist (See Silverman, 2006) researchers could equally well use this criterion. For example, for an emotionalist, specificity is very important because through the specific description of the feelings, he or she can really understand the feelings of the respondent.

The operationalisation of specificity relates closely to the operationalisation of elaborateness and the amount of information. It is also partly dependent on the topic of the interview, so codes on social categorisation are used. The specificity of the information is also measured using two different indicators. Both indicators are functions of other indicators.

The first indicator is the average number of category bound predicates per predicate type. The more predicates that are used per predicate type, the more specific the information on that predicate type has been. Therefore, the average number of predicates per type is a good indicator for the specificity of the information.

The second indicator is the average number of category bound predicates used per category label. The higher the number of unique predicates per category label, the higher the specificity of the information.

2.4 Quality of received information: a summary

In this chapter a set of measures has been presented for comparing the quality of interviews. It was shown that the usual focus on interviewer behaviour is partly relevant because it is based on the assumption that good interviewer behaviour leads to good quality information. But, to be able to assess the quality, it is necessary to analyse the answers of the respondent while taking into account the idea that answer-behaviour is always part of the interaction between interviewer and respondent. Thus, the focus of this research is on the quality of the data. Five indicators will be used: relevancy, depth, amount, elaborateness and specificity of the received information.

The first indicator, relevancy, is based on the inverse of relevant information: off-topic talk. The depth of the information has been operationalised as the amount of personal information turns. The three indicators, amount, elaborateness and specificity, are based on codes that follow from an adaptation of membership categorisation analysis. Through considering the possible objections to the quantification of this qualitative analysis technique, I developed a useful tool for establishing quality in interviews.

Since this research is an experiment in two ways, by research design and by the novelty of the probing tactics as topic under study, it is hard to suggest which of those probing tactics would lead to the highest scores on one or more of the newly developed quality indicators. Therefore, the research is exploratory in its outlook and experimental in its research design.

Notes

¹ A good exception for this is Robert Weiss' 'Learning from strangers' (1994), in which 36 pages are reserved for annotated examples of good and bad interviews. And, he even uses a criterion to assess the quality (see section 2.3).

² See for an extensive list of MCA publications: Paul Ten Have's bibliography on membership categorization analysis. Retrieved 5 March 2009, from <http://www.paultenhave.nl/MCA-bib.htm>

³ And, we even hear that the baby belongs to the mommy. The rules to come to this specific conclusion are described in Sacks' lectures in (Sacks, 1995, p. 248) or Silverman (1998). They are, however, strongly denied as a part of Membership Categorisation Analysis by Schegloff (2007).

⁴ Rubin and Rubin (2005) even include two more criteria for quality: nuance and vividness. The first is rather normative, as if "black and white" answers of respondents are less good and the second is necessary to make good titles and anecdotes, rather than quality.

3 Design and Methodology of the Data Collection

3.1 Introduction

When trying to answer the research question “What effects do the three different probing tactics have on the quality and content of the information received in the interviews?” the best research design is an experimental design. All other types of designs will have serious flaws in establishing causal relations between the probing tactics and the quality or the content of the information obtained through interviews.

In addition to this short introduction, this chapter consists of four sections. In the first section, the general outline of the research design of the main study is shown. In the second section, the design and the results of a pilot study are shown. In the third section, the design, the methodology and the practical organisation of the main experiment are described. The last section is a short description of the interviews conducted within the main experiment.

3.2 General design

As was already stated in the introduction, an experiment is the best way to identify the possible effect of probing tactics. The internal validity of an experiment is much higher than any other form of design. However, for any experiment to be internally valid, the possibility to control intervening or confounding variables is crucial. In this study, which uses a field experiment, this is no different. An experiment is only better than other designs when performed properly.

A difficulty with a field experiment in comparison to a laboratory experiment is that there could be many factors potentially contaminating the effects of the probing tactics. As the research takes place in a social setting, involving live social subjects, the potential threats to internal validity could be substantial.

To eliminate the potential threats to internal validity, several conditions of an experiment should be considered. First it is vital to ensure that the experimental variable is varying according to plan. Therefore, the interviewers should be thoroughly trained in their respective probing tactics; and, it is essential to test whether the interviewers are truly able to perform the instructed probing tactics. *Furthermore*, the performances of any of the different probing tactics should be distinguishable from the other two probing tactics.

The second condition for eliminating threats to internal validity in an experiment is controlling for possible confounding variables. In this study, this could be achieved in two ways, either by trying to keep all other possible confounders equal or through the randomisation of possible confounding variables.¹ For this randomisation, naturally, a

large sample should be taken. Due to the many possible threats in this experiment, both the randomisation and equalisation of as many other factors as possible has been aspired.

3.2.1 Experimental design

In view of the fact that the question in this research is about the effects of probing tactics, all other interview aspects, such as the questionnaire, the scripted questions, the time and location of the interviews and many others, should be kept as constant as possible.

In the design of this research, the central issue to be solved is whether to keep the interviewer characteristics, or the interviewer behaviour, constant. In more practical terms this means I had to choose whether to train all interviewers in all probing tactics so as to neutralise possible interviewer variation or to train different interviewers in different probing tactics to maintain consistency among all kinds of interviewer behaviour effects, such as practice effects or (sub)conscious preference for a probing tactic.

The design I selected in this study is a between-subjects experiment. This means that the interviewers were trained in one of the three probing tactics. The reason to choose this design is that interviewers are more easily and clearly randomisable than (sub)conscious preferences. In addition, I believe that it is less complicated to train interviewers in only one probing tactic than in three different tactics. Imagine all possible complications that occur when an interviewer receives a reasonably comparable training three times. *Moreover*, when an interviewer is trained in only one tactic, he or she will be less prone to accidentally slipping into another probing tactic.

The disadvantage of a between-subjects experiment is that the interviewers differ per tactic. In a design in which the interviewers receive three trainings, interviewer talent and other characteristics remain constant. It follows that if one interviewer is better than others, in a between-subjects design, he or she only belongs to a single tactic. To resolve this problem, I had to select a rather large number of interviewers and randomise the assignment of the interviewers to the different tactics.

As anyone would expect, the three different probing tactics can lead to different effects among different respondents. It is impossible and nonsensical to let one respondent give the same interview three times. Therefore, the only logical solution for dealing with this confounding factor is through randomising the respondents' assignments to interviewers and thus to tactics. Box 3.1 summarises the discussion above.

Box 3.1 Treatment groups, randomised and constant confounders

Experimental variables	Trained probing tactics: accommodating, encouraging, challenging
Randomised confounders	Interviewers assignment to probing tactics Respondents assignment to interviewers
Constant confounders	Other factors, such as questionnaires and interviewing style.

3.2.1.1 Experimental variables: probing tactics

In this section I will briefly discuss the most important features of each treatment group: the three probing tactics and their most important probing techniques.

Accommodating probing tactic

In the first treatment group, working with the accommodating probing tactic, the interviewer withholds evaluative reactions during the use of probing techniques. As a result, the respondent has the freedom to answer without being distracted from his or her planned response. The stance of the interviewer is revealed under no circumstances, although the implicit stance would most often be interpreted as taking the answer for granted, since the interviewer does not object to any response.

Probing techniques that give way to the respondent are typical for the accommodating tactic. Therefore the three minimal techniques, active silence, humming and echoing are exemplary, as are more neutral reactions such as question repetition and summarising. Comments are rather rare in the accommodating tactic, since only a few comments do not explicate a stance. Most expressions are rare as well. However, the ‘expression’ question of comprehension (a question of whether the interviewer understood the answers correctly), is very much a part of the accommodating probing tactic.

Encouraging probing tactic

The second treatment group is the group in which the encouraging probing tactic is used. With this tactic, the interviewer takes the answer of the respondent for granted and comments positively on the quality of the answer. The logic behind this is that positive reinforcement will encourage the respondent to continue answering.

Positive comments, expressions of comprehension and reflections in line with the answer of the respondent are typical probing techniques for the encouraging probing tactic.

Challenging probing tactic

The third treatment group uses the probing tactic in which the answer is not automatically taken for granted. This stance is then subtly exposed to the respondent, in order to make the respondent reassess his or her answer, resulting in more solid and considered answers.

Probing techniques that are typical for this probing tactic are critical or ‘negative’ comments and requests for reasoning such as why-probes. Most of the explicit display of the stance used while probing is in the challenging intonation when using any probing technique.

3.2.1.2 Training

The trainings for each tactic are particularly important, since the experimental variable ‘the probing tactic-as-conducted’ is dependent upon ‘the probing tactic-as-instructed’. This means that not only the probing tactic should be trained differently, but that the *ceteris paribus* aspects of the experiment should remain truly ‘paribus’. Therefore, great attention will be given to the form and the content of the training. Both the training and the instruction manual would, moreover, be pretested in a pilot study.

3.2.2 Design of the interview study

As with all experiments on methods of data collection, this research requires a second design as well. In this experiment it is a design for an interview study. To be able to conduct the interviews that would be necessary for the experiment, a decision concerning population and a selection method should be made. Naturally, the interview should have a topic, and the questionnaire should be developed. As part of the ‘all other aspects are equal’ strategy, the questionnaire should contain a large proportion of scripted questions and topic bridges, while leaving space for improvised probing.

In the following sections, the interview topics, the questionnaire, the research population, the respondent selection and the interviewer selection are discussed.

3.2.2.1 Interview topics

As was already shown in section 2.2.1, the interview type that is used in this study is a concept clarification interview on social categorisation. In other words, the purpose of these interviews is to explore how respondents define and discuss social categories. These types of interviews are particularly interesting for in-depth analyses of discursive constructions of social categories, as well as quick qualitative pre-test studies for meanings attached to social categories as they are used in closed questions of survey interviews.

In her study on the meaning attached to attitude objects (as well as in her case social categories) in closed questions, Carabain (2007) concludes that those meanings can differ and thus lead to differently interpreted questions. Therefore, it is interesting to study social categorisation, i.e. how people categorise themselves and others and how different meanings are attached to these categories.

In this study, open questions and probing are used to explore the different meanings and predicates that respondents attach to social categories. It seems logical that the probing tactic has an effect on this process of meaning construction. It is also possible that the effects of the different probing tactics are dependent on, or influenced by, the topics of social categorisation that were chosen.

Correspondingly, it is important to test for any topic dependent differences between the quality effects for the three probing tactics. While all interviews were on social

3. Design and Methodology of the Data Collection

categorisation, three social categories were selected as ‘topics’ for the interviews. The choice for topics is based on two considerations. The first consideration is whether the category is controversial or not. The second consideration is whether the category is personal or not.

The first topic, the social categorisation of *Amsterdammers*, was chosen, because it is a fairly uncontested and neutral topic. Moreover, I expected that people from Amsterdam would be interested and eager to be interviewed about what an Amsterdammer is.

The second topic, the social categorisation of *Friends* was chosen since this is a fairly personal topic. When discussing how many friends a respondent has, or what a friend is, the respondent has to reveal rather personal information. At the same time, this topic seems rather easy to talk about because it seems to belong to common knowledge.

The third topic was chosen because it is a more controversial topic: the social categorisation of *Allochthons*. According to the CBS (Statistics Netherlands), an allochthon is “every person living in the Netherlands of which at least one of the parents was born abroad”². In everyday discourse, allochthons is also a well-known term often used to designate a variety of immigrants and their descendants. The category allochthons is controversial in several senses. It is controversial because it is a central concept in the debate on the multicultural society. It also leads to controversy because in everyday discourse, it is sometimes connotated with problems of the multicultural society or some categories within it. Furthermore, due to the negativity around the term, some people contest the term itself and would want the term to be abolished. Allochthon is habitually used in contradiction to autochthons, which are according to the definition of the CBS people living in the Netherlands of whom both parents were born in The Netherlands.

3.2.2.2 Questionnaire

Since the goal of this research is to analyse the differences in retrieved information that result from differences in the probing tactic used, it is necessary for the interviewers to pose all important scripted questions identically. Therefore, interviewers are not allowed to alter the question formulation or the order of specific sets of questions.

As part of this *ceteris paribus* strategy, not only are the questions scripted, but the introduction and possible topic bridges are as well. The interviews should therefore only differ with respect to the probing tactic used.

3.2.2.3 The research population

The research population is delimited by age, place of residence and autochthony/allochthony. For the interviews, the decision was made to only interview people older than eighteen.

Probing Behaviour in Open Interviews

For practical reasons, I decided that only people living in Amsterdam were interviewed. Again, a mixed population would have added possible confounders. However, the main reason was convenience.

Due to the selection of the allochthon social categorisation topic, the population under research is delimited to an autochthon³ population. Naturally it is possible to interview allochthons about the allochthon social category as well, but due to the controversiality of the term and to decrease the effect of possible confounders, the choice was made to interview autochthons only.

3.2.2.4 Interviewers

The sampling in this experiment has two levels. The first level is the selection of the interviewers. The second level is the selection of respondents to be interviewed by the interviewer and thus interviews. So, interviews are nested under interviewers. As Kish (1965) shows, statistically, the selection of the primary level is most important. Therefore, instead of having a few interviewers doing as many interviews as possible, power increases much more with a higher number at the primary level. This means that by design, as many interviewers should be trained as possible, and the number of interviews they do is less important.

When using only a few interviewers in a between-subjects design, one 'bad' interviewer could be devastating for internal validity. Therefore, subjects have to be selected rather carefully. Furthermore, the number of interviewers to select is not only important for the internal validity, and indirectly for external validity, but a good sample of interviewers is essential for transferability to other studies.

Researchers conducting open interviews most often use three possible strategies for selecting interviewers; 1) they do the interviews themselves, 2) they use experienced interviewers, or 3) they use novice interviewers and ensure a decent training.

Following Dijkstra (1983) and his reasoning, I chose inexperienced interviewers for four reasons. The first reason is a practical reason: it is easier to find the requisite number of inexperienced interviewers than it would be to find that many experienced interviewers. The second reason is that all inexperienced interviewers start at about the same level, whereas experienced interviewers would be much more diverse. The third reason is that experienced interviewers are more costly. And last but not least is that experienced interviewers are harder to train for a task that contrasts their regular form of interviewing.

To assess whether it was feasible to train inexperienced interviewers and more specifically, to see whether it was possible to train interviewers in a certain probing tactic, a pilot study was held.

3.2.2.5 Respondents

When testing the differences between different probing tactics, it is important to account for a rather representative sample of access panel members. Non-response, however, makes simple random sampling rather cumbersome. Therefore, most commercial survey organisations work with access panels. In this research, 'normal' survey research will be followed, and as a consequence, the same type of respondents will be used.

Naturally, the problem with online access panels is that one needs Internet. There is still some non-coverage of the Internet. Besides the non-coverage, the major problem is the respondent's self-selection for the access panel. As Loosveldt and Sonck (2008) have shown, there are several selection steps to be taken. First a respondent needs to be interested in online panel research, or in political or social issues, or else he or she would not surf the websites that invite people to join such a panel. Second, that person needs to actually join the access panel; and third, the person needs to decide to join in on a specific survey topic when invited to do so.

Another disadvantage of using this selection method for open face-to-face interviews is that, on top of the self-selection described above, respondents need to decide to apply online for the face-to-face interview. However, compared to the other options of large non-response, and thus more indirect self-selection and uncontrollable convenience samples, this option is the most preferable.

Bethlehem and Stoop (2007) along with Loosveldt and Sonck (2008) conclude that the representativeness of online panel surveys, even when using weightings, is not sufficient to be similar to the general population. Still, in both articles, the suggestion is made to use online panels for experiments such as the one in this study.

3.3 Pilot study

Since the probing tactics have been developed in this study, it was unknown if interviewers could actually perform the tactics as they were developed. Therefore, I organised a pilot study to assess the feasibility of the training and performance of the probing tactics. The primary questions for the pilot study were: can interviewers be trained in one probing tactic and sustain this behaviour throughout an interview? And, are these probing tactics distinguishable enough in practice?

Another threat to internal validity, the role of unexpected confounders, was tested as well. Besides, the pilot study was an ideal opportunity for investigating many other aspects, such as the instruction manual, parts of the training and the questionnaire. Through the pilot study, I was also able to pilot the organisation of the experiment, the transcription and the development of the coding system for the main study.

3.3.1 Selection of interviewers

For the pilot study, nine young ‘autochthon’ women of Dutch origin who were between the ages of 18 and 32 were selected to conduct the interviews. I chose these particularities to stabilise the possible confounders of gender, ethnicity and age differences. In a small-scale pilot study with only nine interviewers one deviant interviewer could alter the interpretations completely. More interviewers were not an option, nor were they necessary for the main goal of the pilot.

The main goal of the pilot was to find out if it was possible to *perform* the probing tactics as trained. Therefore, actresses were enrolled to conduct the interviews; if actresses could not perform and stick to the role-playing that was demanded by probing tactics, who else could? Beforehand I considered that if these actresses were successful, a second pilot study in which ‘normal’ interviewers would attempt to perform the probing tactics was an option.

Five interviewers were either graduates or students of private or public academies of dramatic art in the Netherlands. Three interviewers were social science students who acted in a student drama club.⁴ All interviewers conducted four interviews.

The interviewers were divided into three groups of three interviewers each. This division was done on their agenda. This led to the coincidence that two of the three social science students were assigned to the challenging probing tactic group. This did not seem to be a problem, since generalisations were not to be made from this pilot. The other social science student was assigned to the encouraging probing tactic.

3.3.2 Selection of respondents

The respondents were recruited through a classified advertisement in a local newspaper ‘Het Parool’ and offered a twenty Euro incentive. All respondents lived in Amsterdam and were of different ages, with different educational and economic backgrounds. The advertisement mostly emphasised ‘life in Amsterdam’ as the interview topic, whereas the interview contained other topics as well.

3.3.3 Training

Each group of interviewers was trained for one day. All three groups were trained in exactly the same way; the trainings only differed in the use of a probing tactic. I conducted all trainings and during the training, colleagues assisted me. During the trainings, several role-play exercises were held.

Since the interviewers/actors of the different groups were sometimes known to each other, the interviewers had to be informed about the goal of the pilot study. This meant that the other two probing tactics were mentioned. However, the interviewers were

assured that their probing tactic would most likely be the best tactic. This trick seemed to motivate the interviewers.

As part of the training, all interviewers held one complete practice interview with a student. Through a closed video circuit the interview was watched and discussed by me and the other two interviewers. Afterwards, our comments were shared with the interviewer, and important parts of the interview were watched and discussed. A lesson learned from the pilot study was to plan extra time for discussion and feedback after doing a closed video circuit interview.

3.3.4 The interviews

All interviews were held in a room at the university on the day after the training. All interviewers conducted four audio and video recorded interviews.

The interviewers were instructed to keep the interview within a time limit of about thirty minutes. After the interviews, the interviewer received feedback. The time schedule was extremely tight. This exhausted the interviewers because after a lengthier interview, or lengthier feedback time, the interviewer hardly had time to recuperate for the next interview. Due to the fatigue, the fourth interview was often not the best interview.

3.3.5 Question development

A second goal of the pilot study was to pretest the scripted questions. Questions were adjusted as soon as problems around the questions arose. The 36 interviews led to 14 significant changes, such as adding, deleting or rephrasing questions or scripted topic introductions.

Ten of the significant changes were the direct consequence of the single sided information given to the respondents on the interview topic: The interviews were advertised with the topic 'life in Amsterdam'. Some respondents were irritated by the fact that one third of the interview was on Amsterdam, while the other two thirds were on different topics. To handle this irritation, the topics were connected by rephrasing them within an Amsterdam context. Scripted topic introductions were added, deleted or rephrased. A whole new topic consisting of two questions was added at the end of the interview to shift the frame entirely back to Amsterdam. The adjustments turned out to be rather successful; after the adjustments, there were no more complaints about the other topics.

3.3.6 Analysis of feasibility

During the interviews it was already apparent that the interviewers were capable of probing in their trained tactic. From the observations through the closed video circuit in

the adjacent room, it was clear that the probing tactics were also feasible to sustain for the duration of the interview. The actors performed their roles convincingly. Of course there were interviewer mistakes, such as suggesting answers, skipping questions and quick answer satisfaction, but in general, the interviewers did not seem to make many mistakes in the application of the assigned probing tactics.

Obviously there was quite some variety among the interviewers. Some had great acting skills and performed well in the probing tactic but did not always apply them to the content of the answers. Only one interviewer had great difficulty in performing her role as an interviewer using an accommodating tactic. She could hardly restrain from commenting on the respondent's answers and once even showed her disagreement with a respondent. All others could quite effortlessly adhere to the instructed tactics.

Due to the video recording all interviews could be watched again later, so the judgement could be based on more than a single viewing. Some parts of the interview were extensively analysed later.

In the end, the best interviewers were not the (best) actors but the social science students with drama experience. Given their study background, they related to the topics on a deeper level, and the eloquence of their probes was simply of a higher calibre. For these reasons, it was a pity that two of them were assigned to the challenging probing tactic.

Of the 36 interviews, 26 randomly chosen audiotapes were transcribed and coded using ATLAS.ti⁵. The main objective for the transcriptions was the development and some pretesting of the coding system.

Although both the observations during the interviews and later viewings of the videotapes convinced me of the feasibility, an extra quantitative indication would be valuable as well.

These indicators also show that the probing tactics were successfully performed. When reviewing the section of the interview in which the interviewers asked questions and probed on the sensitive topic friends, the differences between the behaviour of the interviewers in the three different probing tactics were clear. This section seemed to be the most difficult part of the interview for the interviewers to sustain their tactic (and was hardly touched by the changed questionnaire). From the analysis of codes assigned to speech turns in the 26 transcribed interviews, we can find examples of the differences between the tactics of the pilot study.

The first example is silence. As a probing technique, "Active silence" is the most non-directive probing technique available. This makes it a very good technique for accommodating interviewers, since it allows the respondent "to proceed in whatever direction is most interesting or meaningful" (Gorden, 1992: 149). The advantages of this technique for the accommodating probing tactic were part of the interviewer training.

The interviewers acted as instructed because in the interviews, the accommodating interviewers used the active silence technique far more often than the encouraging and challenging interviewers. This is demonstrated (see Table 3.1) by the non-parametric Kruskal-Wallis test, in which the interviews are ranked on the number of silences and then the average rank of the probing tactic is compared to the average ranking of the other tactics.

Table 3.1 Kruskal-Wallis analysis of frequency of active silence grouping variable: probing tactic (N=26 Interviews)

	Accommodating (N=11)	Encouraging (N=7)	Challenging (N=8)
Active Silence	18.6	11.1	8.7
$\chi^2(2)=8.74$ $p=0.013$			

However, the accommodating interviewers were not just actively silent, for they also used other probing techniques. Some of the most distinguishing are neutral minimal responses such as “oja”(“oh yeah”) and hums, which were used more often than any other probe. One of the accommodating interviewers hummed in almost 65% of her probes.

The challenging probing tactic can easily be distinguished when looking at why-probes. Interviewers interviewing with the challenging probing tactic use significantly more why-probes than interviewers using the encouraging or accommodating probing tactic. The average length of the probes used by the challenging interviewers is significantly longer than those used in the other two probing tactics. This is probably because the probes used in this tactic tend to be more specific, since the respondents are asked to reinterpret their own answers.

These indicators, as well as the judgments made while watching the videos, led to the conclusion that it is possible to train novice interviewers to perform and sustain a probing tactic throughout the interview. The fact that amateur actresses with a social science background outperformed professional actresses provided beneficial information for the selection of interviewers for the main research. It was not necessary to do a second pilot study, to see if ‘normal’ interviewers could perform the probing tactics as well, a social science background and good communicative skills happened to be more important.

3.4 Methodology of the main experiment

The main field experiment consisted of only one manipulation, the training in a certain probing tactic. The experiment was a between-subjects experiment. This means that all interviewers were trained in only one of the three probing tactics. In total 36 interviewers were trained, each of them doing six interviews, resulting in 216 interviews.

To tackle the difficulty of training twelve interviewers at once, the decision was made to split up the number of interviewers linked to the different probing tactics. This

meant that instead of three groups of twelve interviewers, six groups of six interviewers had to be trained.

The danger of creating more training groups for the experimental variable is that the time span between the first and the last training would be too large. To ensure that no history effects would be part of the design, the groups had to be trained within a time span of three weeks and in a mixed tactics sequence: accommodating first, encouraging second, challenging third, and then the same order repeated. History effects are not singled out, but at least the time span is the smallest possible.

An important improvement with the creation of more interviewer groups is that the order effect of the training becomes less important, since the trainings for each tactic take place twice. Thus the possible effect of improved trainings due to the increased experience of the trainer, or relapse due to tiredness from training interviewers will be more or less equally divided over the tactics. One other advantage of training every tactic twice is the spread of risks. The tested factor will not be completely dependent on a single training.

3.4.1 Organisation of the experiment

The data collection for the field experiment took place in May and June, 2005. To decrease the effects described above, a very tight schedule was set up. Every group of interviewers was trained for two consecutive days. After the training, the interviewers had exactly two weeks to complete the six interviews. The interviews could be held at any time the interviewer and the respondent were able to meet. The interviewers were not allowed to do more than three interviews on a day. Most interviewers kept to a maximum of two interviews per day.

After two or three interviews, the interviewers had to return for feedback on their interview behaviour. The days in between the training days were used for feedback. The schedule could not have been tighter.

3.4.2 Treatment

3.4.2.1 Interviewer instruction

The manipulation in this research starts with the different trainings in the different probing tactics. As described in section 3.2.1.2, the experimental variable, therefore, is not the probing tactic as it was *conducted* but as it was *instructed*. However, if the interviewers from the main study were able to probe according to the probing tactic as instructed, as the interviewers in the pilot study had been, the difference between the instruction and the conduction would disappear. This issue will be discussed and assessed in Chapter 5.

3. Design and Methodology of the Data Collection

By design, all other interviewing behaviours, such as interviewing style, introducing the interviewing, posing questions and even probing techniques were identical for all interviewers, no matter what probing tactic they were assigned to. In other words, the most important possible confounders in interviewing behaviour were kept constant.

In both the instruction manual and in the training itself, the logics of the probing tactics⁶ are used in the instruction of the probing techniques. There are two rationales for this combined discussion of the probing tactic, its logics and the probing techniques. The first rationale is to normalise the probing tactic to the interviewer. The second rationale is that by doing so, all probing techniques are drenched in the logic of the tactic, helping the interviewer to get trained in the correct stance and the implicitness or explicitness of that stance.

3.4.2.2 Training

As explained above, the interview training was stretched over two days in comparison with the pilot study. This was done for three reasons. The first reason is that training six people instead of three makes practicing twice as long. Second, the training was set up to be diverse in its didactical form; instead of just instruction, time was made for watching interview video's, role-playing and discussing the role-plays. The third reason was that the pilot study had shown that the time for practicing and feedback is really necessary and needs to be fully utilised.

I trained the interviewers. For the first batch of training groups, 1, 2 and 3, a 'sidekick' assisted me on the first training day. The role of the sidekick was mainly to help set up the role-plays and give feedback on the role-plays. Sometimes the sidekicks helped me by asking questions for clarification. There was no difference between groups with or groups without a sidekick for the content of the training.

Like in the pilot study, the trainings for the different groups were largely the same. The trainings only differed on the probing tactic. However, all probing techniques were only approached and explained from the logic of the particular probing tactic in which the interviewers were trained. Other functions of probing or other argumentations were ignored. For instance, accommodating interviewers were told that giving positive comments on the answers of interviewers is suggestive interviewer behaviour, while challenging interviewers heard that these positive comments are too weak to provoke the respondents, and instead they had to give critical comments. Encouraging interviewers were told that positive comments are perfect for motivating a respondent.

Training day one

The pilot study had also shown that motivation and team spirit is crucial for a good training and motivated interviewing. Therefore in the training, much effort went into

Probing Behaviour in Open Interviews

teambuilding and motivating the interviewers. The first training day started with a team building game, in which the interviewers had to introduce their neighbour after a short exercise in careful listening. Just as in the pilot study, we had teambuilding group lunches during the training days.

First the interviewers were introduced to the study, the organisation of the fieldwork, the probing tactic they were going to use and some very general goals of the research. This was followed by an introduction to open research interviews and the purposes of open research interviews, rapport and the personal interviewing style. The idea behind this extensive introduction was to thoroughly introduce the interviewers to the goals of interviewing, to ensure that their improvisation was in line with the substantive goals of the interview.

Only after this general introduction did the interviewers lay their eyes on the interview. We started with the introduction and some role-playing on setting the atmosphere and introducing the interview.

Following this introduction, the topics of the interviews were extensively explained. All questions on the questionnaire were explained, discussed and practiced in role-plays. Special attention was paid to asking each question in a natural way, so the questions, reactions and responses of the interviewer did not seem 'over-scripted'. Naturally, the interviewers were forbidden to alter the meaning of the question. Therefore, the main technique they were trained on was paraphrasing previous answers to introduce the next question, for example, "You just told me that you've good contact with your neighbours, do you have allochthon neighbours?"

After learning how to ask questions, the interviewers were taught how to probe. The probing tactic was explained as a general stance to be sustained throughout the interview and, as described in section 1.5, to be solely focussed on the content of the answers, not on the person. All different probing techniques were approached from the stance characteristics from the probing tactic in question. Thus, in the accommodating and encouraging probing tactic trainings an active silence was approached as offering an opportunity and time to disclose, while in the challenging probing tactic trainings, it was explained as a confrontational device. With all groups, most of the afternoon was spent in training and practicing the various probing techniques to be used for challenging, encouraging or accommodating the respondent.

The first day ended with how to make the decision to stop probing. Again, this was more or less the same for all tactics: after the respondent seemed to reach saturation, the interviewer had to ask whether the respondent had anymore to say and then offer a summary of the previously given answer. Only if the respondent replied that the answer was complete could the interviewer move on. Every interviewer was instructed to use this

stop-criterion with every topic. The interviewers were told not to use the stop-criterion for just any questions, but only for the important ones.

Training day two

The second training day started with an explanation of the recording devices. The recording devices were Philips Digital Voice Tracers 7680. Since the audio is crucial for the research, the interviewers received an extensive training on how to operate the device. Throughout the day, all interviewers practiced using their Voice Tracers.

After that explanation, the main points from the previous day were reiterated. Role-plays were held again, and this time the respondents were asked to behave 'difficult', such as being talkative, dominant or shy. All probing techniques were practiced and discussed, again with a strong focus on the probing tactic.

In the afternoon, practice interviews were held. These practice interviews were exactly the same as a normal interview except that the respondent was coming to a room in the university, and the interviewer was being watched through a closed video circuit by all interviewers and me (see for a more extensive argumentation Uhrenfeldt, Paterson, & Hall, 2007). The respondents were students and PhD-students unknown to the interviewer. The respondents received a five Euro incentive. In this adjacent room, the performance of the interviewer was closely followed and criticism and praise were given. After the interview, the other interviewers and I gave feedback on the performance of the interviewer. Sometimes parts of the interviews were replayed, so the interviewers could review themselves. Every interviewer did a practice interview that was analysed by five peers and one or two experts. This *real* practice interview turned out to be very educational, both for the interviewer and those observing.

3.4.3 Questionnaire

For any interview using scripted questions, the necessity of pretesting the questions and the questionnaire is indisputable. In this research, most pretesting of the questions had been done in the pilot study, as was already discussed in section 3.3.5. The last few questions that were added later have been assessed for their quality through colleague reviews in addition to three test interviews with students.

An important lesson learned from the pilot study was that the three topics ought to be connected. Therefore, in the introduction, in the questions and in the topic introductions, links between the different topics were made.

Introduction of the questionnaire

Obviously the questionnaire starts with an introduction (Box 3.2). The introduction is short, but clear and contains the necessary elements.

Box 3.2 Introduction of the interview

<p>In dit onderzoek houden we interviews naar wat mensen vinden over Amsterdam en Amsterdammers, over de sociale contacten en over verschillende bevolkingsgroepen. Het gaat om open vragen, zodat u uw eigen mening of verhaal goed kwijt kunt. De interviews blijven anoniem, dat betekent dat wanneer mensen het onderzoeksrapport lezen, zij niet uw naam te weten komen. Het wordt opgenomen, zodat ik niet zoveel hoeft te schrijven. De eerste vragen gaan over Amsterdam en uw buurt.</p>	<p>In this research we conduct interviews on what people think of Amsterdam and Amsterdammers, on social contacts and on different population groups. The interview consists of open questions, so you can express your opinion or tell your stories properly. The interviews will be kept anonymous, which means that when people read the research report, nobody will find out your name. It is recorded, so I do not need to write that much. The first questions are on Amsterdam and your neighbourhood.</p>
---	---

From the pilot study it was apparent that respondents felt it was important that they know what the interview would be about. Therefore, the introduction carefully presents the three main topics, without mentioning friends and allochthons explicitly, yet hinting with vaguer terms. This has been done to provide for the possibility that respondents answer questions to the Amsterdam and friends parts of the interviews from an allochthon frame of reference.

The presentation of the topics is followed by an explanation of the question types and indirectly urges the respondent to speak frankly. Naturally the anonymity is explained, and respondents are made aware of the audio recording device.

Questionnaire parts and questions

In total there are 28 scripted questions in the questionnaire. These 28 questions can be distinguished in six parts. The first part contains questions on Amsterdam, Amsterdammers, the neighbourhood and the neighbours, (question 1 to 9). The second part is a series of questions on the topic of friends, (question 9 to 15). The third part consists of questions on the third topic, allochthons, (question 15 to 20). Then two questions (20 and 21) are posed on the ‘Wij-Amsterdammers’ policy. The fifth part contains the background questions (22 to 26), and the last three scripted questions are part of the after-talk.

Questions on Amsterdam and Amsterdammer

The first 8 questions on Amsterdam and Amsterdammers are presented in Box 3.3. Questions 2a, b and c are possible alternatives, meant to encourage respondents to talk. The interviewers were free to select any of them, as long as they spent some time questioning the respondent about living in Amsterdam.

Box 3.3 Amsterdammer questions

Question 1	Hoelang woont u al in Amsterdam?	How long have you been living in Amsterdam?
Question 2a	Hoe vindt u het om in Amsterdam te wonen?	What do you think about living in Amsterdam?
Question 2b	Zijn er dingen die u prettig vindt aan het wonen in Amsterdam?	Are there things you find pleasant about living in Amsterdam?
Question 2c	Zijn er dingen die u onprettig vindt aan het wonen in Amsterdam?	Are there things you find unpleasant about living in Amsterdam?
Question 3	Voelt u zich een Amsterdammer?	Do you feel Amsterdammer?
Question 4	Wanneer vindt u iemand eigenlijk een Amsterdammer?	When do you consider someone an Amsterdammer?
Question 5	En wat is typisch voor een Amsterdammer?	What is typical for an Amsterdammer?
Question 6	Woont u met plezier in uw buurt?	Do you enjoy living in your neighbourhood?
Question 7	Heeft u goed contact met uw burenen?	Do you have good contact with your neighbours?
Question 8	Heeft u allochtone burenen?	Do you have allochthon neighbours?

Question 3 is the question that is aimed at provoking the respondent to think about social categorisation of Amsterdammers. This question, as well as 4 and 5 are bookmarked as important questions and are central to this topic. Question 6 and 7 are added to invite the respondents to get talkative. Question 8 already hints at the later topic of allochthons.

Questions on friends

The questions designed for analysing the social categorisation of friends (Box 3.4) are the second central topic for this research. The purpose of questions 9 and 10 is comparable to that of question 3, which is to provoke respondents to think about a category that they have possibly taken for granted. If someone has no friends, question 10 could be skipped, but question 11, 12, 13 and 14 could still be posed, since even if someone has no friends, one could discuss the category. Question 13 introduces a standard category to which friends are compared. Question 14 is comparable to question 8, but in this case functions more directly as a bridge to the introduction of the final topic.

Box 3.4 Friends questions

Question 9	Heeft u vrienden in Amsterdam?	Do you have friends in Amsterdam?
Question 10	Kunt u aangeven hoeveel vrienden u in Amsterdam heeft?	Could you indicate how many friends you have in Amsterdam?
Question 11	Wanneer noemt u iemand een vriend?	When do you call someone a friend?
Question 12	Vindt u dat er verschillen zijn tussen typen vrienden?	Do you think that there are differences between types of friends?
Question 13	Vindt u dat er verschillen zijn tussen een kennis en een vriend?	Do you think that there are differences between an acquaintance and a friend?
Question 14	Heeft u allochtone vrienden in Amsterdam?	Do you have allochthon friends in Amsterdam?

Questions on allochthons

The third topic is the controversial topic in the interview. In the pilot study, some respondents reacted rather fiercely towards the topic and complained that the topic of allochthons was not related to the topic of Amsterdammers; therefore a topic

introduction was scripted. This topic introduction turned out to be successful in the pilot study.

Box 3.5 Allochthon questions

Topic introduction A	Amsterdam is een stad waar veel allochtonen leven. Daarom wil ik het nu graag daarover hebben.	Amsterdam is a city where many allochthons live. Therefore, I would like to talk about that.
Question 15	Waar denkt u zelf aan bij de term allochtonen?	What do you have in mind when thinking of the term allochthons?
Question 16	Zijn er volgens u verschillen tussen allochtonen en autochtonen?	Are there according to you differences between allochthons and autochthons?
Question 17	Zijn er volgens u verschillen tussen allochtonen onderling?	Are there according to you differences among allochthons?
Topic introduction B	Mensen hebben heel verschillende gevoelens over allochtonen;	People have very different feelings about allochthons;
Question 18	Als u denkt aan allochtonen, wat voor gevoelens roept dat bij u op?	When you think of allochthons, what kind of feelings come up?
Question 19	We hadden het net over Amsterdammers, kunnen allochtonen Amsterdammers worden?	We were just discussing Amsterdammers; can allochthons become Amsterdammers?

In comparison with the first question asked about the other two topics, the first question for allochthon is not provoking, but is directly descriptive. This is intentional because I wanted a description before they gave an evaluation that would lead the respondent to answer all questions from that frame of reference. Comparable to question 12 and 13 for friends, the respondents are asked to discuss allochthons internally and in comparison with autochthons, the common sense category used to pair up with allochthons.

Since this topic is so controversial, a second topic introduction has been scripted, presenting the possibility to utter any evaluation in question 18. With question 19 the circle is closed again, as Amsterdammers and allochthons are connected to each other.

Questions on ‘Wij-Amsterdammers’

Half a year before the research, Amsterdam was shocked to witness publicist and filmmaker Theo van Gogh murdered by an Islamist with Moroccan origins, from Amsterdam. In the days, weeks and months afterwards, the local government was frightened for inter-ethnic or inter-religious tensions and even conflicts. In this period of fear, much focus has been on the action programme ‘Wij-Amsterdammers’ (We-Amsterdammers), launched by the local government. This action program had four goals: fight terror, discourage radicalisation, prevent polarisation and mobilise positive powers in society. The means to pursue these goals was by encouraging a ‘We-feeling’ among different groups in Amsterdam; by strengthening social contacts and ties between different ethnic or religious groups.

Since the three topics of the interview and the Wij-Amsterdammers action programme have similarities, these questions on the programme were added, again to frame the three topics into one overarching topic.

Box 3.6 Wij-Amsterdammers questions

Question 20	Kent u het actieplan 'Wij Amsterdammers'?	Are you familiar with the action programme 'Wij-Amsterdammers'?
Question 21	Wat vindt u van dat actieplan?	What do you think about that action programme?

Background questions

The four background questions were deliberately placed at the end of the questionnaire, so as to prevent respondents from being influenced by the style of the questions in the interviews and then only offering factual and superficial responses. The personal background questions are rather standard: profession, educational level, voting behaviour and age.

Questions during the after-talk

After the final background questions were answered, the interviewers had to close their ring binder and pose three more questions. These questions were on what the respondent thought of the conversation, the questions and the interviewer. Even these questions were scripted in advance, so the answers to these questions could also be compared.

3.4.4 Selection and assignment

As was shown in section 3.2.2.4, this research contains a dependent structure, in which interviews are nested under interviewers. In such a structure, it is important that the first level is as large as possible. Therefore, it was decided that 36 interviewers were to be selected. More would be too difficult to train and manage, while less would result in less power.

The number of interviews per interviewer was set at six. This makes the total number of interviews to be held in this study 216.

The assignment of the interviewers to the six training groups had to be unsystematic except in regards to gender. To restrain a possible confounder of gender differences in instruction groups and therefore in the different tactics, gender was equalised between the groups. All instruction groups consisted of four female and two male interviewers. The assignment of the respondents to interviewers was random.

3.4.5 Selection of interviewers

Since I chose to use inexperienced interviewers, recruiting the 36 interviewers was rather easy. To homogenise the interviewer population in order to test whether differences in interviews were actually the results from differences in probing tactics and not from interviewers, I selected students. Students, generally, are about the same age, they have about the same educational background, about the same intellectual capacity and about

the same interests and often come from similar social backgrounds. Due to the topic of the interviews, the interviewers, by design, all had to be 'autochthons'. Since social science students outperformed the other interviewers in the pilot study, I chose for social science students.

Since most social science students are women, I decided to enrol 24 women and 12 men. The students were recruited through personal contacts: e-mails to lecturers of several departments asking them to spread the word to students, flyers, posters and an announcement on the social science faculty education web pages.

The people that reacted to the flyers were students or recent graduates. All students and recent graduates that applied for the interviewing jobs were asked about their ethnic background and their interviewing experience. The selection was very strict concerning ethnic background, since one of the interview topics is social categorisation of allochthons. Other selection criteria included: communicative skills, year of study (at least in the second bachelor year) and having only basic interview experience.⁷ In total, about 70 people applied for the interview job. Of these 70 applicants, 36 were selected based on the above criteria, on their date of application and on the dates that they were able to join in the training.

Eventually the 36 selected interviewers followed the training and conducted their interviews. The average age of the interviewers was 24.7, and the standard deviation was 2.8. There are no significant differences in the distribution of the interviewers over the three probing tactics for the variables: experience in interviewing, educational background, age and of course gender.

Of all interviewers, 34 interviewers were able to interview at least six people. Two interviewers could only hold five interviews. Two interviewers did extra interviews, but these are left out of the analysis.⁸ Therefore, the practical N is 214 interviews.

All interviewers received 20 euro per interview plus 40 euro per training day if all 6 interviews were completed.

3.4.6 Selection of respondents

While it is possible to perform a random sample from a population, all practically and financially possible samples lead to different problems, such as large numbers of non-response for a simple random sample or a very convenient response of interviewer acquaintances in a quota sample. Therefore, the most sensible selection method is to ask respondents from an access panel if they are willing to be interviewed. Besides, as was previously stated in section 3.2.2.5, nowadays the use of access panels is also the most common selection strategy used by survey research institutes.

Furthermore, because of the restrictions on certain characteristics of the respondents, due to the topics chosen, selecting respondents through an online panel was

3. Design and Methodology of the Data Collection

the only possibility. It would be rather awkward to interview someone who is an allochthon according to the CBS-definition, about feelings towards allochthons, so a selection on this characteristic would be essential as well. The choice for Amsterdam was made for two reasons. First, it is logistically better, and second, it was expected that people were eager to talk about being Amsterdammers.

The respondents came from an online panel of O+S Amsterdam, the research and statistics bureau of the Amsterdam municipality. All members of this panel received a questionnaire for other purposes, and the last question of the online survey was the following:

Box 3.7 Question posed to access panel members O+S Amsterdam

Wat is volgens u nou een Amsterdammer?	What is an Amsterdammer, according to you?
O+S Amsterdam werkt samen met de Vrije Universiteit (VU) aan een interviewstudie met daarin o.a. bovenstaande vraag. De interviews gaan over Amsterdammers, contacten in de stad en de multiculturele samenleving.	O+S Amsterdam and the Vrije Universiteit are cooperating on an interview study with the above question. The interviews are on Amsterdammers, contacts in the city and the multicultural society.
Het onderzoek loopt van 9 mei tot 14 juni 2005. Het interview duurt maximaal een half uur. U hoeft de deur niet uit, want de interviewer zal na een telefonische afspraak bij u langskomen. U krijgt een vergoeding van vijf Euro in de vorm van een VVV Iris-cheque.	The research will run from May 9th up to June 14th 2005. The interview will take half an hour maximum. You do not need to go out, because the interviewer will come to you after a telephone appointment. You will receive a shopping coupon worth 5 euro (Iris-cheque of the tourist information board).
Wilt u meewerken aan een interview?	Do you want to cooperate in an interview?
Ja ⇒ Wat is het telefoonnummer waarop de interviewer u kan bereiken voor een afspraak? Nee	Yes ⇒ What's the telephone number the interviewer can contact you at for an appointment? No

There was a very quick response and with 313 telephone numbers collected, O+S took the question offline. The question was unfiltered, which resulted in the fact that from the 313 respondents, 54 were allochthons. After filtering them out, the total amount of volunteers was 259.

These 259 people were randomly (using Microsoft Excel randomiser) assigned to interviewer numbers. Most 'non-response' was due to being unable to make an appointment; less than 10 were mere turndowns.

To check for the representativeness of the sample in this research, the 214 respondents are compared to official statistics of (autochthon) Amsterdammers. Four variables are compared, the first of which is gender.

The sex ratio (m/f) of the 'autochthon' population in Amsterdam is 0.94 (O+S, 2005). Whereas of the 214 interviews held, 113 were with females (sex ratio 0.89). So women are slightly overrepresented in the sample.

Table 3.2 Age distribution of the respondents in comparison to the age distribution of 'autochthon' Amsterdammers.

Birth Cohort	Number of respondents	Percentage of total number of respondents	Percentage of total number of 'autochthon' Amsterdammers #
Upto 1930	2	0.9%	10.8%
1930-1940	11	5.2%	9.1%
1940-1950	35	16.6%	14.4%
1950-1960	65	30.8%	16.8%
1960-1970	57	27.0%	20.5%
1970-1980	38	18.0%	21.8%
1980-1985 ¹⁰	3	1.4%	6.7%
Total	211	100.0%	100.0%

#Source for the 'Authochthon' Amsterdammers: O+S (2005)

In Table 3.2., the second variable that is used to compare the sample with official statistics is the age of the respondent is presented. I compared the birth cohorts of the respondents with the percentage of 'autochthon' Amsterdammers born in the same cohorts. From the table, it is clear that in the sample, the respondents that were born in the 1940's, 1950's and 1960's are overrepresented in comparison to their respective age groups in Amsterdam. Respondents from the 1920's and the 1980's are rather underrepresented in the sample.

An assumption about online access panels is that certain political preferences are overrepresented, while others are underrepresented in self-registration panels. For the sample in this study, this turned out to be partly true, as can be seen from Table 3.3.

Before comparing the results, three remarks need to be made. The first is that the 2006 results are for the entire population of Amsterdam voters and not only for autochthon Amsterdammers. The second is that the official statistics are the results of real elections, whereas in this research, some respondents chose a political colour or two parties instead of a single party. The last remark is that there was a one and a half year time lapse, which was politically turbulent and led to a changed political landscape. In spite of these critical remarks, the table shows some tendencies of the respondents and the general Amsterdam population, which are interesting enough to compare.

In the sample the CDA, Christenunie and VVD voters are underrepresented. For the VVD, when adding the possible voters with the general inclinations liberal and right wing, this could possibly be representative enough. However, for the CDA and Christenunie, including the possible voters who are generally inclined to vote for a Christian party, the sample is still not representative. The respondents are rather left wing when compared to the general Amsterdam population in the election. Furthermore, Groen Links is highly overrepresented among the respondents.

Table 3.3 Answers to the question on most likely party the respondent would vote for if parliamentary elections would be held now, compared with the population results of the parliamentary elections in 2006

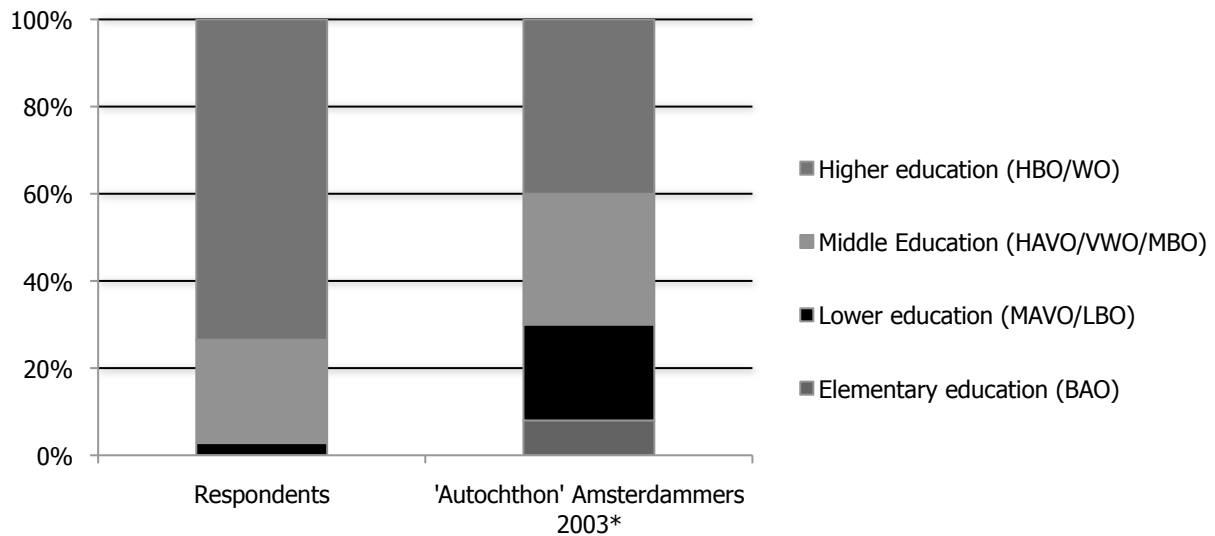
		Frequency	Percentage	Results of the Parliamentary elections 2006 for the Amsterdam population in percentages #
Parties	CDA	4	1.9%	7.1%
	Christenunie	1	0.5%	1.1%
	VVD	15	7.0%	10.3%
	D66	9	4.2%	3.3%
	PVDA	47	22.0%	22.4%
	SP	27	12.6%	13.8%
	Groen Links	36	16.8%	9.3%
	Wilders/PVV			3.3%
	SGP			0.1%
	Partij voor de Dieren			2.6%
	Other Parties			1.0%
General inclinations	Liberal	4	1.9%	
	Christian	1	0.5%	
	Right wing	1	0.5%	
	Left wing	25	11.7%	
	Does not know yet	30	14.0%	
	Not Voting or Missing	14	6.5%	25.6%
Total		214	100.0%	100.0%

#Source for the Amsterdam results of the parliamentary election 2006: O+S (2006)

No respondent indicated plans to vote for Group Wilders (his party, the PVV, did not exist yet), the SGP or The Partij voor de Dieren. The combined category for non-voters or simply missing data account for only 6.5%, whereas for the general population of Amsterdam, the percentage of non-voters was already low with 25.6%. Quite a large percentage of respondents (14%) were in doubt about what party to vote for. However, this percentage could not explain the differences between the population and the sample. The explanation I find most plausible is that the self-selection for the access panel, as well as for the interview, are conducive to more left wing and 'politically active' (in the sense of indicating to vote) applicants.

For the fourth variable to compare the sample with the Amsterdam autochthon population is shown in a figure.. In Figure 3.1 two columns are presented, showing the distribution of the education level among the respondents and the general population of 'autochthon' Amsterdammers.

Figure 3.1 The distribution of the education level of the respondents (n=211) and the 'autochthon' Amsterdammers



#Source for the 'Authochthon' Amsterdammers: O+S (2004).

From Figure 3.1 it is clear that among the respondents, almost 75% were highly educated, whereas for 'autochthon' Amsterdammers, this is 40%. Only 6 respondents had a lower education, whereas *no* respondents had only an elementary education.

To conclude, it is clear that the respondents active in this research are generally of higher education, more left wing and more politically involved than the population of 'autochthon' Amsterdammers.

3.5 The interviews

The interviews were held at respondents' homes, within two weeks after the training. All interviews were recorded using digital recording devices. In the home situation, the respondents generally felt at ease, and none complained about the audio recording of the interviews. Only once did a technical (or personal) failure lead to the loss of important information.

After 2 or 3 interviews, the interviewers returned for feedback. This was usually three to four days after the training. Two interviewers returned after the fourth interview.

The feedback usually lasted about one hour, sometimes more. The feedback consisted of feedback on the organisation of the six interviews, the interviews and interviewing in general, along with very specific feedback on some parts of the interview. This was done through first downloading the audio files from the Voice Tracer to the computer. Then, after setting the play head of the audio player to somewhere in the beginning of an interview, we listened to a few minutes of the interview, after which, I commented on it. This was repeated for several parts of the interview.

3. Design and Methodology of the Data Collection

Most feedback concerned the amount, necessity and types of probes, the probing tactics, suggestive probing and the amount of off-topic talk. During every feedback session, remarks were made on the interview timing, since the instructed 30 minutes seemed very hard to stick to. Interviewers tended to interview longer than that.

After finishing all six interviews, the interviewers returned the questionnaires, which were only used for jottings, all post-interview debriefings¹¹ and the Voice Tracers.

Notes

¹ There are other alternatives for controlling for confounders, by either using case-control studies or performing many small-scale laboratory experiments with different control groups, both to establish the effects of as many factors as possible. This however will lead to many (costly) experimental groups and a low ecological validity.

² <http://www.cbs.nl/nl-NL/menu/themas/dossiers/allochtonen/methoden/begrippen/default.htm?ConceptID=37> (Accessed on April 25th 2008)

³ Using the definition of Statistics Netherlands.

⁴ Due to illness of one of the professional actresses, a colleague PhD-student was called in, who did not have a background in acting. She followed the complete training for the Accommodating tactic with the other two interviewers and conducted four interviews.

⁵ Software created by Thomas Muhr at Scientific software. For the coding and analyses in this research version 5 was used. See www.atlasti.com for more information.

⁶ See section 1.5.3 for the description of these logics.

⁷ One professional military interrogator (who studied social sciences) was turned down, as well as many students who were too young or too old. One exception was made for an older German born student, aged 36. He was so interested and so enthusiastic that he joined in.

⁸ One female interviewer left one respondent shortly after starting the interview because the respondent made her feel very uncomfortable. She interviewed another respondent on her list. Another interviewer did one extra interview because he accidentally made an extra appointment, and he loved doing the interviews. The extra interview and the interview with the scary man are not considered in the analysis.

⁹ O+S uses the category label Nederlanders (Dutch), when considering Autochthons.

¹⁰ Since the respondents had to be 18 years and older, in the sample, the youngest respondent was born in 1986. However, the official statistics are only available every 5 years. Therefore, this individual is left out of the analysis. The age of two respondents is missing.

¹¹ When the interviewers finished an interview, they filled in a small self-evaluation form. The results of this self-assessment are shown in Chapter 5.

4 Design and Methodology of the Data Analysis

4.1 Data? What are the data?

It might seem rather peculiar to start an introduction to a chapter on the methodology of data analysis with a question on the very nature of the area under discussion. But posing the question ‘What are the data?’ is not as peculiar as it seems. In this research asking ‘What are the data?’ is a different question at various levels of analysis. On these different levels, the form of data is different and has its own difficulties. In this chapter, all levels of data are discussed, and it is shown that the peculiar looking question is the very reason to spend an entire chapter on the methodology of data analysis.

Interviews as data?

This research is on the effects that probing tactics have on the quality and the content of the information obtained through an interview. This means that the interview process is the area under analysis. However, it is impossible to unobtrusively observe an interview in a home setting, regardless of whether or not the observation itself would be detailed, valid and reliable enough for drawing any conclusions. As it is impossible to analyse interviews directly, one needs a recording of the situation, so it can be played over and over, for greater reliability, detail and evidence.

Recordings as data?

So, if the interviews themselves are not the data under scrutiny, maybe the audio-recordings are. As was shown in section 3.4.2.2, much effort was put into assuring the correct use of the Voice Tracer recording device, so that the interaction was captured correctly.

The problem with an audio recording is that it is selective, even in an interview. During the interaction, things occur before and after the recording is turned on, and non-audible interaction, which is to say, body language, disappears completely. Since the interviewers in this research recorded the complete interviews, including the introduction and the after-talk, most relevant discussion has been recorded; thus in Sacks’ words, “other things, to be sure, happened, but at least what was on the tape had happened” (Sacks, 1984, p. 26).

Recordings, however, are not the data used in this research, since the analysis of digital recording was due to technical limitations of the Qualitative Data Analysis Software available in 2006 unproductive.¹ Besides, analysing written text is much faster and still more reliable, because it could be read at various speeds and repeatedly as needed.

Transcription as data?

So, in this research, audio recordings are not considered to have direct analytic utility. However, to create a transcript, transcription as an interpretative action is performed. Therefore, transcription is a form of analysis. Section 4.2 shows how this analysis has been done as reliably as possible.

The most interesting analysis however, is in the coding, which takes place *on* the transcripts.

Codings as data?

The production of codes (coding) is, just like transcription, a form of both analysis and data creation. While coding the transcription, any analyst *interprets* the transcript, the codes and the coding rules and *applies* the coding rules. By performing both activities, the analyst creates new data in the form of codes and codings (the application of the codes on the transcript). The simpler a coding system, the less interpretation is necessary, and the more coding becomes an application of rules. The more in-depth and multifaceted a coding system becomes, the more interpretation will be involved. The application of rules could easily be checked by using reliability analysis. However, the interpretation is probably best performed in discussion. Naturally, both activities occur at exactly the same time and are only analytically distinguishable.

Numbers as data?

The final level on which data in this study could be analysed is on the level of numbers. By converting the codes to variables (mainly a technical application of logic rules), a new level of data is created. This level is, mostly in aggregated form, central to the answering of the research question as well as the analysis of the reliability.

4.2 Transcription

In this research, the key variables are constructed through coding the transcripts from audio-recorded interviews. This means that not only is a high quality audio recording crucial for the quality of the transcript, but the quality of the transcript itself is equally crucial for the reliability of the analysis. Transcripts that miss parts of an interview could lead to different conclusions.

In conversation analysis, the role of transcription has been problematised. Hutchby and Wooffitt (1998) treat a recording as a 'reproduction' of an event, whereas a transcript is considered to be just a 'representation'. However, Ashmore and Reed (2000) would hold that the preference for a transcript over a recording is dependent on the phase that the research is in. In order to make this representation as realistic as possible and to

standardise detailed transcriptions, Gail Jefferson developed a transcription ‘system’(see for the summary Jefferson, 2004).

In their work on discourse analysis, Potter and Wetherell (1987) show how decisive the format of a transcript can be for the interpretation given to it. They compare verbally transcribed audio without detail on intonation and interaction to a Jeffersonian transcription, pointing at enormous differences. Almost 20 years later, Potter and Hepburn emphasise their perspective with even stronger words: “the analysis of broader patterns and ideological talk should be able to deal with the specifics of what is going on in the talk, rather than simply a reconstructed distorted and version of it” (Potter & Hepburn, 2005, pp. 8-9). Based on conversation analysis and other forms of interpretative thinking, others, such as McLellan, MacQueen and Neidig (2003) and Bird (2005) stress the link between transcription and analysis and offer a warning about approaches that act as if transcription is just some technical issue *before* analysis. Following their precaution, I treat the transcripts as fallible representations of the recordings that are the closest form I can get to ‘reality’. To increase the reliability of the transcription, each transcription was extensively checked. To randomise errors, many transcribers were enrolled.

The digital audio files were transcribed by students or recent graduates. Again, they were contacted through various channels; some were interviewers, others were found through colleagues, an advertisement on the faculty website and through snowballing. The interviewers were contacted first, since they knew the research, knew the interviews and knew me, so they might have wanted to help. Sixteen of the 36 interviewers transcribed interviews for me; some did just one, while another individual transcribed 19 interviews. Former interviewers were never asked to transcribe their own interviews.

Every transcriber received a document with instructions for the software and the transcription rules. The main rules are presented in Box 4.1.

Box 4.1 The main transcription rules

- Turn taking; interviewer turn ‘I:’, respondent turn ‘R:’ and other peoples turns ‘A:’ or ‘B:’ or ‘C:’.
- Overlapping talk; placing the overlapping parts of both speakers between straight brackets []. Indentation is not necessary.
- No formatting.
- The complete interview had to be typed out, including pretalk, after-talk and off-topic talk.
- The transcript should be verbatim, including ‘uhm’s’, ‘uh’, ‘mh mh’s’, ‘phh’ et cetera.
- The spelling should be correct.
- All forms of humming are transcribed as ‘mh’ or ‘mh mh’.
- All forms of ‘uh’ and ‘eh’ are transcribed as ‘uh’.
- Transcriber comments are placed between curved brackets: ‘(knock on door)’.
- A notable pause is transcribed by ‘(silence)’.
- Laughter is transcribed as ‘(ha ha)’ or ‘(he he)’.

These rules are not as detailed as Jeffersonian transcription rules primarily because the research question does not require a fine grained analysis of the interaction. Besides, it

would be impossible to transcribe 136 hours in such a detailed manner. However the rules are still far more detailed than those used for most transcription in qualitative research.

The transcribers transcribed the recordings at home. The audio files were downloaded through a secured website and transcribed using Express Scribe² for playing the audio files and Microsoft Word for typing them. After finishing an interview the transcriber e-mailed it to me directly.

After receiving the first transcribed interview, I thoroughly checked the interview by listening to the audio file, while reading the transcript completely to increase the reliability of the transcription and to check for the correct application of the transcription rules. Each transcriber received feedback on his or her first interview before he or she was allowed to transcribe a second interview. Some transcripts were returned to the transcriber with my remarks and a few corrected pages, so the transcriber could correct his or her transcription.

All later interview transcripts were at least partly checked, by randomly placing the play head on the timeline and checking a few minutes from that point forward. This was repeated at least twice per interview. If irregularities were found, another part was checked as well. If the interview contained too many errors, I either corrected the transcription myself or sent it back to the transcriber. All interview transcripts were corrected on formal errors, like incorrect formatting, mistakes in turns and overlap.

The transcribers were paid per minute of the interview. This worked well, since the transcribers worked with concentration and efficiency. Furthermore, due to the intensive checking, they delivered good quality.

In total, 28 people transcribed interviews. Two transcribers only transcribed one interview. Five transcribers were dismissed (or quit themselves) before their fifth interview, since their work kept containing too many errors. All their interviews were corrected either by themselves, another transcriber, or me. One transcriber finished ten interviews and handed the transcriptions in together. These ten transcriptions were very poor and were thus returned to the transcriber. After his corrections, the transcriptions were still poor and were therefore sent to a good transcriber, who corrected them all.

Eight other transcribers did more than 10 interviews. They were all very good transcribers. The maximum number of transcriptions done by one transcriber was 22.

4.3 What are the units of observation and analysis?

Related to the question posed in section 4.1, is the question of what the units of observation and analysis are. It might be clear that for the different levels presented above, different units of observation or units of analysis are to be recognised. For instance, the unit of observation for transcribing an interview is every single sound the

interviewer or respondent makes. Every sound, then, has to be interpreted and transcribed; thus, in transcribing, the unit of analysis is also every sound.

Generally when developing a coding scheme, every analyst, both in qualitative and in quantitative research, has to define the boundaries for the units of observation. In most qualitative content analysis, the size for every unit is decided on each time a code is used to describe a certain piece of text. It follows that every time a coder codes a text fragment, he or she has to decide on the exact size of the unit of analysis as well. The theoretical boundary, however, is always restricted by decisions made before the coding. In quantitative content analysis, due to the necessary comparability between the coded fragments, the exact size of these fragments is also decided on *before* the actual coding.

For the comparability in this research, the size and boundaries of all fragments were predefined, in conjunction with the development of the coding scheme. I made the decision to code the interviews on the level of speech turns. Therefore, the boundary for a unit of observation is every turn switch between the speakers in the interview. For example, there is an observable boundary when the interviewer finishes a question and the respondent starts uttering, since the interviewer turn stops, and the respondent turn starts. The advantage of this decision is that interviewer and respondent turns are clearly distinguishable, and for the coding, two sets of codes could be used. So, the unit of observation is the coding unit, which is the interviewer or respondent speech turn.

The unit of analysis used in this research, however, is often not the single turn. In most cases, it is a part of the interview about a certain topic or the entire interview. It simply depends on the relevant question to be answered. Every switch in unit of analysis will be clearly marked.

4.4 Coding

The problem of a coding system for 214 open interviews is exactly that: a coding system for *214 OPEN* interviews. There seems to be a trade off between a *comparable coding system*, necessary for the experiment and an *in-depth, rich, complex coding system*, necessary to analyse open interviews that contain qualitative data. The latter is simply impossible; to be able to analyse so many open interviews efficiently, a reduction is indispensable. The former may lead to a superficial and less informative coding system.

An extra difficulty that goes along with creating the coding system is that the system should be applicable to the three different topics: Amsterdammers, friends and allochthons. Last but not least, the coding system should be reliable and trainable to coders, while at the same time it should account for the complexity of open interview interactions.

On the bases of transcripts and audio collected as part of the pilot study, a coding system was developed to tackle these difficulties. Central to the coding system developed

here is the use of three main sets of codes that are used to facilitate the analysis and to answer different questions. Altogether, the main sets are as parsimonious as possible, while still giving opportunity for in-depth analysis. The codings are highly comparable and potentially rich.

The first main set of codes comprises the administrative or *general codes*. The purpose of these general codes is mainly to facilitate coding, later analysis and to distinguish important parts in the interview transcripts. The second main set of codes was created to analyse the *interviewer behaviour*, the experimental variables. These codes are applied to the interviewers' questioning and probing behaviour. Here the coding unit is the interviewer turn. The third main set of codes is called the *respondent answer codes*. These codes are applied to respondent turns. The respondent answer codes are used to answer the main question of this research: what effects do the probing tactics have on the quality and the content of the information obtained through the interview?

The software chosen to generate the codings and analyse parts of the data was ATLAS.ti. At the time of the pilot study, other software for Computer Assisted Qualitative Data Analysis (CAQDAS) was not as advanced and versatile as ATLAS.ti. Therefore, this choice was relatively simple.

The coding of the interviews took place in January, February and early March, 2006. In the months prior to the actual coding, this phase was organised by planning the coding, assigning transcriptions to the software, general coding and with technical preparation. Almost all general codes were created and inspected by me. Using these general codes, I created two general project files³. In one, coders could read the complete transcripts but only code interviewer turns for interviewer behaviour. In the other, coders could read the complete transcripts as well but could only code the respondent turns for the respondent answer codes.

The coding was planned to consist of two phases; the interviewer turns were to be coded in January, and the respondent turns in February. The two phases were identically planned; they started with instruction, including a manual and a codebook, a practice interview and group discussions for inter-coder agreement.

The planned time for the whole coding operation was about 1100 hours. In order to finish this within a proper time span, seven coders were needed to work approximately 20 hours a week. Less coders would increase the possibility of risks associated with negative effects from a single coder and would require a longer time span and/or more hours per week. More coders would be more difficult to train, and the coding period would have been either uncontrollably short or longer but less intensive for the individual coders. In other words, the period and the intensity of time, the manageability and the risks decided how the balance between *intra-coder* (stability) and *inter-coder* reliability was maintained.

All seven coders received a file containing 33 or 34 transcripts. Three interviews were used for reliability testing and training purposes. The other 30 or 31 of these were assigned using a systematic non-random selection with an interval of 7. The sequence of the transcripts to be coded was randomly determined.

The coders and I shared two adjacent rooms at the university, so the coders could discuss problematic codings, and I could supervise the coding.

4.4.1 General codes

The first main set of codes, the general or administrative codes, was mostly created before the coding period. These codes could be divided in three types. The first type is a set of codes that is created on the *level of the interview transcript* as unit of coding. These codes function as variables on the complete interview⁴. Examples are age and gender of the respondent, interview number and interviewer number.

The second type of general codes contains codes that are used to label different parts of an interview. For example, the part of the transcript on Amsterdammers was selected and turned into a record⁵ that could be coded. The codes linked to the parts of the interview were extremely helpful for the organisation of the coding and for later analysis on the *parts of the interview* level. All parts of the interviews were coded.

To filter irrelevant parts of the interview, the coders have coded off-topic parts of the interview using the rules for relevancy, as described in section 2.3. The off-topic parts were not coded with the interviewer behaviour codes or the respondent answer codes, but later, the number of turns that were enclosed by the off-topic parts was used for the relevancy quality criterion.

The third type of codes consists of codes on *interviewer and respondent speech turn level*. Turns were selected as records and thus became codeable units of observation. Due to the transcription rules, it was possible to automatically create these records by searching and automatically coding the interviewer turns and the respondent turns. Alongside the interviewer and respondent turn codes, some paralinguistic utterances were also 'auto-coded'. Other automatically coded speech features include overlaps, silences and laughter (also on turn level).

The turn level was the main level of coding. All interviewer behaviour and respondent answer codes were used on this level. A single turn could be coded with as many codes as necessary. Hardly any interviewer turns were left uncoded, whereas respondent turns containing no codeable information, but were still relevant interview talk, were left uncoded.

So, through the general codes, the three units of observation and analysis are created: the general variables for the interview-as-unit, the records for the parts-as-unit and the records on the lowest level, the interviewer and respondent turns-as-unit.

4.4.2 Interviewer behaviour codes

The second main set of codes used in this research is related to interviewer behaviour. These codes are used to determine whether or not the interviewers followed the instructions. Two sets of codes are sufficient for answering this question. The first set contains all codes on the interviewer mistakes, while the second set includes all codes on probing techniques.

The first set contains all codes that are used for the analysis of *interviewer mistakes*. These are codes on mistakes made in questioning and probing. The codes on *questioning mistakes* relate to *question presentation* and *question number*⁶, both checking for reformulations of the scripted questions and even for skipping questions. Obviously the questioning mistakes are only used on the turns containing the scripted questions (or the reformulated versions of it).

For the *probing mistakes*, the focus lies on four types of inappropriate interviewer behaviour: Giving *suggestions* (including the *introduction of terms by the interviewer*) showing *agreement* with the statements of the respondent, giving *wrong clarifications* and making *two or more requests in one turn*. The probing mistakes could be used on any interview turn. In this coding set, the codes are not mutually exclusive, for example an interviewer turn could contain both an agreement and a wrong clarification. For a schematic presentation of all interviewer mistake codes see Box 4.2.

Box 4.2 Schematic representation of the first set of interviewer behaviour codes: interviewer mistakes

Interviewer mistakes	Questioning mistakes	Question presentation Question number
	Probing mistakes	Suggestion (including Term Interviewer) Agreement Wrong clarification Two or more requests in one turn

The second set contains codes based on the *probing techniques*, as described in section 1.4.1. All codes for probing techniques are presented in Box 4.3. Most of these codes for probing technique are not mutually exclusive; this means that in a single turn an interviewer could use several probes. What most often occurs, is that interviewers start with a hum and follow that with another probe. If a probing mistake is phrased within a probe, the turn could be coded with both probing technique and probing mistake codes. When two probes of the same type of probing techniques are used within one turn, the code for that technique is only used once. If it is not a simple repetition, but a double-barrelled probe, it is also coded as a coding mistake. When two or more different requests are used within one turn, this is automatically a probing mistake.

Two codes that are ‘not probing techniques’ are included in this set. These are codes for behaviour that is used by the interviewer to ensure rapport: *meta-remark*, which is a remark on the interviewing situation and *bridge to a new topic*.

Box 4.3 Schematic representation of the second set of interviewer behaviour codes: probing techniques

Probing Techniques		(Active silence) ⁷
	Minimal probes/ continuers	Echo Hum Negative comment Positive comment
		Unfinished question / sentence
	Question repetition	Question reformulation Question repetition, same meaning Question repetition, verbatim
		Follow-up question
	Requests	Request for elaboration Request for specification Request for specification by example Request for specification by contrast Request for specification of own opinion Request for reasoning Why-probe Request for experiences Request for feelings
	Expressions	Expression of doubt Expression of lack of comprehension Expression of comprehension Question of comprehension Expression of perceived inconsistency
		Summary/ paraphrase
	Reflections	Reflection following respondent Reflection countering respondent
	Non-probing behaviour	Meta-remark Bridge to a new topic

Obviously there are some minor differences between Box 1.2 and Box 4.3. The first is that active silence is not considered for analysis⁸. Second, the comments and expressions are a little bit more specific. Third, in Box 4.3 the requests are categorised and differentiated from the follow-up question. Although theoretically the requests for reasoning, experiences or feelings are follow-up questions, for coders it was expected to be clearer to group them under requests, since all requests were also coded for specificity.

To identify the specificity of the request, which means whether the request is undirective or directive all types of requests were coded with a small group of codes. If the request is directive, the *term* used to make it directive could either come from a previous answer of the respondent or from a preceding question from the questionnaire. If the interviewer introduced the term, it is counted as an interviewer mistake (see section 5.2.2). The codes are schematically represented in Box 4.4.

Box 4.4 Schematic representation of codes for the specificity of the request (directive or undirective)

Specificity of request	Term from answer respondent Term from question No terms used (undirective)
-------------------------------	--

In addition to the codes for the specificity of the requests, the coders had to code all *requests for feelings* and *experiences* for the *direction* of the probes. Interviewers could ask for feelings and experiences without mentioning a *direction*, or using a *positive* or *negative direction*, or even both (see Box 4.5).

Box 4.5 Schematic representation of codes for the direction of request for feelings and request for experience

Direction of request for feelings and request for experience	Negative Positive and Negative Positive No direction mentioned
---	---

Perhaps, it seems surprising that there is no set of codes to code for the probing tactics used in the interview. There are two reasons for this. The first reason is that I did not want the coders to guess which tactic was used during the interviews. They might simply read and code according to their private hypotheses. Secondly, by looking at the probing techniques the interviewer used, an analyst should be able to deduce the probing tactic in which the interviewer was trained. This will be tested in Chapter 5.

4.4.3 Respondent answer codes

The *respondent answer codes* are the key important codes for answering the research question. To analyse the effects of probing tactics on the quality and content of information, the line of reasoning from Chapter 2 is followed. There are three sets of respondent answer codes: answer codes, personal information codes and membership categorisation codes

The first set of respondent answer codes are simply labels for the contents of the different answers provided by the respondent. This is exactly what most researchers would do, when coding answers from open interviews. These *answer codes* were mostly created prior to the coding phase. Not all questions were intended to get relevant information, so only the answers to relevant questions were coded.

The second set of respondent answer codes consisted of codes on *personal information*. The most important of these, the personal information code is used as the indicator for *depth* in the interview, as described in section 2.3. The coders used this code on every turn in which the respondent gave some new information about his or her personal life or feelings. During the pilot study, a strong correlation was found between the variable based on codings for each fragment of personal information, on the one

hand, and the variable based on turns containing personal information ($r=0.81$ $p<0.01$ ($n=27$)) on the other. Therefore, only the code on turn level was used. Other codes for the personal information were too specific or unreliable and were not used in the analysis.

Last but not least, the final set of the respondent answer codes is based on membership categorisation analysis and used to analyse the *amount*, *specificity*, and *elaborateness* of information produced in the interview. The *membership categorisation codes* consisted of two subsets: *category labels* and *predicate types*. All codes are schematically represented in Box 4.6.

The coders had to code each respondent turn for the *category labels* that were *explicitly used by the respondent*. If category labels were not previously used in any other interview, the coders created new *category labelled [codename]* codes. Therefore, the first subset of membership categorisation codes consisted primarily of inductively created codes. The coders were allowed to create new category label codes in cases when the respondent not only mentioned a category label, but also used a predicate bound to it. When respondents had previously used a category label and then referred to it later with an index, such as ‘them’ or ‘that people’, the turn was coded with a *category indexed [codename]* code.

The second subset of membership categorisation codes concerned the predicates that respondents attached to a specific category. This subset was developed beforehand, partly based on experiences in the pilot study. From the pilot study, it proved unproductive to create predicates inductively and analyse quality of the responses with them. When coding the friends part of the interview, this led to hundreds of codes, after only a few interviews. These codes were also hard to classify at a higher level, since they varied across too many dimensions. Moreover, for the other topics in the interview, this would have led to even more codes. Therefore, a classification principle was created beforehand. The category bound predicates were clustered in *predicate types*, which were partly based on literature on membership categorisation (e.g. Sharrock, 1974; Watson, 1978; Wowk & Carlin, 2004), partly on the experience in the pilot study and completed with some predicate types that were developed during the coding phase.

A special predicate type is *relational feature*. This type was developed during the coding phase, to account for the difficulty that many respondents had not only with predicated categories, but also (and quite frequently for friends) with the relation between categories. In membership categorisation analysis literature, many authors mention friends as an example of a membership categorisation device (Nikander, 2000; Psathas, 1999). Some authors even show how it is invoked when categorising (Rapley, McCarthy, & McHoul, 2003; Sacks, 1972). However, although in membership categorisation there is a strong focus on the relational aspects of categories, no term was developed to account for the predication of relations rather than categories⁹. Consequently, this predicate type

Probing Behaviour in Open Interviews

is special, since it is used on the relation with or between categories rather than on the category itself.

The predicate types were used in every instance a predicate type was used within a turn. So, if a respondent used three different category bound activities, the predicate type activities code was labelled with a three.

Box 4.6 Schematic representation of the second set of respondent answer codes: membership categorisation codes

Membership Categorisation Codes	Category Label	Category Labelled [codename] Category Indexed [codename]
	(Category Bound) Predicate Types	Activity Appearance Beliefs Birth Locality Competences Culture Economic Feature Educational Feature Other Features Feelings Knowledge Locality Preferences Relational Feature Rights & Obligations Time
		Number of times a Predicate of a certain type was used within a turn

4.4.4 Selection of coders

The coders were either known to me or recommended by colleagues. They were selected for being serious, reliable and smart students. The team consisted of 7 people at first. Five finished their master degree recently, and all were social scientists: two sociologists, one psychologist, three organisation anthropologists and one anthropologist.

At the end of coding the files on interviewer behaviour, two coders quit coding. They were replaced with two anthropologists, of which one was still a student. This team stuck together until the end.

4.4.5 Training of the coders

The coders received two trainings. As with the training of the interviewers much stress was put on teambuilding and getting across the important ideas rather than only providing rule based coding instructions. I gave a very short description of my research without explaining the probing tactics. More stress was put on interviewer behaviour and the content of the answers.

4. Design and Methodology of the Data Analysis

Unsurprisingly, the basis of the training was a coding manual, which contained explanations about the code sets, a codebook, software instruction and theoretical background.

This short explanation of the research was followed by an introduction in ATLAS.ti; the main concepts, the terminology and the functions important to the coders were explained. All coders received documentation on the software and the ATLAS.ti manual was available.

After this instruction, the coders received an explanation of the different sets of codes. All codes were explained, beginning with the notions underlying them and their contextualisation within their sets. There were no references made to the different probing tactics. From the files, the coders could not read which tactic was used in which interview.

For the coding phase, I found it very important to create a good atmosphere, since the coders would have to work together for two months in a job that would be tedious at times. Therefore, group lunches and teambuilding exercises were held throughout.

The afternoon of this training was used to explain the various codes in vivo, using examples from interviews and ATLAS.ti for practicing coding.

The first interview transcript was identical for all coders. This transcript was individually coded and was then thoroughly discussed in small groups, under my supervision.

The second training was on the coding of the respondent answer turns. This training was much less technical but far more theoretical. Again, it started with an outline of the research and the general function of coding the answers of the respondents. Three different sets of codes were introduced, each with a set of instructions: the answer codes, the personal information codes and membership categorisation codes.

The answer codes as well as the personal information codes were fairly easy to explain, and this coding was expected to be very swift and easy. Thus the main focus was on the last family of codes, the membership categorisation codes. Since membership categorisation analysis requires a different approach to the interview transcripts, the training for this set of codes was more theoretical as well.

The planned sequence for this phase of coding was to begin with the answer codes and the personal information codes for each interview and to then proceed with the membership categorisation codes to code the same interview.

Again, all coders practiced first, by coding an identical interview transcript. The differences between the individual codings were compared and extensively discussed.

4.4.6 Coding process

4.4.6.1 Coding process, phase 1: interviewer behaviour codes

The coding in January 2006 had an unexpectedly quick start. Most coders found it hard to toggle between coding the quality of the questioning and coding the probing techniques used by the interviewer. Therefore, all coders first coded all turns of the interviewers with regard to the number of the questions asked and the quality of the questioning. This turned out to be an excellent training exercise for acquiring basic knowledge of coding in ATLAS.ti.

After finishing coding all questions and the questioning quality, all coders received a table, from which they could see which questions were missing in their coded interviews. From this table, they checked their own codings.

Almost the entire interview was coded for interviewer behaviour. This meant that the coders coded every speech turn of an interviewer, starting from the introduction and continuing until the after-talk. Naturally the focus was on the probing, but since most utterances coming from interviewers are part of the probing, all turns would be coded.¹⁰ See the screenshot in Figure 4.1 for an example of the way the coding on the interviewer behaviour took place practically.

The coding of the interviewer turns took less time than expected. However, there was quite a varied range of coding speeds among the different coders. One coder did not fully complete her own file and was helped by one other coder, who finished four randomly chosen interviews. The fastest coder needed 50 hours to complete all of her interviews, while the slowest coder took 85 hours for the same amount of interviews. The slow coder, however, performed his task very conscientiously, so that checking his work was hardly worth the additional effort.

Some errors in the coding and the fast average speed of the coding of the interviewer turns led to the decision to have all coders first check their own file after finishing and then check another coder's file. This was a bit overcautious because it took a lot of time and did not lead to many corrections. The coders checked the files in two ways: they conducted an integral check of an interview, and they used combinations of codes useful for searching specific incidences¹¹ to check for common errors. After the peer verification, I checked all files again using different queries of codes. Most files took one or two hours to check.

4.4.6.2 Coding process, phase 2: respondent answer codes

To prevent the necessity of these extensive checks in the second phase, the group work was intensified for the second coding phase. Moreover, instead of discussing one practice interview and one interview towards the end, the coders had to code three identical interviews; they coded one at the start of this phase, another one-sixth of the way into it and the last when they were one-third of the way through this phase. These interviews were thoroughly discussed afterwards.

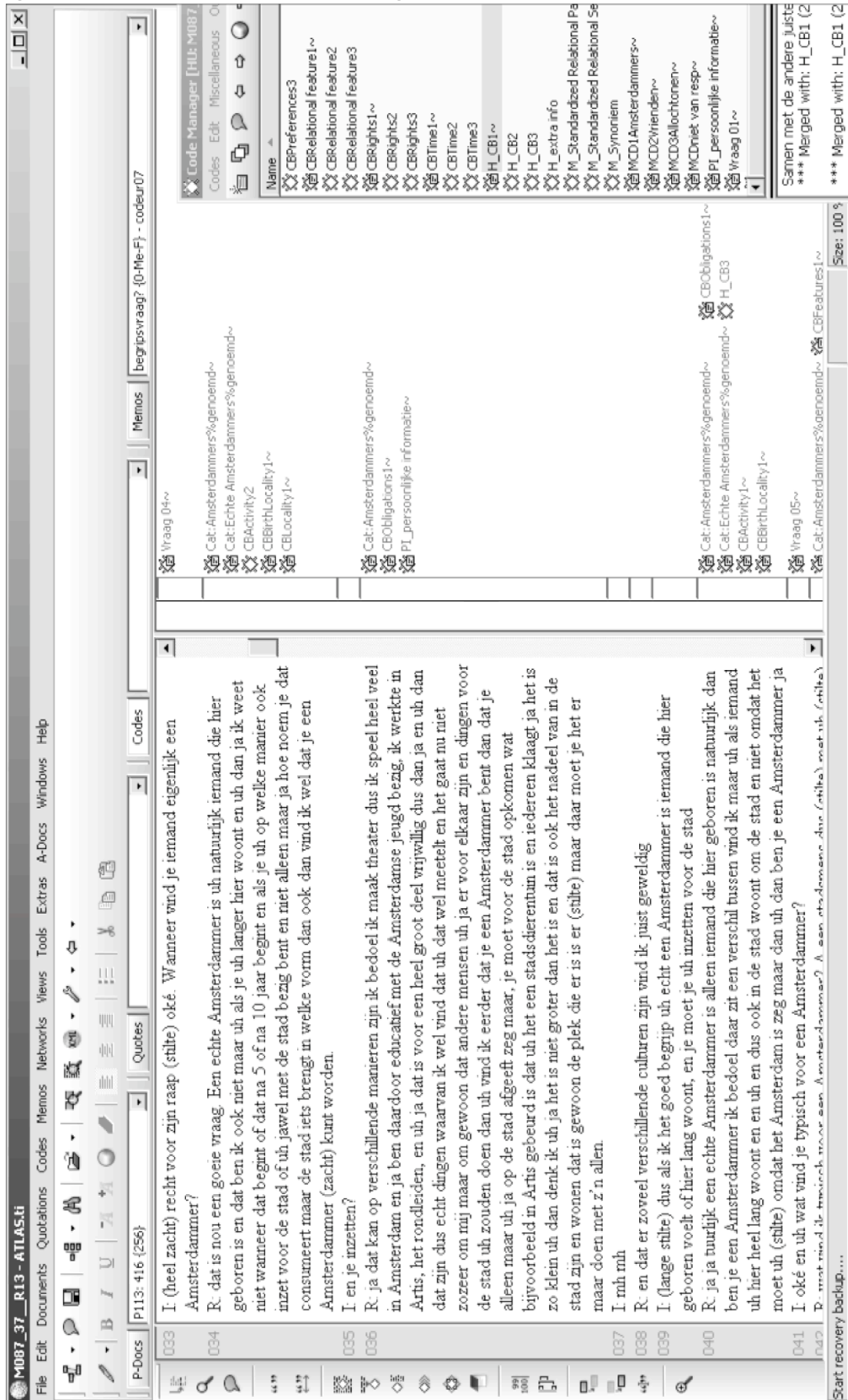
The planned sequence for the coding phase was to use the different code sets consecutively per interview. This meant that the coder had to go through the same interview twice, first using the personal information codes and the answer codes and afterwards again using the membership categorisation codes. The reason is that while toggling a code set is technically simple, toggling a mindset turned out to be very confusing.

So, the coders started with the personal information codes and the answer codes and coded all interviews in accordance with the random sequence that was distributed among them. Only after finishing all interviews they coded using the membership categorisation codes, following the same sequence. In Figure 4.2 an example screenshot of ATLAS.ti with both personal information and membership categorisation codes are shown.

Unfortunately, a mistake in the definition of one of the main codes of personal information codes led to some uncertainty among some of the coders. The mistake was corrected in the training, but combined with the difficulty to toggle ones mind between personal information codes and membership categorisation codes, this led to quite a slow start in coding the respondent turns. However, after resolving these errors definitively, the personal information codes turned out to be very swift and easy to code. The application of the answer codes did not cause any serious problems.

The most exciting part of the coding phase was coding the membership categorisation codes. The category labels and even the indexes to the categories were fairly easy to code. However, coding the predicate types turned out to be somewhat difficult. The coders could easily detect a predicate when used. However, it turned out to be hard to interpret which predicate type was used. As planned, some types of predicates were added, and this helped noticeably. Usually, the coders easily detected the repetition of predicates that were used.

Figure 4.2 Screenshot of ATLAS.ti showing respondent answer codes



4.5 Reliability

In this research, inter-coder-reliability plays an important role, since all analyses and thus conclusions heavily rely on the coding that is done. In the different analyses conducted in this research, resemblances of both quantitative content analysis and qualitative data analysis could be found. Both research traditions have their own way of tackling reliability issues. In this research, I combined both traditions to increase and test for reliability.

4.5.1 Reliability tests

The quantitative content analysis tradition for accounting for inter-coder-reliability is through post-hoc tests of reliability, such as Cohen's Kappa and Krippendorff's Alpha (Krippendorff, 2007). Using a measure such as Krippendorff's Alpha, the agreement between multiple coders could be established, and chance is taken into account.

In this tradition, very strong emphasis lies on reliability as reproducibility of the codings on the basis of the coding instruction. To get this reproducibility, it is extremely important that the coders have independently coded their documents.

In my research, the coders coded the two interview transcripts for the interviewer behaviour codes and the three interview transcripts for the respondent answer codes independently. So for establishing the inter-coder-reliability coefficient, the criteria were met (Krippendorff, 1980). However, the coders could be aware of the fact that all coders had to code these interview transcripts, so it is presumable that they concentrated more on their codings. For the first transcript this awareness was higher than it was for the later transcripts, because with the later some of the coders only noticed it after coding the transcripts independently.

Since I have chosen for the conservative Krippendorff's Alpha, I find a coefficient above 0.6 acceptable (Lombard, Snyder-Duch, & Bracken, 2002) and above 0.8 excellent. A coefficient above 0.9 is probably merely indicative of the simplicity of that code.

4.5.1.1 Reliability of interviewer behaviour coding

As discussed above, all seven coders had to code two identical interviews for the interviewer file *independently*. These independently coded interviews are tested for their reliability using Krippendorff's Alpha for nominal codes (Krippendorff, 1980, 2007).¹² The reliability analysis is conducted for all codes and all interviewer turns, except for the introduction and the after-talk. In total, 343 turns were analysed.

The results of the coding reliability test for the compared coded transcripts are comforting. Out of all 37 codes that were used in these interviews by any user, twelve codes scored below 0.6. Eleven codes that scored below 0.6 are each used less than four times in these 343 cases. When a code occurs only once or twice, the effect of a mistake leads to extreme results in the Alpha's. The effect of one mistake is naturally higher in the

cases with an average below four than it is when codes occur often. Besides, coders are less trained in linking codes to situations that happen to occur less often, and thus, they frequently choose different codes.

Since the first interview was coded directly after the training, some coders still misinterpreted some coding rules. For instance, two coders also used the code positive comment for 6 turns containing just a “ja” (“yeah” or “yes”), which resulted in a rather low (0.1) coefficient. Additionally, “I: Ja” turns were also seen as suggestion (0.3) or agreement (0.2) at times¹³.

This mixing up is probably the reason for the low score of reflections (0.4); they are often mistaken for summaries (0.8) and vice versa.

For some codes, the Alpha coefficients are very high. Sixteen scored above 0.8. Of course, some codes are easy to apply, such as the question numbers and the quality of the presentation of the questions (all above 0.95). Why-probes (1.0) as well as requests for examples (1.0) are also easy to determine, since both have particular words to be used in such a probe: ‘why’ and ‘example’. Coders clearly found these easy to use.

Other requests are more difficult to determine. The request for elaboration however, still has a high Alpha coefficient of 0.9. The coders were very agreeable on whether a probe was a request or anything else.

Since all hums were transcribed using hm hm (or sometimes accidentally mh mh, mhmh or hmhm), interviewer hums were coded automatically.

4.5.1.2 Reliability of respondent answer codes

For testing the reliability of the respondent answer codes, I selected three interview transcripts to be coded by all coders. On the codings of the three interviews, several Krippendorff’s Alpha’s were calculated. The Alpha coefficients are calculated by using all possible respondent turns within the parts to be coded. So for the Personal information codes and the membership categorisation codes, this applied to all turns of the respondents in the Amsterdammer, friend and allochthon parts. In total these were 484 units.

The result for the *depth* indicator is rather impressive: The inter-coder reliability for personal information is 0.9, whereas the application of this code seemingly uses a lot of interpretation.

As explained in section 4.4.3, the set of membership categorisation codes consists of two subsets: The predefined but rather broad predicate types and the ‘in vivo’ created category labels.

The predicate type codes are analysed for their Alpha’s on the three interviews, by combining all occurrences of a specific predicate type within one turn. For instance, whether an activity is mentioned once, twice or even six times within the same turn, it is

in this case only labelled as activity. Fifteen predicate types that are used in these interview transcripts (out of all 16) have coefficients above 0.6. One coder erroneously used an instance of educational feature, whereas no other coder did.

The category label and index codes were not predefined. This could have led to very different categories, since much interpretation, or at least formulation on behalf of the coders, is needed. However, the Alpha coefficients calculated on the three coded interview transcripts are rather high for all categories (16). The coefficient was below 0.6 for only one code: the category index to autochthons.

4.5.2 Reliability improvement beyond post-hoc tests of the instruction

4.5.2.1 Improvement by group work

In contrast to the independence that is necessary for the reliability tests of the coding *instruction*, in qualitative data analysis, it is very common to code in groups and to let these groups interact to improve the reliability and validity of the coding *in practice*. In their survey of the use of CAQDAS¹⁴, Lee and Fielding (1995) found that CAQDAS is often specifically used to facilitate group work for better inter-coder reliability. Team research leads to discussions on the meanings of codes, and the outcome of this has revealed more consistent coding.

As can be read in section 4.4.5, much effort was put into increasing the coherence of the team. The main reason for building a team was not just to promote nice working conditions and create a good atmosphere in itself, but in keeping the coders sociable, cooperative and motivated, it intended to increase both intra-coder and inter-coder reliability. In a good team people dare to ask questions, discuss their doubts and give feedback to others.¹⁵

The group work consisted of three types. The first type of group work was called *jurisprudence meetings*, following Van den Berg and Van der Veer (1986), who further suggest that working this way improves the reliability beyond simple post-hoc reliability tests. In these planned meetings, the independently coded identical interviews were thoroughly examined, and all coded turns were meticulously discussed. All jurisprudence meetings were under my supervision. In the jurisprudence meetings, additional rules of application could be developed. Sometimes these rules concerned very peculiar occurrences of one code, while other times the jurisprudence formed an important addition to one of the coding rules as it had been formulated beforehand. These meetings helped remarkably to keep the coding consistent.

The second type of group work consisted of planned *'problem' meetings* in which all turns coded with a special 'problem' code were discussed. Most coders used memos to explain their doubts on how to code a certain turn. These meetings were very useful at

the start, but later on, these problems were solved by the coders themselves. These meetings only led to additional rules on very peculiar situations.

The third type of group work occurred on an *unplanned* basis and was the main reason why the second type became less important. When one of the coders ran into a problem, he or she often directly asked other coders or me how we would code it. This led to intensive interaction throughout the day. Due to the sociable atmosphere, every coder would occasionally ask the others how they would code a certain turn. Sometimes eight people decided on one turn. This does not seem to be very efficient, but most of the time coders shared their difficult turns in order to find someone else who could quickly refer to an earlier decision that was made on a similar quotation, and this was much faster than puzzling about it for minutes and interpreting it in an overly profound way. Of course these unplanned group discussions were not always the most efficient with the group's time, but I was impressed by the self-regulation of the coders.

All ad-hoc codes or additions to the coding rules that came out of any group work were gathered and distributed on paper and as memos in the coder files by me. The coders helped tremendously, especially with the membership categorisation codes, by gathering lists of turns and their respective predicate types, which were again distributed among all coders.

4.5.2.2 Improvement by checks and comparison

Besides the group work, there were three other ways that the inter-coder and intra-coder reliability of the coding were ensured and accounted for. One way consisted of *checking ones own file* on some codes and queries of codes. All coders had used the 'problem' code, and they were asked to check these after finishing the file. All coders also had to make some adjustments due to jurisprudence, and they all compared the relevant codings in their file for consistency, the intra-coder reliability. The table on questions, as was mentioned in section 4.4.6.1, was also a form of checking the coder's own coding.

The second way to account for the reliability of the coding was to have every coder check some interview transcripts of another coder at least twice. This *peer reviewing* led to discussions and sometimes corrections of the coding. The coders checking the other coders' files were also asked to identify problematic codes in the interviews they checked and to check for these as they continued checking subsequent interviews. Both ways, however, took place after the bulk of coding was done: at the end of January, for the interviewer turn codings, and in March, for the respondent turn coding. Conversely, once in a while, some coders voluntarily asked another coder or me to do peer reviewing on parts of the interview they were coding.

The third extra way of checking reliability of the coding was done by me. As I supervised the coding, *I monitored the coding* of each coder by checking their files at least

once a week. These checks were partly technical and procedural, checking the integrity of the file and the amount of codings, the sequence followed by a coder, and in part, they regarded content of the coding. If questions arose, they were discussed with the coder.

All ways to increase reliability described above are based upon group work, checks and development of coding rules. This means that some concepts were much clearer when coding the last interview transcript than they were when coding the first interview transcript. In other words, the validity of the codes grew through the development of the codes and their rules. Through the possibility of CAQDAS to track code developments, consistency comparisons¹⁶ are easy to perform across the interview transcriptions. As described above, these checks were continuously done. Therefore, I would argue that the above-presented methodology not only increased the reliability, but also the validity of the codings.

4.6 From codes to statistics

The latest part in the analysis of the data is the conversion of the codes into SPSS data files. Fortunately the ATLAS.ti export function to SPSS has improved exponentially over the years, making it possible to export hundreds of codes and 108066 coded records, such as speech turns and parts to an SPSS data file. This conversion is mainly a technical operation in which most codes are converted into dichotomous variables with the turns and parts as cases. This means that if a code is used on a turn the value is one, while when the code is not used the value is zero. In this conversion only some codes are defined as not dichotomous, but are nominal or even ordinal. For most of the further analyses, the data were aggregated to *parts of the interview level* or *single interview level* and analysed mainly using analyses of variance or chi-square analyses, as will be described in Chapter 5, 6 and 7.

Notes

¹ See for instance Alan Stockdale's contributions to the Atlas.ti forum <http://forum.atlasti.com> for many discussions on the use of audio recordings as data.

² From NCH software: <http://www.nch.com.au/scribe/> (accessed 4th December 2008)

³ Called Hermeneutic Units (or HU's) in ATLAS.ti idiom.

⁴ These variables used on complete documents are called Primary Document Families in ATLAS.ti idiom.

⁵ A quotation, which is a text fragment in ATLAS.ti idiom, and which functions as a record in a database.

⁶ Missing question numbers indicate skipped questions.

⁷ Notable pauses were transcribed by the transcriber using "(stille)". The coding of active silence was done automatically and thus 100% reliable. However, an analysis of variance and checks of the audio led to the conclusion that the transcriptions were not reliable on that part. Some transcribers scored significantly lower in recognising notable pauses; therefore, active silence has been left out of further analyses.

⁸ Notable pauses were transcribed by the transcriber using "(stille)". The coding of active silence was done automatically and thus 100% reliable. However, an analysis of variance and checks of the audio led to the conclusion that the transcriptions were not reliable on that part. Some transcribers scored significantly lower in recognising notable pauses; therefore, active silence has been left out of further analyses.

⁹ Matthews (1983) arrives at the same conclusion in interviews with older adults: people talk about friends *as particular individuals* or about friends *as relationships*. In the sociology of friendship one of the main issues since the 1980's is that the focus in much research on friendship has shifted to the relationship of friends rather than the variables attributed to individuals. According to Blieszner and Adams (1992) and extensively shown in Adams and Allan (1998), due to this shift in focus many more methodological approaches, such as network analysis studies, qualitative interview studies or ethnographies have been used to analyse friendship as a relation in context (See for the importance of context in friendship also Rawlins, 1992).

¹⁰ Within the first week, two codes for probing techniques were added to all files: *meta remark* (something like: "this interview is going fast") and *expression of doubt* "do you really mean that?". Both codes did exist in a previous version of the coding system, but the second code was accidentally left out, while the first was left out, since it seemed to occur very rarely, and thus, I planned to code these instances with a miscellaneous code.

¹¹ These combinations of codes are queries of coded quotations, using proximity or boolean operators codes and they were saved as 'supercodes' (as they are called in ATLAS.ti). For more information on queries see the ATLAS.ti manual available at www.atlasti.com.

¹² I used the KALPHA Macro for SPSS (www.spss.com), developed by Andrew F. Hayes, as described in (Hayes & Krippendorff, 2007).

¹³ As Houtkoop-Steenstra (1996, p. 219) has put it: “The Dutch ‘ja’ is highly ambiguous. This holds when used in turn-initial position in particular, and the speaker has not yet continued his or her turn. The interactional function of ‘ja’, becomes clearer once the speaker continues. A ‘ja’ may be used as an acknowledgement token, or to agree with the prior speaker. It may also be used as the beginning and the postponement of a non-agreeing action.”

¹⁴ Computer Assisted Qualitative Data Analysis

¹⁵ The second day I suggested that the coders might decorate the room if they like by putting a few posters on the wall, if they were a bit fed up. Half an hour later I returned, and the room was full of posters. This event somehow skyrocketed the idea of a team, and all coders mentioned this afterwards as extremely motivating.

¹⁶ This consistency comparison is not to be confused with constant comparison; the latter type of comparison aims at a more abstract conceptual development besides the development of codes and consistency in the coding (Glaser & Strauss, 1967; Boeije, 2002). Both types of comparisons are not only done post-hoc, but generally continue throughout the analysis.

5 Quality of Interviewer Behaviour

In this research, the input for the various probing tactics is the training as it has been described in Chapter 3. Therefore, the so-called treatment groups are based on the trainings. However, this does not mean that in the analysis I will only look at the answers of the respondents. The interview is an interaction between interviewer and respondent and this should be taken into account in the analysis. Authors from various backgrounds and with diverse perspectives have shown that focussing on the responses of respondents alone provides a one sided assessment of quality effects of interviewer behaviour (Sacks, 1995; Rapley, 2001; Van der Zouwen, 2002; Dijkstra, Van der Veen, & Van der Zouwen, 1985). Therefore, the focus of this chapter will be on the analysis of interviewer behaviour.

Quality of interviewer behaviour is harder to assess in open interviews than quality of interviewer behaviour in closed interviews because in closed interviews, deviations from the standardisation of interviewer task behaviour indicate a potential loss of quality. For example, this is the case if questions are not read as scripted, answer categories are not presented as scripted and/or the assistance in the answering is not adequate. In open interviews, the interviewer behaviour that follows the scripted question is not standardisable, and therefore, the quality of the probing behaviour is less easily determined. This is probably one of the reasons that interviewer behaviour in closed interviews is studied more frequently¹ (e.g. Dijkstra, 1983; Van der Zouwen, 2002; Foddy, 1998). However, several conversation analysts as well as discursive psychologists have qualitatively assessed interviewer behaviour in open interviews (e.g. Houtkoop-Steenstra, 1996; Abell, *et al.*, 2006).

In this chapter, there are three goals that have been set. The first goal is to describe and assess the quality of the interviewers' relevant scripted behaviour. By scripted behaviour, I mean the reproduction of all sentences and questions that were printed in the questionnaire. As was shown in Chapter 1, open interviews are characterised by a large amount of *non-scripted* behaviour, since reactions towards respondents' answers are not written on the questionnaire, and neither are any other aspects of the specific probes. Nevertheless, the scripted behaviour is relevant for the research question, since in this experiment it is important to establish if the interviewers actually posed the important questions and to determine if they presented these exactly as they were scripted; otherwise, the input before the probing would be different and less comparable.

The second goal of this chapter is to come to a *general assessment* of the quality of the interviewing in an interview, based on two aspects: a *two-sided assessment* of the interview and scores on *mistakes in non-scripted interviewer behaviour*. The two-sided assessment is done, on the one hand, through the assessment by the respondent that occurs during the after-talk, and on the other hand, through the self-assessment of the interviewer, which takes

place after the interview. The scores that are given for the mistakes made in the unscripted interviewer behaviour are determined by looking at the interviewer behaviour. The assessment of the mistakes will give answers to the following questions: How often do the mistakes occur? Will it be necessary to exclude interviews due to bad scores? For instance, an interview is not useful when an interviewer is continuously suggesting answers or showing agreement with the respondent.

The third goal of this chapter is to describe how and test whether the interviewers followed the instructions for probing. Did the interviewers probe at all? And, if so, what probing techniques did they use? Were the interviewers sufficiently able to probe according to the instructed probing tactic?

As was explained in section 4.4.2, each probing tactic should be distinguishable on the basis of the probing techniques used by the interviewer. In other words, by assessing the techniques the interviewer used, one should be able to evaluate the performance of the instructions for the probing tactics.

5.1 Quality in scripted interviewer behaviour

The first category of mistakes concerns the way scripted questions are posed by the interviewer. In posing questions, an interviewer can make two mistakes: illegitimately skipping a question or reformulating a question with a different meaning. In the following subsections, both mistakes are shown and statistically assessed.

5.1.1 Illegitimately skipping questions

In the trainings, the interviewers were specifically instructed to ask those questions that can be considered crucial for the concept of social categorisation. The total number of important questions is fourteen. On the questionnaire used by the interviewers, these questions were marked, so the interviewers were well aware of these important questions.

In some instances, interviewers could legitimately skip questions, like when the answer was really clear: If a respondent explains that he has *no* friends in Amsterdam, it is unnecessary to ask for *the number* of friends this person has in Amsterdam. All coders were instructed to code every question for the question number, *and* if a question number was missing, they had to check whether the skipping was legitimate or not.

Table 5.1 shows that illegitimate skipping of important questions is rare for these important questions.

Table 5.1 The number of important questions asked in the interviews (N=214 interviews)

Number of important questions asked	Frequency	Percentage
12	1	1%
13	4	2%
14	209	97%
	214	100%

In just one of the 214 interviews, two important questions were illegitimately skipped, while in four interviews one single question was skipped. It follows that in 97% of the interviews, all important questions were posed. There is no relation between interviewer characteristics and skipping questions, since no interviewer illegitimately skipped important questions in more than one interview.

5.1.2 Reformulation of questions

When the goal of an interviewing study is to compare the answers given in different interviews, it is ineffective to pose questions with different meanings to different respondents. In some forms of survey research, this is a deadly sin: for instance compare Fowler and Mangione (1990). In this experiment, it is a deadly sin as well precisely for the reason of comparability. Therefore, it is obvious that interviewers who present questions as scripted are performing better than interviewers who reformulate the question and alter its meaning.

All questions posed in the interviews were coded on their number, the legitimacy of skipping them, and then also on the wording used by the interviewer. The coders were instructed to label the wording per question. Questions that were presented verbatim were coded as such. Questions with minor changes in the wording, such as small additions like “uh”, “so”, were coded as reformulations with the same meaning. Both question presentations are considered ‘well-presented’. However, reformulations with altered meanings are mistakes and were coded as such.

To assess the quality of the presentation of the questions, the best measure is to look at the mistakes made in the formulation: the reformulations with altered meanings. In Table 5.2 the scores are shown.

From this frequency table it appears that in 85% of the interviews, the interviewers made no mistakes in the presentation of the 14 important questions on social categorisation. In 4 interviews, the meaning of two of the questions was altered. In 15% of the interviews, one or more wrong reformulations were made.

Table 5.2 Frequency table of the number of reformulated questions with altered meanings in the interview (N=214 interviews)

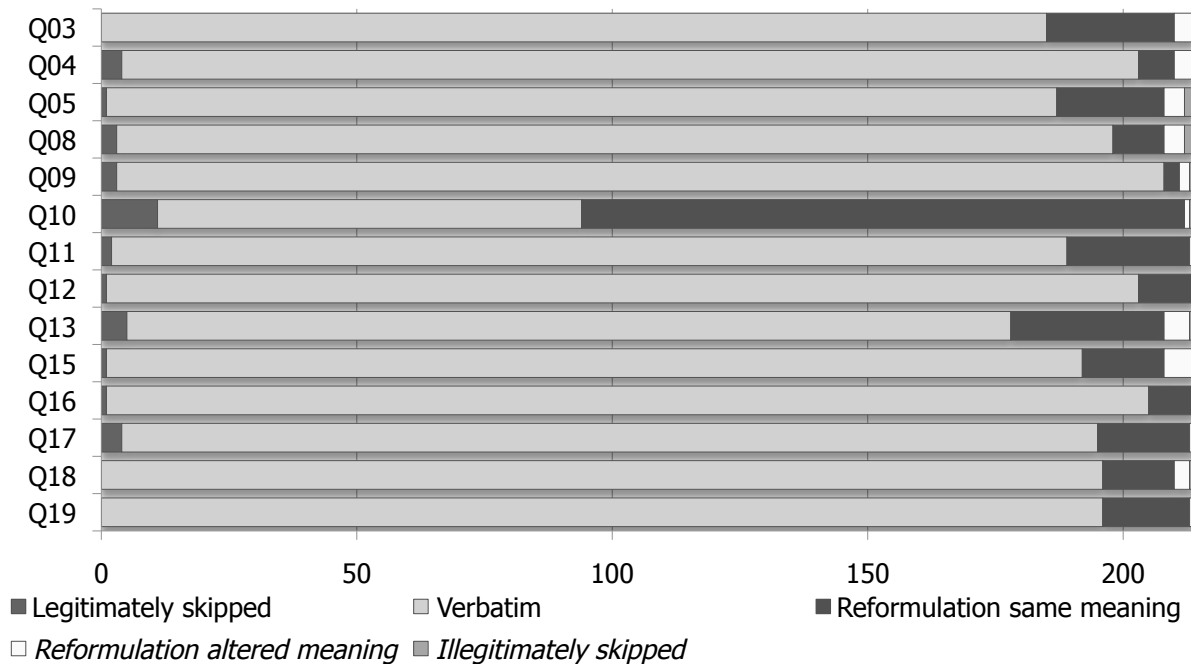
Number of reformulations with altered meanings	Frequency	Percentage
0	182	85%
1	28	13%
2	4	2%
	214	100%

These results are rather comforting; in general, the interviewers posed their questions as they were scripted in the questionnaire. To analyse which questions turned out to be vulnerable for a reformulation with an altered meaning, the analysis has to be taken to another level, the level of the question, to be precise.

5.1.3 Quality of the questioning per question

Naturally, all conclusions drawn above on the important questions are based on pooled data, on the level of *all important questions per interview*. If we take another unit of analysis and look at the different questions, the results are even more comforting.

Figure 5.1 Quality of the questioning per question of the questionnaire (N=214 interviews)



In Figure 5.1, for every important question, the quality of the questioning is depicted. The first value stands for the amount of legitimately skipped questions. The second value represents the amount of verbatim questions, and the third indicates the amount of reformulated questions with the same meaning. So these first three values represent correct interviewer behaviour. The fourth value is a reformulation of the question in such

a way that it alters the meaning. The last number represents illegitimately skipped questions.

From the figure one can see that not many mistakes were made. Of the total number of questions, only 1.5% were either skipped or reformulated with an altered meaning.

There is a large amount (118 of the 214) of reformulations with the same meaning in question 10, “Could you indicate how many friends you have in Amsterdam?”. This question was often shortened, since the preceding question already indicated that the friends talked about were in Amsterdam. So in this case the context of previous utterances was important on deciding whether the meaning was altered or not. Since Amsterdam was mentioned in the preceding question, not mentioning Amsterdam in this question, did not count as a reformulation with an altered meaning. Qualitative analyses of the reformulations with altered meanings show that such alterations are truly marginal. These questions are mostly just slightly reformulated, but strict coding rules made them reformulations with altered meaning. For instance question 13 in interview 313 turn 163 was altered from “Do you think that there are differences between an acquaintance and a friend?” into “*Because*, do you think that there are *clear* differences between an acquaintance and a friend?”

I may conclude that the deviations from scripted questions are rather minimal. If deviations occur, they are just minor, and the rate of 97.3% of well-presented questions, either verbatim (86.5%) or with minor reformulations (10.8%), is extremely good. Reformulations with an altered meaning occurred in 1.2% of the questions and only 0.3%, are illegitimately skipped. So, the interviewers behaved as instructed in asking 98.5% of the questions; hence, they performed close to perfection.

5.2 Quality in non-scripted interviewer behaviour

Non-scripted interviewer behaviour is, as could be understood from the introduction to this chapter, all interviewer behaviour that was not fully scripted beforehand, which is to say all interviewer behaviour in which some improvisation was necessary. This of course does not mean that the interviewer could just do anything he or she thought was good. Obviously there were rules marking the boundaries of tolerable behaviour, and there were recommendations on behaviour, such as what makes a good performance of the interviewing style, and then more generally what makes good interviewing. In the first subsection the recommended behaviour is assessed by looking at the self-rating of the interviewer along with the assessment that was indirectly done by the respondent. In section 5.2.2, the interviewer mistakes are analysed.

Because probing is so central to this study, I will deal with this type of non-scripted behaviour in a separate section later in this chapter, in section 5.3.

5.2.1 Assessment of the interview

5.2.1.1 Interview assessment by the interviewer

After each interview, the interviewers were obliged to fill in a debriefing questionnaire. The questionnaire contained seven questions relating to the interview that were to be answered on a five-point scale (1 to 5) and one open question that asked for “further remarks”. The purpose of each question was explained to all interviewers beforehand, during the trainings. Only three questionnaires were not handed in. The questions posed on this questionnaire are presented in Box 5.1:

Box 5.1 Questions of the post-interview debriefing

1	How do you think the interviewing went?	Not good at all	Not so good	Fair	Good	Very good
2	How do you think the probing went?	Not good at all	Not so good	Fair	Good	Very good
3	How do you think the Probing Tactic went?	Not good at all	Not so good	Fair	Good	Very good
4	Do you think the respondent was a difficult or easy person to interview ?	Very easy	Easy	Average	Difficult	Very difficult
5	Do you think the respondent was nice ?	Not nice at all	Not so nice	Fairly nice	Nice	Very nice
6	How was the rapport ?	Not good at all	Not so good	Fair	Good	Very good
7	How do you think the interview went?	Not good at all	Not so good	Fair	Good	Very good

As can be read in Box 5.1, the first three questions gradually became more specific in relation to the interviewer performance and more specific parts of the interviewer behaviour: the interviewing in general, the probing (techniques), and finally the probing tactic. The interviewers turned out to be very pleased with their own interviewing skills. Not once did an interviewer rate him or herself as ‘not good at all’ on one of these three questions. The averages² were 3.8, 3.7 and 3.6 for questions one, two and three respectively, all with standard deviations of 0.5.

The two questions on the respondent, questions four and five in Box 5.1, show that on average, the interviewers not only appreciated themselves, but their respondent as well. Most respondents were considered to be nice and easy to interview. The mean for respondent likeability is 3.9, and the mean for respondent difficulty is 2.7, so just a little closer to ‘Average’ than to ‘Easy’.

Naturally, the most interesting question for interviewing style is question six, with which interviewer-rated rapport is measured. In most of the interviews rapport was rated ‘good’ to ‘very good’ (in almost 86% of the interviews!) So, on average, the interviewers rated the rapport 4.0, which is the highest mean of all questions.

Question 7 is the final general rating of the interview. Again, the average is 3.7 with a standard deviation of 0.5. Thus the interviewers' ratings of the interviews were rather high.

Although the differences in the averages are small, it would be interesting to analyse whether there are differences in the self-rating of the different interviewers who were instructed in one of the three different probing tactics. The non-parametric rank-order Kruskal-Wallis test, presented in Table 5.3, shows some appealing results. Although the results are only significant at the 10% level, for question 5 on the 'likability' of the respondents, it shows that Challenging interviewers rated their respondents as less nice than the encouraging and accommodating interviewers. For question 6 on rapport, again, the challenging interviewers are ranked lower and thus rate the rapport lower.³ Accommodating interviewers rank their rapport highest.

Table 5.3 Mean ranks, N's and Kruskal-Wallis analyses of answers to post-interview debriefing questions. Grouping variable: probing tactic.

	Accommodating (N=70)	Encouraging (N=70)	Challenging (N=71)	$\chi^2(2)$	Sig.
1. How do you think the interviewing went?	104.6	106.1	105.8 [#]	0.04	0.98
2. How do you think the probing went?	103.3	102.3	112.4	1.68	0.43
3. How do you think the Probing Tactic went?	102.7	101.2	112.5 [#]	1.85	0.40
4. Do you think the respondent was a difficult or easy person to interview ?	105.4	101.0	111.5	1.21	0.55
5. Do you think the respondent was nice ?	110.1	113.0	95.1	5.07	0.08
6. How was the rapport ?	113.6	108.1	96.5	5.29	0.07
7. How do you think the interview went?	108.5	102.6	106.9	0.50	0.78

[#] n=70

The conclusion on the high average scores of the interviewer's assessment of the interview could mean two things: Either the interviewers overrated themselves, as Gorden would argue (see the citation on Neophyte interviewer self-rating rapport in section 1.3.3), or they were as good as they tell us they are.

5.2.1.2 Interview assessment by the respondent

When most interviewers finished the general background questions, they closed the ring binder containing the questionnaire, started packing and casually posed three extra questions. The first question was on what the respondent thought of the 'conversation';

the second was on the respondent's opinion of the questions, and the third was: "What did you think about me as an interviewer?"

The open answers to all three questions were coded on five-point scales, ranging from 'very negative' to 'very positive' (1 to 5). The last code was only applied when respondents also explained why it was 'very positive', otherwise 'positive' was chosen. The questions were not always posed nor answered in the clearest of ways, and therefore, the N is rather low.

As was the case with the interviewers' self-ratings, on average, the respondents' ratings for the conversation, the questions and the interviewers were rather high. On average, the conversation was rated a 4.2 (standard deviation 0.6), the questions were criticised more often, scoring a 3.9 (standard deviation 0.9), while the interviewer was rated the highest, with an average of a 4.4 (standard deviation 0.6).

In Table 5.4, the results of the ranking and the Kruskal-Wallis tests are shown to test for a relation between the probing tactics (as they were instructed) and the respondent rating.

Table 5.4 Mean ranks, N's and Kruskal-Wallis analyses of ordinal coded answers to after-talk questions. Grouping variable: probing tactic.

	Accommodating	Encouraging	Challenging	$\chi^2(2)$	Sig.
Conversation	76.0 (n=53)	73.7 (n=55)	75.5 (n=41)	0.11	0.95
Questions	72.4 (n=44)	69.9 (n=53)	77.0 (n=48)	0.82	0.66
Interviewer	91.5 (n=51)	81.1 (n=64)	92.4 (n=60)	2.42	0.30

The results from this test are very clear. There is no effect of the probing tactic on the respondents' 'in the face' after-talk assessment of the conversation, the questions or the interviewer. This result could be either due to politeness of the respondent that would produce such non-variance, or it could be due to the first-rate performance of the interviewers on interviewing style and interviewing behaviour.

It would be wonderful to believe that the interviewer performance was outstanding simply on the basis of the assessments of both the interviewer and the respondent. However, due to the problems in these assessments, it would be more realistic to assess the quality of the non-scripted interviewer behaviour by considering this behaviour on its own.

5.2.2 Interviewer mistakes in non-scripted interviewer behaviour

As was explained in the introduction to this chapter, the main way to assess the quality of the different interviews is through analysing mistakes. In this section, some specific mistakes that interviewers can possibly make are presented:

- Offering suggestions
- Introducing new terms
- Showing agreements
- Giving wrong clarifications
- Making two or more requests in one probe turn

These types of mistakes are rather common in open interviewing, especially by novice interviewers. For this assessment, the percentage of mistakes is calculated by adding the number of different interviewer turns to the number of mistakes and then dividing this sum by the amount of on-topic interviewer turns. If the percentage of mistakes is considered, as is done in the lowest row of Table 5.5, one can see that mistakes are rare; on average, they occur in only two percent of all turns. Interestingly, there is hardly any difference in the mean proportion of mistakes in interviewer turns over the different parts of the interview.

Table 5.5 Averages of the total number of turns of the interviewer, the turns of the interviewer containing a mistake and the percentage of mistakes per part of the interview (N=214 interviews).

Parts	Average number of on-topic interviewer turns	Average number of on-topic interviewer turns containing a mistake	Average percentage of on-topic interviewer turns containing a mistake
Amsterdam	22	0.36	2%
Amsterdammer	31	0.50	2%
Neighbourhood	8	0.17	2%
Neighbours	12	0.22	2%
Friends	35	0.73	2%
Allochthon Friends	5	0.08	1%
Allochthons	79	1.70	2%
Wij_Amsterdammers	11	0.25	2%
Entire interview	202	4.00	2%

The most frequently occurring mistake is showing agreement with the respondent, which occurs in 0.6 % of the interviewer turns.

5.3 Probing

Open interviewing is much more than just asking questions, as was argued extensively in Chapter 1. The second important component of open interviewing is probing, in terms of both asking probe questions and reacting to the answers that a respondent offers.

In probing, quality is very hard to assess quantitatively, since the situations in which probes come forward are highly complex. The reason that interviewers choose a certain type of probe is difficult to determine, as is whether another probe might have been better, and which probe for that matter. In a conversation, many spontaneous utterances are brought forth. Conversation analysis can be used to analyse such utterances quite

thoroughly, and yet still, it is likely that one could only make modest claims about the appropriateness of a selected probe.

However, to make claims on the appropriateness of interviewer behaviour in 214 interviews, it is impossible to use conversation analysis because that requires a very detailed analysis of the interaction between interviewer and respondent (See Mazeland, 2003; Silverman, 1998). Therefore, I developed the quantitative coding system of interviewer behaviour codes as described in section 4.4.2, as a more limited but suitable tool for answering the research questions concerning the probing tactics.

In the following section, I try to answer the question of what can be said about interviewer probing behaviour in the open interviews in this study.

5.3.1 The amount of probe turns

It seems perfectly plausible that an interviewer who does not probe at all, generally speaking, receives less, less specific, less elaborate and less in-depth information than an interviewer who probes in abundance. However, there are situations in which *not*-probing, or using just one single probe is the best approach to interview, such as when respondents give complete answers directly after an initial scripted question. Still, it is highly unlikely that this will occur throughout an interview. Therefore, the number of probes is important.

So, how do I count these probes? In this research, probes are operationalised as probe turns. All turns that do not contain scripted elements in them, such as (parts of) questions, topic-bridges and other scripted interviewer speech turns are defined as probe turns. Therefore, the amount of probe turns per interview is calculated by taking all interviewer turns and subtracting from this all question turns, parts of questions turns and all other scripted parts of the interview.

Since the introduction of the questionnaire does not contain questions or probes, the intro is not interesting for an analysis of the ratio of probe turns per question turn. The after-talk is very interesting, but strictly not considered to be part of the interview, since the interviewers did not have to continue using their respective probing tactics; as a consequence, the after-talk is not coded for interviewer behaviour. The background questions are not open-ended questions, but closed ended questions, in which probing is only necessary if the respondent does not answer adequately. So for the analysis of the probing, the introduction, the after-talk and the background parts of the interview are left out.

From the cell in the third column and the bottom row in Table 5.6, one can see that the average ratio of probe turns per question turn for the main part of the interview is 7.7. This means that on average, 7.7 probe turns follow every single question turn. However, many probe turns are very short. These are the so-called minimal probes and continuers,

such as “hm hm”, “Ja” or short echo’s. Instead of uttering something, an interviewer could just as well use a non-verbal continuer; therefore, a ratio (mostly) based on these continuers is not the most revealing ratio.

Minimal turns are defined as those turns in which the utterance of the interviewer contains less than 10 characters, when transcribed. So when the minimal turns are excluded, instead of using all probe turns, the average ratio decreases from 7.7 to 3.5. This means that on average, 3.5 longer probe turns follow every question.

The interviewers were instructed to probe more in the important parts of the interview. Whereas in the other parts of the interview they could stop probing at any decent point, in the important parts of the interview, they had to use the stop-criterion (see section 3.4.2.2).

In a one-way analysis of variance, there is a significant relation between the independent variable parts of the interview and the dependent variables for the ratio of probe turns and the ratio of longer probe turns. For the ratio of probe turns, the F is fairly high, although the effect size is limited, $F(7, 1693)=47.3$ $p<0.001$ $r^2=0.16$. However, the ratio for longer probe turns shows a slightly stronger effect, $F(7, 1693)=62.5$ $p<0.001$ $r^2=0.21$.

Table 5.6 Means of the Ratio of Probe Turns and the Ratio of Longer Probe Turns per part of the interview and the entire main section of the interview (N=1701 parts of 214 interviews).

Parts	N	Mean ratio of probe turns	Mean ratio of longer probe turns
Amsterdam	214	5.0	2.1
Amsterdammer	214	9.4	4.5
Neighbourhood	211	7.0	2.7
Neighbours	212	5.3	1.9
Friends	214	6.2	3.2
Allochthon Friends	209	4.3	1.9
Allochthons	214	12.2	5.4
Wij_Amsterdammers	213	5.2	2.5
Main section of the interview	214	7.7	3.5

In Table 5.6, the mean ratios for the different parts are shown. From this table, it is clear that the interviewers took the most important parts of the interview (the parts on the three social categorisation topics, printed in bold) seriously; on average, the interviewers used a higher amount of longer probe turns in the important parts of the interviews.

5.3.2 Probing techniques

In the previous section, it was shown that interviewers probe more in the important parts of the interview, and thus, they did what they were supposed to do. However, other than knowing if the interviewers used longer probing techniques and more turns containing a

probing technique, nothing has been specified about these techniques. *Which* probing techniques the interviewers used, is still to be determined.

Since the interviews are of different length when calculated by the number of turns, absolute numbers of the frequency of the use of certain probing techniques is not a clear measure. Therefore, percentages of the turns containing specific probing techniques (probing technique turns) of the total amount of probe turns are used. Naturally, the percentages are very small, since no interviewer only gives negative comments, and no interviewer will continuously request for specification.⁴

In Table 5.7, all types of probing technique turns are presented. From the table, one can see that the continuer humming is by far the most popular probing technique turn. In 16% of the cases, this probing technique turn is not a minimal turn and is thus combined with another probing technique, such as a request, within one turn.

As can be seen from Table 5.7, there was a widespread use of different probing techniques. After humming, the other probing technique turns that were most widely used are summaries and echoes. Other probing techniques are much rarer. A negative comment is extremely rare; only in 17 turns in all 214 interviews was this probing technique used.

The requests are, when considering how often they were used in an average interview, not rare at all. Requests for specification were used in almost 6 turns in an average interview, whereas requests for elaboration were used in more than 7 turns. Requests for reasons and their shortened variant, the why-probe, were used rather often as well, about 4 turns per interview, when considered together.

Table 5.7 Number of probing technique turns and the percentage of probing technique turns of all probe turns

		Number of probing technique turns	Percentage of probing technique turns of all probe turns
Minimal probes/ continuers	Echo	2520	6.6%
	Hum (including 8624 minimal hums)	10268	26.8%
	Negative comment	17	0.0%
	Positive comment	808	2.1%
	Unfinished question / sentence	1352	3.5%
Question repetition	Question reformulation	57	0.1%
	Question repetition, same meaning	191	0.5%
	Question repetition, verbatim	169	0.4%
	Follow-up question	587	1.5%
Requests	Request for elaboration	1616	4.2%
	Request for specification	1219	3.2%
	Request for specification by example	407	1.1%
	Request for specification by contrast	153	0.4%
	Request for specification of own opinion	88	0.2%
	Request for reasoning	599	1.6%
	Why-probe	287	0.7%
	Request for experiences	167	0.4%
Request for feelings	220	0.6%	
Expressions	Expression of doubt	90	0.2%
	Expression of lack of comprehension	60	0.2%
	Expression of comprehension	356	0.9%
	Question of comprehension	210	0.5%
	Expression of perceived inconsistency	40	0.1%
	Summary /paraphrase	6015	15.7%
	Reflection following respondent	981	2.6%
	Reflection countering respondent	109	0.3%
Non-probing behaviour	Meta-remark	249	0.7%
	Bridge to a new topic	710	1.9%
Total		29545	77.2% ⁵

As described in Chapter 4, the coders had to distinguish the specificity of all the request probes. From Table 5.8, one can infer that the interviewers, on average, used slightly more unspecified questions, such as “Could you tell me a little more?” than specified questions, like, “What do you mean by sports friends?” Of all requests used, on average 51% were unspecified, 20% returned to a term from the question and 28% were directive requests on terms used by the respondent.

Table 5.8 Count and distribution of the specificity of the request (directive or undirective).

		Count	Percentage of request turns
Specificity of request	Term from answer respondent	1263	28%
	Term from question	904	21%
	No terms used (undirective)	2292	51%
	Total	4459 ⁶	100%

In section 4.4.2, it was explained that all requests for feelings and experiences had to be coded for the positive or negative direction of the probe. As Table 5.9 illustrates, on average, interviewers rarely mentioned any possible direction of a feeling or experience. While trying not to be suggestive, the interviewers were even precautious in asking for either positive or negative feelings or experiences.

Table 5.9 Count and distribution of direction of request for feelings or request for experiences.

		Count	Percentage of request for feelings or experiences turns
Direction of request for feelings or experiences	Negative	21	6%
	Positive and Negative	30	8%
	Positive	24	6%
	No direction mentioned	298	80%
	Total	373 ⁷	100%

As explained above, the quality of the probing cannot be fully inferred by simply looking at the number of turns in which probes are used. Nor can it be fully assessed by the number of turns in which different types of probes are used. The quality of the probes can only be accounted for by considering the answers of the respondents because only then can we know whether the probing was accurate. This will be done in Chapter 6.

5.3.3 Probing tactics

In the instruction, the interviewers were not only trained in properly phrasing a question and in using adequate and various probes, but most importantly, they were also trained in the probing tactic that they were assigned to. Now that we have seen what probing techniques the interviewers have used, it is important to analyse whether the interviewers used the probing techniques that are specific or preferable for their particular probing tactics.

5.3.3.1 Sets of preferred techniques

The different probing tactics differed in regards to the various techniques that I explained to the interviewers as being preferable. All interviewers were trained in and practised the different techniques, but as explained earlier, through the training and the instruction manual, I kept bringing some preferred techniques to their minds. In the box below, the preferred techniques for each probing tactic are presented.

The codes for specificity of the requests are shown in italics, since these are not probing techniques but attributes of the request, which were trained as preferable for the challenging probing tactic.

Box 5.2 Preferred probing techniques per probing tactic

Accommodating	Encouraging	Challenging
Question of comprehension	Expression of comprehension	Expression of lack of comprehension
Echo	Positive comment	Negative comment
Hum	Reflection following the respondent	Reflection countering respondent
Summary		Expression of perceived inconsistency
Question reformulation		Expression of doubt
Question repetition, same meaning		Request for reasoning
Question repetition, verbatim		Why-probe
		Request for specification
		<i>Term from question</i>
		<i>Term from answer of respondent</i>

5.3.3.2 Quality of the probing tactics using multivariate analysis of variance

From these sets of probing techniques, I can test whether the probing tactics were performed as intended. By using multivariate analysis of variance for the different sets of techniques, I can both test whether the interviewers complied with the training given and see what techniques are important for defining the probing tactic in practice.

All predefined sets of probing techniques for the probing tactics differ significantly, as can be seen from Table 5.10. As a test statistic, I have chosen for the Pillai-Bartlett Trace, since it is more robust when using equal sample sizes and when violating the assumptions of both the normality of the distribution and the homogeneity of the covariance-matrix (Field, 2005, p. 594). Neither of these assumptions is met in my data.

Table 5.10 Pillai-Bartlett Trace for all sets of probing techniques with probing tactic as the predictor variable (N=214 interviews).

	Pillai-Bartlett Trace	F	Hypothesis df	Error df	Sig.	Partial Eta²
Set of preferred techniques Accommodating (N=71)	0.55	11.2	14	412	0.00	0.28
Set of preferred techniques Encouraging (N=71)	0.54	26.9	6	420	0.00	0.27
Set of preferred techniques Challenging (N=72)	0.77	12.9	20	406	0.00	0.38

The statistics show that *for all probing tactics* the interviewers significantly differed with respect to the used preferred techniques.⁸ These results show that there is a significant relation between the probing tactic in which the interviewer is trained in and the preferred set of probing techniques. However, from these statistics, we do not know which probing techniques are the decisive techniques within this set, and we do not know either if they

differed according to the expectation. For this determination, the tests of between-subjects effects are useful.

Table 5.11 shows that the decisive techniques in the first set, the set of preferred techniques for the accommodating tactic, are humming, summary and the question of comprehension: “Did I understand it correctly?” An accommodating interviewer uses a question of comprehension, on average, seven times more often than an encouraging interviewer does and 1.8 times more often than a challenging interviewer. On average, both humming and summarising are used significantly more often in the accommodating tactic. Of all probe turns used in an average accommodating interview, 22.6% contain summaries, and 36.4% contain hums. The relative frequencies of the question repetitions and the echoes do not differ significantly ($p < 0.05$) from those of other probing tactics.

Table 5.11 Statistics for the set of preferred techniques for the accommodating probing tactic (N=214 interviews).⁹

Preferred techniques for the accommodating tactic	Average percentage of probing technique turns per tactic			Tests of between-subjects effects df=2	
	Accommodating	Encouraging	Challenging	F	Sig.
Summary	22.6%	11.4%	19.3%	26.6	0.00
Humming	36.4%	20.4%	19.5%	23.9	0.00
Question of comprehension	1.4%	0.2%	0.8%	11.5	0.00
Question repetition, same meaning	0.5%	0.4%	0.7%	3.0	0.06
Question repetition, verbatim	0.7%	0.5%	0.5%	1.1	0.34
Question reformulation	0.2%	0.2%	0.2%	0.8	0.74
Echo	7.0%	7.4%	7.7%	0.2	0.79

The significant F and the high Pillai-Bartlett Trace in Table 5.10 and the high F’s in Table 5.12 seem to suggest that the ratios for the preferred techniques for the encouraging tactic are all significantly higher for the encouraging probing tactic.

Table 5.12 Statistics for the set of preferred techniques for the encouraging probing tactic (n=214 interviews).

Preferred techniques for the Encouraging tactic	Average percentage of probing technique turns per tactic			Tests of between-subjects effects df=2	
	Accommodating	Encouraging	Challenging	F	Sig.
Positive comment	0.9%	3.7%	1.0%	71.9	0.00
Expression of comprehension	1.4%	1.1%	0.5%	8.4	0.00
Reflection following respondent	1.7%	2.7%	2.9%	5.6	0.00

However, upon taking a closer look, it is apparent that the average percentage of turns that contain an expression of comprehension for all probe turns is higher with the accommodating tactic than with the encouraging tactic. Besides, the average percentage of turns with the reflection following the respondent is lower here than with the challenging probing tactic.

The only reason why the F-statistic for the set of preferred techniques for the encouraging tactic is high and significant and also why the Pillai-Bartlett Trace is relatively

high is due to the extensive use of positive comments with this probing tactic. As a crucial part of the encouraging probing tactic, this technique is the decisive technique within this set.

The set of preferred techniques for the challenging tactic again shows how almost all techniques in this set are significantly different per tactic. Besides, the means of the percentages per interview are all highest in the challenging tactic. As expected, the percentage of turns containing directive requests that use terms from respondents' answers were much higher in the challenging interviews than in those conducted with the other two probing tactics. Challenging interviewers probe more specifically, and this can also be seen from the other decisive techniques that were used, such as the why-probe, the request for specification and the request for reasoning.

Table 5.13 Statistics for the set of preferred techniques for the challenging probing tactic (N=214 interviews).

Preferred techniques for the Challenging tactic	Average percentage of probing technique turns per tactic			Tests of between-subjects effects df=2	
	Accommodating	Encouraging	Challenging	F	Sig.
Term from answer respondent	2.5%	1.9%	7.4%	102.3	0.00
Why-probe	0.2%	0.2%	2.1%	78.7	0.00
Request for specification	2.4%	2.0%	6.8%	77.3	0.00
Request for reasoning	1.0%	1.1%	3.4%	42.2	0.00
Reflection countering respondent	0.1%	0.1%	0.7%	12.8	0.00
Expression of perceived inconsistency	0.0%	0.0%	0.4%	11.0	0.00
Term from question	2.9%	1.8%	3.4%	10.2	0.00
Expression of lack of comprehension	0.1%	0.1%	0.3%	9.1	0.00
Expression of doubt	0.1%	0.3%	0.3%	4.5	0.01
Negative comment	0.0%	0.1%	0.1%	2.9	0.06

Turns containing the probing techniques expression of perceived inconsistency, expression of lack of comprehension and negative comment are seldom recorded because such utterances are so confrontational. An interesting probing technique is the expression of doubt technique. I expected this probing technique to mostly correlate to the challenging probing tactic. However, the equal means for the challenging and the encouraging probing tactics are surprising. Again, a post-hoc explanation is that many interviewers in the encouraging tactic use "Are you serious?" as a positive comment fairly often, while in the challenging tactic, the intonation for the same phrase is different.

5.3.3.3 Quality of probing tactics using a classification technique

The results presented above seem very comforting, but in my opinion, they are not convincing enough to confirm that the probing tactics were 'well-performed'. We now know that on average, the probing tactics were performed correctly in the interviews.

Probing Behaviour in Open Interviews

However, I decided which probing techniques to include or exclude in the preferred sets for the analyses.

Therefore, it would be valuable to be able to *predict* which probing tactic the interviewer was supposed to perform in a specific interview on the basis of all probing techniques that were used in the interview. This prediction is possible using a canonical discriminant analysis.

For this analysis, I used the training in the probing tactic as a grouping variable and *entered* all percentages for the turns containing probing techniques as independent variables (see Table 5.7 or Box 4.3) *together*, including the percentages for turns containing specificity: term from answer of respondent, term from question and no term used (undirective). Based on linear combinations of these variables canonical discriminant analysis finds axes that separates the three probing tactics best¹⁰. Then based on these canonical functions, it uses probabilities to classify all interviews. This classification shows a fine result: 92% of the interviews are classified according to the correct probing tactic, just from the probing techniques that are used in the main part of the interview. The classification results for the probing tactics results are presented in Table 5.14. The numbers of the correctly predicted interviews are in italics.

Table 5.14 Classification results for the discriminant analysis of all probing techniques for the three probing tactics (N= 214 interviews)

Probing Tactic	Predicted group membership			Total
	Accommodating	Encouraging	Challenging	
Accommodating	<i>65</i>	6	0	71
Encouraging	6	<i>65</i>	0	71
Challenging	4	1	<i>67</i>	72
Total	75	72	67	214

From the classification results in Table 5.14, one can see that the discriminant analysis based on the probing techniques never classified an interview that was held using the encouraging or the accommodating probing tactic as a challenging interview. It follows that the challenging probing tactic stands out as a tactic that is really different from the other two.

When inspecting the very few misclassifications, they appear to be the result of either an extreme use of minimal probes or an unexpectedly large use of a single probing technique. For instance, request for elaboration was a probing technique that was used unexpectedly often by the interviewers assigned to the accommodating tactic, so that when an encouraging interviewer used the request for elaboration technique a lot, this interview was then classified as accommodating.

5.4 Concluding remarks on interviewer behaviour

The interviewers behaved according to the instructions they received in the training. Almost no questions were skipped or altered in meaning, and if it happened, it was mostly not that important. In the non-scripted interviewer behaviour, the interviewers also hardly made any mistakes; suggestion, agreement, wrong clarifications, introduction of new terms by the interviewer and two or more requests in one probe turn hardly presented as issues.

Although the interviewers were told during the training to use any of the probing techniques when appropriate, it is natural that some probing techniques should be more preferred in a certain probing tactic than others. So, preferred sets of probing techniques for the different tactics were tested using a multivariate analysis of variance. The results were comforting, as the tactics differed significantly, meaning that the training led to significantly different interviewer behaviour. A discriminant analysis that was used for reclassifying the probing tactics led to fine results. By just using different probing techniques as variables, the linear combination of these led to 92% accuracy of the predictions. This means that the interviewers did what they were supposed to do, and the experimental variable was indeed varying.

In the next chapters, we will see what these differences lead to in terms of differences in the quality and content of the answers of the respondents. For now, we know that in terms of quality, the behaviour of the interviewers was satisfactory.

Notes

¹ There could obviously be many other reasons, for instance, another reason would be that there are simply more survey interviews available for secondary methodological research.

² Obviously these variables are ordinal and not scale variables, and thus, it is not right to simply use descriptive statistics on them. For the sake of readability, however, I chose to do so anyway.

³ The Pearson's correlation between the answers to questions 5 and 6 is $r=0.52$ ($p<0.001$).

⁴ Being less precise and not correcting for the length of the interview leads to nearly equivalent results for all analyses in this chapter.

⁵ The percentage base used here is the *number of probe turns* ($N=38\ 281$) since the variable is not the frequency of *probing techniques* used but is the frequency of turns in which a probing technique was used and thus called the *probing technique turns*.

⁶ The difference between the sum of all requests in Table 5.7 and the sum of the three specificity of the request codes in Table 5.8 is partly due to the interviewer mistake introduction of new term by the interviewer" (126 turns) and partly a result of interviewers using more than one request within one probe turn (171 turns).

⁷ The difference between the sum of request for feelings and request for experiences in Table 5.7 and the sum of the direction of request for feelings or experiences in Table 5.9 is a consequence of interviewers using both requests within one probe turn. Of these 'double barrelled request turns', no direction was mentioned in eleven interviewer speech turns, and in three turns, the interviewer probed for both directions.

⁸ For all three sets of preferred techniques, and within all three probing tactics, the interviewer effects were significantly different. Some interviewers simply performed the probing tactics on average better than others. Theoretically, a multi-level analysis of variance would be more correct, since the interviewer effects on the sets of probing techniques are significantly deviating and the interviewers are nested within one of the three probing tactics. However, since the top level contains only three values (for the probing tactics) and each interviewer only interviewed 5 or 6 respondents, a multi-level analysis is not the most informative analysis in terms of power.

⁹ The assumptions of both the normality of the distribution and of the homogeneity of variance are violated. However, according to Andy Field "In terms of violations of the assumption of the homogeneity of variance, ANOVA is fairly robust when sample sizes are equal"(2005, p. 324). For all analyses in this research, both Kruskal-Wallis tests (for the violation of the assumption of the normality of distribution) and Welch F tests (for the violation of the assumption of the homogeneity of variance) were done, leading to similar results.

¹⁰ The summary, the tests of these canonical discriminant functions and the coefficients are presented in the following tables. The first table shows the initial statistics for the discriminant analysis. The first variable accounts for 66.6% of the variances. The Wilks Lambda's in the second table are low, indicating the clear relation between the many variables (which results in this high degrees of freedom). As could be seen from the significance tests both dimensions are explanatory for the group differences.

Function	Eigenvalue	Percentage of Variance	Canonical Correlation
1	3.32	66.6%	0.88
2	1.66	33.4%	0.79

Test of Function(s)	Wilks' λ	χ^2	df	Sig.
1 through 2	0.09	480.0	60	0.00
2	0.38	192.6	29	0.00

In the table below, the standardised discriminant function coefficients are presented, which could be seen as equivalents to standardised beta's in regression. Since all probing techniques (and the extent of specification of the request) have been entered in the analysis, some coefficients are low. More parsimonious analyses using a stepwise enter of probing techniques led to similar predictions.

Standardised Canonical Discriminant Function Coefficients

	Function 1	Function 2
Term from answer respondent	0.36	-0.20
Term from question	0.05	-0.07
No terms used (undirective)	-0.11	-0.09
Expression of lack of comprehension	0.01	-0.23
Expression of comprehension	-0.02	0.21
Question of comprehension	0.06	0.23
Negative comment	-0.17	0.02
Positive comment	-0.40	-0.57
Echo	-0.23	-0.09
Unfinished question / sentence	0.24	-0.13
Hum (including 8624 minimal hums)	0.23	0.74
Meta-remark	-0.05	0.14
Follow-up question	0.38	0.10
Reflection following respondent	-0.07	0.00
Reflection countering respondent	-0.05	-0.10
Summary /paraphrase	0.21	0.52
Expression of doubt	0.08	-0.06
Request for specification by contrast	0.11	0.14
Request for specification of own opinion	-0.04	0.03
Request for experiences	-0.03	0.06
Request for feelings	0.18	-0.08
Request for reasoning	0.10	0.08
Request for specification	0.36	0.11
Request for elaboration	-0.14	0.45
Request for specification by example	-0.09	0.10
Question reformulation	0.04	-0.05
Question repetition, same meaning	0.01	0.06
Question repetition, verbatim	0.06	0.17
Why-probe	0.58	-0.25
Expression of perceived inconsistency	0.36	-0.12

6 Probing Tactic and Quality of Information

In the previous chapter, it was demonstrated that the training of the interviewers had the desired effect: the interviewers correctly asked the questions from the questionnaire, using the probing techniques that are preferred for the probing tactic they were trained in.

Since the interviewers behaved as they should have behaved, now the proof of the pudding is in the eating. In this chapter, I test the effects of the probing tactics on the five criteria for the quality of the information retrieved through the interview, as they have been described in section 2.3: relevancy, depth, amount, elaborateness and specificity of the received information.

6.1 Relevancy of the information

As an indicator for the quality of the received information, relevancy seems so obvious that it is often overlooked. Every researcher, however, only wants information that is relevant. Naturally, for the sake of rapport, interviewers may allow irrelevant talk, but when it comes to the important parts of the interview, the interviewer should not go astray and keep respondents on track as much as possible.

As was explained in section 2.3, relevancy is operationalised inversely, which means that *off-topic* talk has been coded. In general, off-topic deviations rarely occur. On average, 1.6 % of the turns in the interviews are off-topic.

When comparing the effects that the probing tactics have on off-topic talk in the main part of the interview, while leaving aside the introduction, the background part and the after-talk, an interesting result surfaces. No matter which probing tactic was used, the averages of the amount of relevant information obtained turn out to be almost equally low. This is measured with the percentage of off-topic turns of all speech turns. When doing an analysis of variance for the percentage of off-topic turns of all turns, the results are $F(2, 213)=0.44$ ($p=0.64$).

Accordingly, there is no difference between the three probing tactics in terms of spinning off more (or less) relatively irrelevant interview talk.¹

6.2 Depth of the information

The quality issue of depth has been dealt with by defining depth as uncovering Personal information, such as personal events or personal feelings.

In the coding, the code for personal information has been used on turn level, which means that a coding indicates that this particular speech turn of the respondent contains personal information. This may lead to a problem when considering the different number of speech turns and the different length of those speech turns. A long turn can contain a lot of personal information, while it can also contain just one minor detail, but in both

cases, it has been coded as personal information. A story cut into pieces by an interviewer who uses continuers will lead to a higher score on the amount of personal information than will that same story, had it been told in just a few lengthier turns. But, as I have stated in section 4.4.3, in the pilot study, the correlation between the codes used on turn level and codes used on occurrence level was high; moreover, I use percentages of personal information turns of all respondent turns to control for the effects for the amount of probe turns used by the interviewer².

Table 6.1 shows that the percentages for personal information turns of all respondent turns hardly differ between the accommodating, encouraging and challenging probing tactics. The standard deviation of the total mean of the percentages of personal information turns of all respondent turns is considerable: 9%.

Table 6.1 Anova table of the percentages of personal information turns of all respondent turns for the main parts of the interview (N=214 interviews)

	Mean accommodating	Mean encouraging	Mean challenging	Standard deviation of the total mean	F	Sig.
Percentage personal information turns of all respondent turns	26%	24%	24%	9%	0.64	0.53

Although the three probing tactics have proven to be clearly dissimilar to each other (see section 5.3.3), the effect of the probing tactics on the personal information or depth is insignificant.

These results are both intriguing and comforting. It is intriguing in light of the general logic in use in qualitative interviewing that says that when an interviewer encourages a respondent while probing, the respondent will ‘open up’. It is comforting to know that in order to reach depth in an interview, an interviewer can encourage, accommodate or challenge without any repercussions for the depth of a response.³

6.3 Amount of the information

The amount of information brought forward in the interview is the third quality criterion to be used when evaluating the quality of the information received in interviews. As was discussed in section 2.3, the amount of information depends on the goal of the research as well as the range of the central interview question. In this research, the goal of the interview was to reach completeness of information on the respondents’ social categorisation of Amsterdammers, friends and allochthons. The goal was instructed to the interviewers, they had to acquire as much information about the categories as possible.

As a concept, the amount of information has been operationalised through the variables of the number of (non-repeated) category bound predicates mentioned by the

respondent. When looking at the results from the analysis of variance, in Table 6.2, it is clear that there is no difference at all between the probing tactics.

Table 6.2 Anova table of all (non-repeated) category bound predicates for the three topic parts of the interview

All category bound predicates	Mean accommodating	Mean encouraging	Mean challenging	Standard deviation of the total mean	F	Sig.
Amsterdammers (n=213)	15.1	16.6	16.1	8.7	0.54	0.58
Friends (n=213)	20.8	21.3	21.7	11.0	0.13	0.88
Allochthons (n=211)	34.7	40.0	40.8	23.0	1.46	0.24

The standard deviation of the total mean of the three probing tactics is fairly large, which shows that there is quite some variation between interviews that are conducted using the three different probing tactics. The standard deviation for friends and allochthons is highest for the challenging tactic, which indicates that in some interviews in which this tactic was used, a large number of predicates were mentioned, while in other interviews, these were hardly used.⁴

One could argue that even though there are no differences between the probing tactics in terms of the amount of new information brought forward in the interviews, it is important to consider the efficiency of the probing tactics, as well. A high percentage of repetition of the same information could mean redundant information, which could be said to have been obtained inefficiently. However, one could also argue that repetition of information is sometimes used as a rhetorical device and as such, is far from redundant. Therefore, I will just use the term repetition.

When comparing the percentage of repetition of category bound predicates of all category bound predicates (Table 6.3) for the interviews in the different tactics, the results for the amount of repeated information again suggest that the probing tactics are almost equivalent. The standard deviations of the total mean vary between 10 and 15%.⁵

Table 6.3 Anova table of the percentage of repeated category bound predicates of all category bound predicates for the three topic parts of the interview

Percentage of repeated category bound predicates of all category bound predicates	Mean accommodating	Mean encouraging	Mean challenging	Standard deviation of the total mean	F	Sig.
Amsterdammers (n=213)	14%	17%	14%	15%	1.45	0.24
Friends (n=213)	10%	10%	11%	10%	0.55	0.58
Allochthons (n=211)	9%	11%	9%	10%	1.38	0.26

Hence, for the analysis of the amount of information delivered through the interview, different probing tactics do not lead to different amounts of information, nor do they result in the acquisition of varied amounts of repeated information.

6.4 Elaborateness of the information

In the previous section, the amount of information has been analysed. It is possible that on average, of the three different probing tactics, one might not lead to a higher amount of information, while it might still lead to more elaborate information. This measure of elaborate information is operationalised as the number of different predicate types. In Box 4.6 on page 84, the 16 possible predicate types are presented. Thus, for each topic part of the interview, the maximum possible score for the number of different predicate types is 16.

In Table 6.4, the average number of different predicate types is shown for each tactic. For Amsterdammers, this varies between 5.1 and 5.5, for friends, it varies between 5.9 and 6, and for allochthons, it is around 9.1. These results show that there are no differences between the three probing tactics.

The difference between probing tactics in terms of the standard deviation is minimal, except for the allochthons topic: the standard deviations are 2.6 for the accommodating probing tactic, 2.8 for the encouraging probing tactic, and 3.1 for the challenging probing tactic.⁶

Table 6.4 Anova table of the number of different predicate types for the three topic parts of the interview

Number of different predicate types	Mean accommodating	Mean encouraging	Mean challenging	Standard deviation of the total mean	F	Sig.
Amsterdammers (N=213)	5.5	5.1	5.5	2.1	0.94	0.39
Friends (N=213)	5.9	5.9	6.0	1.8	0.05	0.95
Allochthons (N=211)	9.2	9.1	9.1	2.9	0.04	0.96

In Table 6.5, a comparable measure for elaborateness is analysed. Instead of looking at the number of different predicate types, this measure concerns the number of different category labels a respondent uses. Whereas in the previous measure, the number of different predicate types has a maximum of 16 for the elaborateness of the respondents, for this measure, there is no maximum. This could possibly lead to larger differences. However, when we look at the means of the probing tactics, we see again that the differences are extremely small.⁷

Table 6.5 Anova table of the number of different category labels for the three topic parts of the interview

Number of different category labels	Mean accommodating	Mean encouraging	Mean challenging	Standard deviation of the total mean	F	Sig.
Amsterdammers (N=213)	3.0	3.1	3.2	1.8	0.13	0.88
Friends (N=213)	3.2	3.5	3.1	3.2	1.63	0.20
Allochthons (N=211)	11.3	11.8	11.6	1.9	0.38	0.68

In this study, there are no differences in the average measures for the elaborateness of the information: the number of category labels or predicate types. Nevertheless, it might be possible that while the numbers are equal, the content of the social categorisation or predication will vary. Then, other mechanisms are at work. This will be tested in Chapter 7.

6.5 Specificity of the information

The final criterion for an evaluation of the quality of the information received and to test if there are any differences in the effects of the three probing tactics is the analysis of the specificity of the information. The measure used is probably the best measure for specificity since it is most closely connected to how respondents talk in interviews; they name a certain category and relate predicates to it. The more they do this, and the less they repeat themselves, the more specific they are. As with the other indicators for the quality of the received information, the results in Table 6.6 yield insignificant differences.⁸

Table 6.6 Anova table of the number of category bound predicates per category label for the three topic parts

Number of predicates per category label	Mean accommodating	Mean encouraging	Mean challenging	Standard deviation of the total mean	F	Sig.
Amsterdammers (N=213)	6.2	6.7	6.5	4.3	0.24	0.79
Friends (N=213)	6.9	6.9	7.3	3.8	0.36	0.70
Allochthons (N=211)	3.2	3.4	3.5	1.7	0.43	0.65

The second measure for specificity concerns the predicate types: The average number of (non-repeated) category bound predicates per predicate type. The more different predicates are used for one predicate type, the more specific predication a respondent uses while discussing the topics. For instance, a respondent can name several different rights and obligations of friends, such as giving emotional support, calling you every now and then, and having the right to call you for help when moving to a new house. All these predicates will fit into the rights and obligations predicate type, but differ in content.

Table 6.7 Anova table of the number of category bound predicates per predicate type for the three topic parts

Number of predicates per predicate type	Mean accommodating	Mean encouraging	Mean challenging	Standard deviation of the total mean	F	Sig.
Amsterdammers (N=213)	2.7	3.3	3.0	4.3	4.53	0.01
Friends (N=213)	3.5	3.5	3.6	3.8	0.08	0.92
Allochthons (N=211)	3.6	4.3	4.2	1.7	3.09	0.05

Table 6.7 is the first table in this chapter in which significant results appear. For two topics, Amsterdammers and allochthons, the encouraging and challenging probing tactics led to slightly higher averages for the numbers of predicates per predicate type. In other words, respondents tend to talk more specifically about certain predicate types, when they are probed encouragingly or challengingly.⁹

However, these final results show differences with a rather small explained variance, of ($r^2=0.04$) for the Amsterdammers topic, and ($r^2=0.03$) for the allochthons topic. However, regarding the content of the interviews, this might indicate that there are possible effects between the three different probing tactics.

6.6 Concluding remarks on the effects probing tactics have on the quality of the information

In this chapter, it was shown that the probing tactics, although they were performed well, generally, did not have effects on the quality of the information brought forth in the interviews. This can only lead to the conclusion that, although we are often warned about effects of interviewer behaviour, in the three different tactics, on average, about the same level of quality of information is reached. The standard deviations show that within the probing tactics, there are large differences in variance between the interviews, which show variation in the dependent variables.

The open interview is probably a robust form of interviewing, in which respondents can answer relatively independently from the stance and reaction of the interviewer who is probing. Perhaps this is due to the feedback options the respondent has, in order to explain the social categorisation. The robustness of the open interview, in quality terms, does not automatically mean that respondents can withstand the effects of the probing tactics, regarding the content of the talk. So, the question remains if respondents answer differently and use different predicate types when they are interviewed with a challenging, encouraging or accommodating probing tactic. That final part of the research question is to be answered in Chapter 7.

Notes

¹ Neither is there a significant difference between the interviewers that interviewed within one of the three tactics. This is the result of additional analyses, in which all ANOVA's used in this chapter are also tested for within probing tactic differences between interviewers. Naturally non-parametric tests are more suitable, but since the results dare not that different, I use the same tests for interviewer effects as used in analysing the effects of the probing tactics.

² This choice was made on the basis of a qualitative analysis and after looking at the Pearson's correlation between the amount of personal information and the amount of minimal turns. The r-coefficient was fairly high, 0.7 ($p < 0.001$), indicating the effect of cutting stories of the respondent by the interviewer. An ANCOVA with the probing tactic as a fixed factor and the amount of minimal turns as the covariate showed that almost all of the effects of the amount of personal information were due to the covariate.

³ Since there are no significant differences between the interviewers within the challenging and encouraging probing tactic for the percentage of personal information they receive through the interviews, this conclusion could (within a limited range) be extended to interviewers. However, among the accommodating interviewers one received only a low percentage of personal information, while some others received high percentages. This led to significant differences between these interviewers.

⁴ In this challenging tactic the interviewers also scored significantly different in the amount of information received in the allochthon part of the interview. However for the friends part of the interview the significant differences between the interviewers was in the encouraging tactic. For all other parts and tactics, the interviewer effects within the probing tactic are insignificant.

⁵ The within probing tactic interviewer effects suggest the same, since these are all insignificant.

⁶ For the Amsterdammer part, the interviewer effects within the accommodating tactic deviate significantly for the different numbers of different predicate types. For the friends part, the same occurs for the encouraging tactic. And for the allochthon part, the accommodating and the challenging interviewers show significantly different numbers of different category bound predicate types.

⁷ There are no significant differences between interviewers within any probing tactic for the number of different category labels.

⁸ For the number of predicates per category label, only within the challenging tactic on allochthons the interviewer effects were significant.

⁹ One interviewer with a high average number of different predicates per predicate type per interview made that there is a significant interviewer effect. Within the challenging tactic on allochthons the interviewer effects were significant in the number of different predicates per predicate type.

7 Probing Tactic and Content of Information

7.1 The comparison of the content of the interviews

In Chapter 6 the focus was on the effects that probing tactics have on five criteria for the *quality* of the information that is offered in interviews. In this chapter, the effects that the probing tactics have on the *content* of the interview are analysed. Since we have seen that the quality indicators hardly differ over the three probing tactics, it is necessary to see if the content of the interviews varies as a result of the different probing tactics used. The three different topics of the interview, Amsterdammers, friends and allochthons, are considered in consecutive order.

The content of the interviews is compared in two ways. The first way is by analysing the answers of the respondents, using specific codes for specific answers. The second way is by analysing the content of the predicate types. These predicate types are based on the same codes as the predicate types in Chapter 6. However, in this chapter, the codes are used differently. The goal is not to find out how many of the types are used or how many predicates are used per type, but to determine which types are used, and to consider how this differs over the three tactics. In Box 7.1 the different predicate types are presented.

Box 7.1 Predicate types

Activity	Economic Feature	Preferences
Appearance	Educational Feature	Relational Feature
Beliefs	Other Features	Rights & Obligations
Birth Locality	Feelings	Time
Competences	Knowledge	
Culture	Locality	

As was shown in sections 2.2.3 and 4.4.3, most predicate types were established in advance, based on membership categorisation literature. Some were created during the coding phase, only after it became clear that the available types were insufficient. In this chapter, the types are simply used dichotomously and not for any frequency analysis: a certain predicate type was used or was not used for one of the three social categorisation topics in the interview.

Naturally, other researchers have developed alternative coding systems for coding social categories. This is especially the case regarding friends, as many authors have tried to establish which meaning various groups of people attach to this social category. For instance, Adams, Blieszner and De Vries (2000) developed a typology of elements of friendship definitions, which is very suitable for coding friends and friendship. However, their coding system is not applicable to other topics, such as allochthons and Amsterdammers. The coding system I use in this chapter is more abstract, but the

advantage is that by using predicate types, it is possible to draw a comparison between different topics. At the end of this chapter, exactly this comparison will be made.

7.2 The Amsterdammer part of the interview

7.2.1 The Amsterdammer questions and answers

The first social categorisation topic in the interview was Amsterdammers. As was already mentioned, the topic was chosen because of its relative neutrality. Since the respondents all lived in Amsterdam, the topic was introduced with four questions (questions 1, 2a, 2b, and 2c) on Living in Amsterdam. The first question, after the introduction, was question 1: “How long have you been living in Amsterdam?” The purpose of this question was to make the respondent feel comfortable. The relatively simple question was posed to get the respondent in a talkative mood for the next three questions (questions 2a, 2b, and 2c). This set of three questions began with a request for an evaluation, which was followed by two requests for description. The questions are presented in the box below.

Box 7.2 Questions 2a to 2c

No.	Vraag	Question
Q2a	Hoe vindt u het om in Amsterdam te wonen?	What do you think about living in Amsterdam?
Q2b	Zijn er dingen die u prettig vindt aan het wonen in Amsterdam?	Are there things you find pleasant about living in Amsterdam?
Q2c	Zijn er dingen die u onprettig vindt aan het wonen in Amsterdam?	Are there things you find unpleasant about living in Amsterdam?

As Puchta and Potter (1999, 2004) have also described, these two types of questions can mobilise respondents into lively discussion and providing rich information.¹ If the respondent was talkative, the interviewer was allowed to skip question 2b or 2c, so the information respondents gave was less important than for them to be talkative.

The important questions were the questions as they are posed below.

Box 7.3 Questions 3 to 5

No.	Vraag	Question
Q3	Voelt u zich een Amsterdammer?	Do you feel Amsterdammer?
Q4	Wanneer vindt u iemand eigenlijk een Amsterdammer?	When do you consider someone an Amsterdammer?
Q5	En wat is typisch voor een Amsterdammer?	What is typical for an Amsterdammer?

The purpose of question 3 was to trigger categorisation by starting with a possible member, the respondent. By starting with the respondent, the source for the first categorisation was the respondent, him or herself. Question 4 was to distil the predicates that the respondent uses for his or her categorisation. Since question 4 could lead to a categorisation frame in which the main focus would be on boundary setting instead of the content of the category, the fifth question was posed.

The answers the respondents gave to question 3, “Do you feel Amsterdammer?” have been coded using three values: Yes, No or Somewhat. As Table 7.1 shows, when analysing the effects of probing tactics on the answers to question 3, the conclusion is rather straightforward: there is no significant difference between the answers of the respondents that result from the three probing tactics used. In other words, the probing tactic has no effect on whether respondents feel Amsterdammer or not.

Table 7.1 Crosstabulation of probing tactics and the answers to question 3 “Do you feel Amsterdammer?”.

Do you feel Amsterdammer?	Accommodating	Encouraging	Challenging	Total
Yes	50	48	42	140
No	14	15	15	44
Somewhat	7	8	15	30
Total	71	71	72	214

$\chi^2(4)=4.6$ $p=0.34$

Since this first analysis is on one of the first questions in the interview, it is conceivable that the interviewer had not yet had enough time to perform the trained probing tactic. Therefore, it would be possible that further on for this topic, there would be effects of the probing tactic. This will be tested using the predicate types for the answers to question 3, 4 and 5.

7.2.2 The predicate types for Amsterdammer categories

To analyse the effects that the probing tactics have on the content of the social categorisation of Amsterdammers, chi-square tests of the occurrences of predicate types per probing tactic are performed.

In this analysis, the answers to all three Amsterdammer questions are analysed together. There are three reasons for this pooled analysis. The first is that respondents could shuffle the questions, for instance, if respondents start to talk about Amsterdammers in prototypical manners, directly after question 3, rather than after question 5. The second reason is that the interviewer either felt the need to alter the sequence of the questions due to some answers of the respondent or -and this occurs most often- the interviewer saves some probes to use them after all three questions have been posed. The third reason is that the predicate types are analysed for all Amsterdammer categories that respondents use. For instance, when respondents juxtapose ‘Real’ Amsterdammers to ‘Import’ Amsterdammers, the predicates for both categories are analysed.

From Table 7.2, one can see that there is only one significant difference between the three probing tactics when looking at the Pearson’s chi-square for all types. I do not take this single significant difference very seriously, due to the chance capitalisation that is at

play. Therefore, for the predicate types on Amsterdammers, there are no serious differences between the three different probing tactics. This means that concerning the neutral question topic, the probing tactics have no effect on the content of the answers.

Table 7.2 Percentages per probing tactic, of interviews in which specific predicate types are used for the Amsterdammer part of the interviews.

	Accommodating (N=71)	Encouraging (N=71)	Challenging (N=71)	$\chi^2(2)$	Sig.
Activity	58%	61%	59%	0.12	0.94
Appearance	10%	6%	10%	1.09	0.58
Beliefs	18%	11%	8%	3.31	0.19
Birth Locality	51%	45%	56%	1.80	0.41
Competences	27%	30%	34%	0.85	0.65
Culture	10%	4%	20%	8.73	0.01
Economic Feature	13%	14%	18%	0.96	0.62
Educational Feature	7%	4%	6%	0.53	0.77
Other Features	96%	100%	99%	3.57	0.17
Feelings	66%	54%	61%	2.39	0.30
Knowledge	21%	15%	17%	0.83	0.66
Locality	80%	76%	75%	0.69	0.71
Preferences	35%	23%	28%	2.80	0.25
Relational Feature	6%	7%	4%	0.53	0.77
Rights & Obligations	15%	23%	14%	2.03	0.36
Time	34%	34%	41%	1.02	0.60

Thus, when comparing the three probing tactics for the content of the answers and the predicate types, there are no substantial differences². The very small differences between the three tactics in both the quality and the content of the Amsterdammers part of the interviews lead to the question if there is any difference between the three probing tactics at all. In the design of the study, however, the three different topics were purposively selected: Amsterdammers is thought to be the easiest social category to talk about, while friends is a more personal topic and allochthons a more controversial topic. Therefore, it is interesting to see where the differences lie for the two other topics.

7.3 The friend part of the interview

7.3.1 The friends questions and answers

The second topic in the interviews is the topic of friends. Since friends is a personal topic, in the introduction to the interviews, the topic was broadly introduced as “social contacts”. After the questions about Amsterdammers, the interviewers posed a question about the neighbourhood and then continued with two questions about neighbours. So, without explicitly mentioning it, the topic of “social contacts” was introduced.

The first question on friends was posed within an Amsterdam framework to keep the interview coherent for the respondents. Question 9 asked: “Do you have friends in

Amsterdam?” Question 9 functioned as a filter question for question 10, which was a very common question used in survey research on friends: “Could you indicate how many friends you have in Amsterdam?” On average, respondents mention 12.6 friends. The number of friends, however, differed between 0 and 60, suggesting that respondents used extremely different categorisations for friends.

Table 7.3 Anova table of the average number of friends per probing tactic, from answers to question 10 “Could you indicate how many friends you have in Amsterdam?”.

	Accommodating (N=66)	Encouraging (N=66)	Challenging (N=67)	F	Sig.
Average number of friends	13.0	14.5	10.3	2.39	0.09

Table 7.3 shows that respondents who were interviewed with the challenging tactic mentioned an average of almost 3 friends less than respondents of the accommodating tactic and 4 friends less than respondents who were interviewed encouragingly. However, these results are not statistically significant for a p-value of 0.05, and thus, they show no difference between the effects of probing tactics.³

The third question of the friends part of the interview was question 11: “When do you call someone a friend?” This question is central to membership categorisation analysis. In answering this question, the respondents came up with their own specific predicates for friends. Question 12, “Do you think that there are differences between types of friends?” was the scripted follow-up question that was asked to prompt respondents to disclose more detailed categorisation information. Question 13, “Do you think that there are differences between an acquaintance and a friend?” referred to the rather common sense boundary between friends and not-friends. The friends part of the interview was closed with question 14, which was intended to simultaneously function as a bridge to the next question: “Do you have allochthon friends in Amsterdam?”

This question is also analysed for possible effects of the probing tactics (Table 7.4). The probing tactic had no significant effect on the answer to the question of whether or not the respondent has allochthon friends in Amsterdam. Thus, the performed probing tactics have no effect on report (or definition) of the number of friends, nor on the report concerning allochthon friends.

Table 7.4 Crosstabulation of probing tactics and the answers to question 14 “Do you have allochthon friends in Amsterdam?”.

“Do you have allochthon friends in Amsterdam?”	Accommodating	Encouraging	Challenging	Total
Yes	31	25	19	75
No	40	43	50	133
Total	71	68	69	208

$\chi^2(2)=4.0$ p=0.14

7.3.2 The predicate types for Friend categories

In this subsection, the analysis of the effects that the probing tactics had on the content of the information is again specified to the predicate types used. Just as in the analysis of the predicate types in the Amsterdammer part of the interview (section 7.2.2), in the analysis of the friends part of the interview, the answers to the different questions are not analysed separately but are pooled over all questions on friends.

From Table 7.5 it is clear that almost all predicate types do not differ significantly across the three probing tactics used. It is clear that some predicate types for friends are simply less relevant, given that some means are rather low.⁴

Table 7.5 Percentages per probing tactic, of interviews in which specific predicate types are used for the friends part of the interviews.

	Accommodating (N=71)	Encouraging (N=71)	Challenging (N=71)	$\chi^2(2)$	Sig.
Activity	99%	89%	97%	8.24	0.02
Appearance	0%	1%	1%	1.01	0.60
Beliefs	4%	3%	4%	0.26	0.88
Birth Locality	3%	0%	1%	2.03	0.36
Competences	7%	1%	4%	2.78	0.25
Culture	0%	0%	1%	2.01	0.37
Economic Feature	7%	7%	1%	3.07	0.22
Educational Feature	7%	8%	3%	2.13	0.34
Other Features	83%	87%	85%	0.52	0.77
Feelings	28%	35%	32%	0.82	0.66
Knowledge	20%	32%	34%	4.18	0.12
Locality	72%	54%	65%	5.22	0.07
Preferences	30%	37%	48%	5.14	0.08
Relational Feature	96%	92%	97%	2.49	0.29
Rights & Obligations	63%	65%	54%	2.24	0.33
Time	49%	55%	56%	0.79	0.67

One predicate type, however, does differ significantly across the three tactics. This predicate type is activity. From a close analysis of the interviews where this predicate type is missing, I conclude that this is mostly a result of formulation⁵, rather than completely different meanings.

As was seen in the Amsterdammers part of the interview, for the friends part of the interview there are hardly any differences between the three probing tactics in regards to both the answers to the interview questions and the use of predicate types. This leads us to consider the final and controversial topic of allochthons to find out if there are any differences in this topic that result from using the challenging, accommodating or encouraging probing tactic.

7.4 The allochthon part of the interview

7.4.1 The allochthons questions and answers

As was the case for the topic of friends, the allochthon topic was presented generally in the introduction of the interview. The interviewers were instructed to introduce the topic as “different population groups” (“verschillende bevolkingsgroepen”). Although this is a more general term, it is commonly used, so most people recognise it to refer to ethnic diversity.

The topic was already introduced in two questions. The first time was question 8 in the Amsterdammer part, which concerned neighbourhood and the neighbours, which asked: “Do you have allochthon neighbours?” The second time, the allochthon topic was presented later in question 14, on allochthon friends, as it was analysed in section 7.3.1. For these questions, respondents not only referred to the categories friends and neighbours but also to allochthons. This means that the respondents were at least able to *use* the social category allochthon⁶.

After this first categorisation of allochthons, respondents were asked to explicate their categorisation of allochthons in questions 15 to 17.

Box 7.4 Questions 15 to 19

No.	Vraag	Question
Bridge	Amsterdam is een stad waar veel allochtonen leven. Daarom wil ik het nu graag daarover hebben.	Amsterdam is a city where many allochthons live. Therefore, I would like to talk about that.
Q15	Waar denkt u zelf aan bij de term allochtonen?	What do you have in mind when thinking of the term allochthons?
Q16	Zijn er volgens u verschillen tussen allochtonen en autochtonen?	According to you, are there differences between allochthons and autochthons?
Q17	Zijn er volgens u verschillen tussen allochtonen onderling?	According to you, are there differences among allochthons?
Bridge	Mensen hebben heel verschillende gevoelens over allochtonen;	People have very different feelings about allochthons;
Q18	Als u denkt aan allochtonen, wat voor gevoelens roept dat bij u op?	When you think of allochthons, what kind of feelings come up?
Q19	We hadden het net over Amsterdammers, kunnen allochtonen Amsterdammers worden?	We were just discussing Amsterdammers; can allochthons become Amsterdammers?

The answers to these three important questions will be used in the next subsection to see whether the predicate types mentioned differ across the three probing tactics. Question 18 was scripted to investigate how respondents use the term allochthons, when formulating their opinions. In question 19, the Amsterdammer and the allochthon topic were coupled.

In considering the effects of the three probing tactics on the content of the answers to the allochthon questions, the answers to two questions have been coded for four aspects. The first answer analysed is the answer to question 18.

Table 7.6 Crosstabulation of probing tactics and the answers to question 18 “When you think of allochthons, what kind of feelings come up?” coded on emotions.

Emotions	Accommodating	Encouraging	Challenging	Total
Mostly Negative	17	24	19	60
Neutral⁷	25	25	29	79
Mostly Positive	11	7	8	26
Total	53	56	56	165

$\chi^2(4)=2$ p=0.36

As can be read from Table 7.6, there are no significant differences between the three probing tactics in relation to the positivity or negativity of respondents’ feelings towards allochthons. Since the question solicits specific feelings, one would expect that respondents would offer descriptions of emotions. Most respondents did indeed (165), but others did not refer directly to emotions, but spoke instead about experiences that illustrated their emotions. Eighty respondents did both. For experiences, as Table 7.7 shows, the effects of the probing tactics are insignificant as well.

Table 7.7 Crosstabulation of probing tactics and the answers to question 18 “When you think of allochthons, what kind of feelings come up?” coded on experiences

Experiences	Accommodating	Encouraging	Challenging	Total
Mostly Negative	14	23	15	52
Neutral	14	13	13	40
Mostly Positive	16	10	11	37
Total	44	46	39	129

$\chi^2(4)= 3.6$ p=0.17

The other question that is analysed for two aspects is question 19, which questioned the possibility of allochthons becoming Amsterdammers. The first aspect analysed is simply the answer to the question. Four values are used as codes for these answers: *Yes / Under conditions (after probe) / Under conditions (without probe)* and *No*. The *Under conditions* values are coded to see if interviewer probing might have had an effect on the answer. From the codes presented in Table 7.8, one can notice an interesting difference between the tactics.

Table 7.8 Crosstabulation of probing tactics and the answers to question 19 “We were just discussing Amsterdammers; can allochthons become Amsterdammers?”.

Can allochthons become Amsterdammers?	Accommodating	Encouraging	Challenging	Total
Yes	45	43	44	132
Under conditions (after probe)	3	3	12	18
Under conditions (without probe)	21	19	13	53
No	2	6	3	11
Total	71	71	72	214

$\chi^2(6)= 13.3$ p=0.04

When using the challenging tactic, the interviewer invited respondents to rethink or nuance their answers more often (or more quickly). Thus, this is the first evidence of an effect of the probing tactic. Nevertheless, when grouping the *Under conditions* values

together, the significance disappears ($\chi^2(4) = 2.6$ $p=0.62$). This leads to the conclusion that this effect is mainly a result of the promptness of the challenging interviewers but does not significantly effect the content of the answers. Due to the possibility to explicate their initial answers, the respondents that were interviewed in the accommodating and the encouraging tactics, could also add conditions.

The second analysis of the answers to question 19 considers if the respondent used the same criteria for an Amsterdammer when speaking on the topic of allochthons as the respondent used while discussing the Amsterdammer topic. The challenging interviewers were specifically instructed to point out possible inconsistencies between the answers to questions 4 and 19, so it would be interesting to see if the challenging probing tactic led to fewer inconsistencies. From Table 7.9, one can conclude that there are not significantly less inconsistencies in the interviews held in the challenging probing tactic. Rather, the number of inconsistent respondents is even higher (though not significantly) in the challenging tactic!

Table 7.9 Probing tactics and the consistency of the answers to question 19 “we were just discussing Amsterdammers; can allochthons become Amsterdammers?” with the criteria used in the answers to question 4 “When do you consider someone an Amsterdammer?”.

	Accommodating	Encouraging	Challenging	Total
Consistent	61	64	60	185
Inconsistent	10	7	12	29
Total	71	71	72	214

$\chi^2(2) = 1.4$ $p=0.49$

So even for the controversial topic allochthons, the probing tactics do not have a significant effect on the emotions, experiences, inclusiveness or consistency expressed by the respondent. This leaves one final question unanswered: are the predicates used for allochthon categories different for the three different probing tactics?

7.4.2 The predicate types for allochthon categories

Since there is also no relation between the probing tactic and the predicate types for the Amsterdammer or friend categories, the question now, of course, is if and how the use of the three distinct probing tactics results in different information in the responses concerning the allochthon categories. For this particular topic, this is an appealing question since the topic is controversial, and therefore, one would expect respondents to be mindful of their words. So, at least for this topic, one would expect different results from the different probing tactics, whether this is due to social desirability, finding common ground or the extra possibility to speak out.

There are many qualitative examples in which respondents try to find a common ground⁸. A clear example of this is the utterance of complaints, as we see in the example presented in Box 7.5, from interview 551. Right after complaining about the costs of the

social welfare system, due to the large number of babies coming from Turks, Moroccans and asylum seekers, the respondent tries to find common ground with the interviewer.

Box 7.5 Fragment from interview 551

404	R: Ben ik de enige die zo praat of uh?	R: Am I the only one who talks like this or uh?
405	I: Nou, nee hoor, dat hoor ik wel vaker hoor, wel hetzelfde.	I: Well, no, I hear it more often, about the same
406	R: Nou, dankjewel, want nu durf je er tenminste over te praten. En, ik bedoel ik kom uit een volksbuurt, mijn vader, die is tachtig, maar die zei dat vijftig jaar geleden al.	R: Well, Thank you, because now one dares at least to talk about it. And, I mean I come from a working class area, my father who is eighty, but he said that 50 years ago already
407	I: mh, ja, interessant dat dat uh	I: mh, Yeah, interesting that that uh
408	R: Ik bedoel als als we hadden geluisterd,	R: I mean, if if we would have listened
409	I: mh mh	I: mh mh
410	R: naar de mensen die het kunnen weten,	R: to the people that could know
411	I: mh mh	I: mh mh
412	R: want het is met alles zo, en het is, dat heeft niks met ziekte of wat ook te maken, maar als iets aanbelt, bij iemand,	R: because it is in everything, and it is, that has nothing to do with sickness or anything else, but if something knocks, at someone
413	I: mh mh	I: mh mh
414	R: dan ga je dan ga je er anders over praten, als het bij jou aanbelt. En als je zit vanuit een ivoren toren,	R: then you will, then you will talk differently, when it knocks your door. And if you are in an ivory tower
415	I: ja	I: yeah
416	R: dan denk je, nou ja, sinds ze asielzoekers zeg maar door heel Nederland verspreid zijn, en hoor je ook veel meer mensen klagen omdat het dan ook ineens bij hun voor de deur gebeurt.	R: you will think, well yeah, since they asylum seekers are scattered throughout the Netherlands, then you also hear many more people complain because it happens suddenly in front of their doors.
417	I: aha	I: aha
418	R: Dan is het verrek,	R: Then what the heck,
419	I: ja	I: yeah
420	R: hé? er wordt gestolen, campings worden leeggeplunderd, en weet je dat ze het zijn,	R: ay? Things are stolen, campsites are plundered, and you know that they are the ones
421	I: mh	I: mh
422	R: maar dan ben je een [racist.]	R: But then you're a [racist.]
423	I: [mh mh]	I: [mh mh]
424	R: En nu is het gewoon dat je realistisch ben en daar ben ik heel erg blij om.	R: And now it's simply that you are realistic and I am very glad about it.

Although the respondent already appreciates the interviewer's acceptance of the answer in turn 406 of the fragment, the positive comment of the interviewer in turn 407 encourages the respondent to speak out. This interview is obviously held in the encouraging probing tactic. The interviewer is following the path taken by the respondent, providing encouragement to continue. So, based on these types of interactional signals that indicate a search for a common ground, one would expect that probing tactics do have an effect on such a controversial topic.

Since examples like the one above seem to illustrate the effect of interviewer behaviour,⁹ one could expect that in a topic this controversial, the probing tactics will have effects on the predicate types and thus on the content of the information provided in the interview. After the example presented above, anyone can imagine that the encouraging probing tactic would lead to more openness and thus a different categorisation than the other two probing tactics.

7. Probing Tactic and Content of Information

Now the proof of this qualitative pudding is in its quantitative eating. Therefore, Table 7.10 presents the percentages of the different predicate types for the allochthon part of the interview for the three different probing tactics.

Table 7.10 Percentages per probing tactic, of interviews in which specific predicate types are used for the allochthon part of the interviews.

	Accommodating (N=70)	Encouraging (N=71)	Challenging (N=70)	$\chi^2(2)$	Sig.
Activity	77%	80%	74%	0.72	0.70
Appearance	56%	73%	69%	5.15	0.08
Beliefs	69%	66%	73%	0.75	0.69
Birth Locality	43%	38%	49%	1.60	0.45
Competences	63%	77%	61%	5.06	0.08
Culture	83%	76%	84%	1.78	0.41
Economic Feature	64%	58%	57%	0.92	0.63
Educational Feature	51%	45%	40%	1.85	0.40
Other Features	100%	97%	97%	2.02	0.36
Feelings	40%	34%	30%	1.58	0.45
Knowledge	16%	14%	17%	0.25	0.88
Locality	97%	90%	94%	3.03	0.22
Preferences	47%	51%	47%	0.24	0.89
Relational Feature	21%	8%	23%	6.16	0.05
Rights & Obligations	53%	48%	50%	0.35	0.84
Time	33%	42%	33%	1.81	0.41

From the table, one can see that only one predicate type is significant for a p-value of 0.05. This predicate type is the relational feature. As was shown in section 4.4.3, this predicate type was created during the coding phase, mainly to provide for the predication of a relation between two members of the categories. For the allochthon topic, this predicate type was sometimes used when respondents were not talking about the allochthon category and its subcategories, but of the relation between allochthons and autochthons, the relationship among allochthons or the relationship between the respondents, themselves, and allochthons. Since this type is rather different and slightly more complex, I do not have a better explanation for this single significant result other than chance capitalisation. Both the predicate types appearance and competences are almost significant and percentage-wise are higher for the encouraging probing tactic.¹⁰

But, rather than speculating on possible explanations for insignificant differences, it is worthwhile to notice that again the results for the allochthon topic are, generally, insignificant. This implies that the different probing tactics have no effect on the predicate types for the controversial topic.

7.5 Concluding remarks on the effects probing tactics have on the content of the interviews

In Chapter 6, it was shown that the quality indicators for the information received in the interviews had no significant results, and the same occurs in this chapter for the content of the three different social categorisation topics, with respect to both the answers to the interview questions and the predicate types used by the respondents.

For all three topics, different respondents varied remarkably in their predication. In the different interviews, the variation between respondents is huge, which becomes obvious in simply reading the transcripts or listening to the recordings of the interviews. When using the predefined coding scheme on predicate types, most types are also different between interviews. However, when analysing the effects of the probing tactics on the predicate types, the results are that there is no significant effect of these tactics.

One final remark has to be made, before reaching a conclusion on the content of the information as a consequence of the three probing tactics. One could imagine that there would be no difference between the tactics because of a lack of variability of the predicate types. However, this was not the case, for an analysis of the differences between the predicate types across the three different topics confirms a strong relation between the interview topic and the predicate types that were used is strong. In Table 7.11, the results for the different predicate types are shown.

Table 7.11 Percentages of the interviews in which specific predicate types are used per topic.

	Amsterdammer (N=213)	Friends (N=213)	Allochthons (N=211)	$\chi^2(2)$	Sig.
Activity	59%	95%	77%	76.75	0.00
Appearance	8%	1%	66%	285.21	0.00
Beliefs	13%	4%	69%	262.12	0.00
Birth Locality	51%	1%	43%	138.50	0.00
Competences	30%	4%	67%	190.55	0.00
Culture	11%	0%	81%	380.18	0.00
Economic Feature	15%	5%	60%	183.57	0.00
Educational Feature	6%	6%	46%	144.04	0.00
Other Features	98%	85%	98%	41.57	0.00
Feelings	60%	32%	35%	42.17	0.00
Knowledge	18%	29%	16%	12.52	0.00
Locality	77%	63%	94%	57.54	0.00
Preferences	29%	38%	48%	17.42	0.00
Relational Feature	6%	95%	18%	418.10	0.00
Rights & Obligations	17%	61%	50%	88.53	0.00
Time	36%	54%	36%	17.70	0.00

Immediately upon looking at Table 7.11, it can be inferred that across the three different social categorisation topics, respondents used very different sets of predicates. The Amsterdammer social category, as well the allochthon category, have a high score for

birth locality and locality, since these are obviously important types for those categories. However, category bound feelings turn out to be used far more frequently when talking about Amsterdammers than when talking about friends or allochthons. The defining predicate types for friends turn out to be the category bound relational feature and rights and obligations, both for obvious reasons. For the allochthon topic, a broader pattern of predicates is used, but appearance, beliefs, competences, culture and economic feature stand out against the other two topics of social categorisation.

So, this topic specific distribution of the predicate types shows that this way of analysing social categorisation is very suitable for detecting differences in the content of information. Apparently, the predicate types are not only analytically useful tools when used qualitatively but also when used quantitatively.

The conclusion of this chapter is rather straightforward: It does not matter if the topic is more or less neutral, personal or controversial, the effects that the probing tactics have on the content of the information are by and large absent. This means that the open interview turns out to be rather robust against the potentially different influences of these probing tactics, on both the quality and the content of the received information.

Notes

¹ Although Puchta and Potter (1999, 2004) did this analysis on focus group interviews, the phenomena they describe occur in an individual interview as well.

² An additional chi-square analysis shows that within each probing tactic, the interviewers do not differ significantly on any of the category bound predicate types for the Amsterdammer part of the interview.

³ The interviewer effects on the number of friends is even less significant: $F(35, 163)=1.13$ $p=0.31$ or non-parametrically Kruskal-Wallis $\chi^2(35)=37.2$ $p=0.37$.

⁴ In the additional chi-square analysis on interviewer effects within the probing tactics for the friends part of the interview, one of these infrequently used predicate types (economic feature) for friends differs significantly between interviewers within the encouraging probing tactic. For one interviewer the predicate is used in three interviews, whereas for the other interviewers it is used in zero or one interview only. Besides this significant result, only for two out of 48 predicate types the interviewer effects significantly. These are locality and relational feature and again it is within the encouraging probing tactic.

⁵ Sometimes activities and other features are used for fairly similar predicates but are still different in the predicate *type*, since the formulation is for activities and action (often a verb), whereas for other features it is a feature (often a noun), without the active element. When activities were missing (in the encouraging tactic), it was mostly due to this formulation issue.

⁶ After question 08, a total of 11 respondents asked for a definition of an allochthon, as did 22 respondents, following question 14. Only one respondent posed the question for the meaning of allochthon in reaction to both questions. Naturally, the response of the interviewers was: “How you define it yourself”? This is a variant of the typical standardised interviewing behaviour WIMTY (“whatever it means to you”) response, as described in (Moore, 2004).

⁷ The neutral class consists of respondents mentioning both positive and negative feelings and experiences and respondents who mentioned feeling neutrally.

⁸ The grounding used here is naturally related to the grounding proposed by Clark and Brennan (1991) but focuses more explicitly on mutual beliefs or opinions.

⁹ Although we obviously will never know whether using the challenging or accommodating probing tactic in *this* interview with *this* respondent would have led to equally racist talk.

¹⁰ When analysing the interviewer effects within the three probing tactics through the additional chi-square analyses significant results show up for two predicate types. These are category bound preferences for the accommodating tactic and category bound appearances for the challenging tactic. For both I believe it is chance capitalisation rather than anything else.

8 Conclusion

8.1 Short summary of the research and its conclusions

Semi-structured interviews with mostly open-ended questions, so called *open interviews*, could be used for the reconstruction of concepts as used by the population under study. This could be the goal in itself as is often the case in qualitative surveys, but they can equally well be used for large-scale surveys with closed questions, for instance in a pre-study meant to inform the operationalisation and the formulation of the closed questions. Or, they could be used after a survey to enable the interpretation of respondents' answers.

Since in open interviews, the answer categories are not predefined, the most important interviewer behaviour in this type of interview is probing respondents after their initial answers to obtain explanations, examples, specifications, elaborations and additions. This probing behaviour is essential for collecting more, more specific, more elaborate and more in-depth relevant information.

Interestingly, probing behaviour in open interviewing has rarely been systematically studied, and more importantly, very seldom studied using an experimental design. For survey interviews with closed questions, however, interviewer behaviour has been studied far more frequently. Some results from these studies with closed interviews could be transferred to open interviews; however, most results are not transferable. In regards to probing behaviour, this is due to the very different objectives of probing in the two interview types. In survey interviews, probing is mainly generated by mismatches between answers and prestructured response options (Ongena, 2005) and thus a distortion of the paradigmatic question-answer sequence that needs to be repaired with probing. In open interviews, probing is mostly generated by incomplete answers. Since most initial answers to open questions are incomplete or need something akin to elaboration, explanation or specification, probes are almost always necessary.

For *survey interviews with closed questions*, it has been well established that both the quality and content of the received information are vulnerable to inadequate interviewer probing. Although probing in interviews with open questions is less vulnerable to errors than the error repair probes used in survey interviewing, it is common sense that asking open questions and probing afterwards gives the interviewer a major role in the data collection. And indeed, the autonomy of the interviewer in open interviews is much higher due to probing.

It could be stated that generally, there is a severe lack of methodological studies on open interviews. Furthermore, literature on *open, unstructured and unstandardised qualitative research* interviewing is only partly relevant. For this research, open questions are used and probing is central; however, in the type of open interviews used in this study,

standardisation of interviewer behaviour, the presentation of the initial questions and the sequence of the topics or even questions is followed more strictly than in many qualitative interviews.

To deal with the lack of knowledge on the effects of different probing behaviours of the interviewer, a field experiment was held. In this field experiment, a standardised questionnaire was used, and the interviewer behaviour for posing the initial questions was standardised as well. For probing behaviour, three different tactics were developed, and interviewers were trained in one of these probing tactics. The use of probing techniques was not standardised, since the choice for a certain probing technique is dependent on both the answers from the respondent and the probing tactic in which the interviewer has been trained.

The research question that has been answered in this thesis consequentially is: “What effects do different probing tactics have on the quality and content of the information received in open interviews?”

To answer this research question, the probing tactics have been defined as an interviewer *stance towards the responses* of the respondent, and *the reaction following* from this stance. I distinguished three probing tactics, based on notions found in a broad range of interviewing literature:

- The encouraging probing tactic, in which the interviewer takes the answer for granted and explicitly reacts accordingly,
- The challenging probing tactic, in which the interviewer does not take the answer for granted and shows this in his or her reaction, and
- The accommodating probing tactic, in which the interviewer does not explicate his or her stance.

In this study, the focus lies on the effects the probing tactics have on both the quality and content of the information received. The quality of the received information is operationalised into five indicators: relevancy, depth, amount, elaborateness and specificity. The content of the interviews is defined as both the content of the answers to some questions and the predicate types used by the respondents in talking about social categories.

The effects of the probing tactics were tested in a field experiment in which six groups of six novice interviewers were instructed in one of the three probing tactics. The intensive training took place over two days.

To achieve the highest internal validity as possible, several measures were taken. First, the interviewers were randomly (only limited by their agendas) assigned to one of the six groups. Second, except for the probing techniques, all interviewer behaviour was highly standardised: The introduction, the interviewing style, the topic sequence, part of the question sequence, question formulation, the stop-criterion and the handling of the

after-talk. Third, the respondents were randomly assigned to the interviewers and thus to the treatment groups. It was planned that all interviewers would interview 6 respondents from Amsterdam. Two interviewers interviewed 5 respondents, resulting in a total number of 214 interviews.

To assure external validity, several measures were taken as well. First, all three tactics are based on concepts found in a broad range of interviewing literature. Second, within the experiment three replications were held on three different topics. The first social categorisation topic was on Amsterdammers, the second on friends and the third on allochthons. The topics were chosen to differ on two dimensions: controversiality and intimacy. Third, the interviewers that were selected were not unlike interviewers employed in other social science studies, i.e. novice interviewers with a social science background. Fourth, the respondents were invited to participate through a sample from an access panel of Amsterdammers.

The interviewers performed very well, and they displayed minimal unintended behaviour, such as mistakes like skipping questions, reformulating questions with altered meanings or making suggestions. Besides, they achieved good results in performing probing behaviour as instructed. So, the experimental variable, the probing tactics, was performed very well.

The interview audio was precisely transcribed and checked. Afterwards, intense coding took place with high inter-coder-reliability and high validity and reliability due to the use of jurisprudence methods.

The effects that the probing tactics had on the quality as well as the content of the received information were analysed using mainly analyses of variance and chi-squares. These analyses demonstrate that there are no differences between the three probing tactics in terms of the five quality dimensions of the information retrieved through the interview: relevancy, depth, amount, elaborateness and specificity. Above that, the three different, well-performed probing tactics lead to no differences within the content of the interviews, neither for the coded answers to the questions nor for the different predicate types respondents bind to categories. Interestingly, these lacks of differences apply to all three social categorisation topics, leading to unequivocal results for a neutral, personal and controversial topic, respectively.

These results are comforting for many interviewers and researchers who use open interviews since there are no significant effects on either the content or the quality of the received information of the different probing tactics.

8.2 Wider potential

Naturally, there are boundaries to the external application of these results. In many other studies as discussed in Section 2.1.2, there is strong evidence that the interaction leads to

at least a certain amount of co-construction. The interviews held in this study are different from interviews held elsewhere, if not for the simple facts that in other open interviews, other topics are used, other respondents are interviewed and other interviewers are employed. It is also because the interviewers in this study were trained in a specific way and the interviewer behaviour was highly prescribed. Therefore it is important to explore the bandwidth of the results of this research

In this research, all three topics were on social categorisation since establishing the meaning of, and ascription to, categories or concepts is central to concept clarification interviews. The three selected topics (Amsterdammers, friends and allochthons) were chosen to differ on the level of intimacy and controversiality, in order to deal with possible confounders as a consequence of these dimensions. Naturally, one could think of more personal topics or more controversial concepts to be clarified, or even interviews with a totally different focus or range, as shown in Box 1.1. However, based on the total lack of differing effects of the three different topics and a broad range of neutral, but also very personal stories, as well as very controversial opinions offered, I believe that the results are generalisable not only to concept clarification interviews but also to broader and less focussed interviews.

One of the predicates often used for speaking about Amsterdammers was that Amsterdammers call a spade a spade. It is possible that my respondents, who were all living in Amsterdam, were sharing this category bound feature and spoke plainly, regardless of being challenged, accommodated or encouraged. The political activity of the respondents in terms of expressed voting intentions is also higher than that of the average Amsterdammer. The high education of most respondents could have been influential as well. It could be that due to their education, they are more aware of the probing tactic used on their answers, and as a consequence, they are less impressionable. In any case, there are a few aspects that could possibly influence the effects of the probing tactics on the quality and content of the information retrieved.

As was shown in the previous section, the interviewers selected in this study were not unlike interviewers hired by other researchers in universities. Therefore, I do not consider the interviewers to be possible threats to the external validity of the results in this study.

As was shown in Chapter 5, the interviewers performed nearly perfectly in posing the questions as scripted, without making interviewer mistakes. The probing behaviour shown in this research varied enough in techniques, and the interviewers probed as they were instructed, using the techniques as they were instructed. The probing tactics were also performed as instructed. The discriminant analysis in section 5.3.3.3 shows clear differences between the tactics.

All research that is available on interviewing shows that certain prerequisites are essential for interviewing performance. Without a good training, interviewers simply do what they consider good interviewing and behave especially different on questions that assume a great deal of interviewer activity (Billiet & Loosveldt, 1988). In open interviewing, almost all questions demand a fair amount of interviewer activity, so training for these kinds of interviews is expressly important. Other researchers analysing quality in open interviewing found that other than a good training, a quality monitoring system also has strong improving effects on interviewer behaviour (Lamb, *et al.*, 2002).

Now, exactly these two aspects, the training and the performance of the interviewer, were successful in this study. In other open interview studies (*ceteris paribus*), when the same level of interviewer accuracy is reached, I do not expect differences.

When other types of open interviews are used, the results could be different. In this research, the interviewer behaviour was standardised on many aspects: The introduction, bridges and question formulation were scripted and were to be presented verbatim; the topic sequence, the question sequence (with some possibilities for deviation in some situations), and the interviewing style were predetermined; the interviewer rules on suggestion and when to stop probing were strict, and even the behaviour for the after-talk was restricted. In other open interview types, this standardisation could be different, leading to different results. However, in general, I do not believe that a replication with less standardisation will produce very different results, since the process of answering open questions is more or less the same, as will be argued in the next section.

8.3 Probing tactics and feedback options for respondents

Since others (Baker, 2004; Rapley, 2001; Roulston, Baker, & Liljestrom, 2001; Roulston, 2006) have argued that in general co-construction takes place in individual interviews, it is intriguing that when analysing a large number of open interviews, there is no effect of probing tactics on the quality and content of the information. On the basis of the results shown here, it could be argued that these authors probably select strong cases of co-construction, since there are on average no differences in the co-constructions as a consequence of the probing tactics.

The question therefore remains as to what explanations might be found for the fact that there are no differences in the effects that the distinct probing tactics have on the quality and content of responses for any of the three topics.

This is more intriguing when considering the numerous studies on survey interviews with closed questions, where large differences can be seen to result from only minor changes in interviewer behaviour, question formulation and question sequence. Moreover, the lack of effects of the probing tactics is especially striking when considering

some common sense 'quantitative' objections to open interviews, which broadly posit open interviews as subjective, biased and person dependent¹.

A possible hypothesis to explain this paradox is that respondents in survey interviews are restricted in their answers due to the format of closed questions. In answering open questions, respondents have the possibility to nuance previously formulated answers, put them in perspective, continue on them or even renounce them. This leads to the idea that the question format enables respondents to become more robust against other possible intervening factors like probing tactics. The lack of differences in the effects that these tactics have on the quality and content of the information received during an interview suggests that the form of open interviews prompts respondents to speak out more freely, and thus, their responses will be more 'credible' (according to Guba and Lincoln (1985)) than might have been expected. By using their own words to answer open questions, respondents are more capable of expressing themselves in open interviews. This should lead to the conclusion that the feedback possibility for respondents answering open questions is what makes the open interview more robust than standardised interviews that use closed questions. This line of argumentation leads to the expectation that in all interviews that contain a substantive amount of open questions, regardless of their specific topic or ultimate intention, different probing tactics will not affect the quality and the content of the information.

8.3.1 How to ensure robustness

To ensure the robustness of the open interview, the researcher needs to deal with three essential preconditions: preparing the study, training the interviewers and monitoring of interviewers.

For the first precondition, the preparation of the study, the most important issue is to formulate good questions. For years now, scholars who analyse the interaction occurring through surveys with closed questions have been stressing this point. Suchman and Jordan (1990) show that survey questions often contain too many specifications of conditions and, as such, are rather awkward and difficult for interviewers to parse. A related issue that Houtkoop-Steenstra (2000) shows is that the question formulation (audience design) is often too broad, leading interviewers to reformulate questions for their specific respondents (recipient design). Others (Van den Berg, Mehcz, Houtkoop-Steenstra, & Holleman, 2002; Van den Berg & Carabain, 2003) have shown the vast amount of mistakes that are made by question designers. In other words, it is not that easy to formulate good questions of any kind. To ensure that respondents have the option to offer feedback, the questions need to be truly open and clear.

In addition to the clarity within the formulation of the questions, the whole structure of the questionnaire certainly needs to be in good order. The sequence of the

topics, the sequence of the questions, the routings and the bridges between the questions, along with the introduction and the introductory questions all need to be well designed. To put it briefly, if the design of the questionnaire is not conducive to the desirable ends, it will result in interviewers making ad hoc adjustments, according to their own ideas.

The methodology I used to ensure the quality of the questionnaire was threefold. It consisted of careful designing, peer-reviewing by colleagues working at the social research methodology department and pre-testing on *three* specific occasions. The first pre-test was carried out before the pilot study. The second pre-test I carried out during the pilot study, while watching the actors using the questions; the questionnaire was improved along the way. The final pre-test that led to minor changes on the questionnaire was held before the main study. The outcome was a solid questionnaire with only minor mistakes.

The second precondition, and the most important one for ensuring successful interviewer behaviour, is that interviewers enable respondents to nuance, elaborate, specify or simply make additions to their answers. This was necessarily emphasised in the training. Training is especially crucial for novice interviewers since it provides the foundation of their interviewer behaviour. I believe that several choices were central to the success of the training.

The first choice was to extensively prepare the content of the training. As shown in section 3.4.2, the training and the instruction manual were well prepared and even piloted, so they contained hardly any loose ends.

The second and probably most important choice was the decision to aim for interviewer commitment. This was accomplished by keeping the training groups small (6 students) and working on team spirit and mutual feedback. In creating the commitment, it was essential to keep the interviewers involved in the goals of the research, by sharing responsibility, while showing contagious enthusiasm and maintaining a very good understanding of the aims of the interviews. Naturally, the creation of understanding was most explicit and precise.

This understanding is the third (and again crucial) choice made for the training. Following recent work in survey methodology, the focus was on understanding, rather than trying to protocol, all interviewer behaviour. For instance, the importance of standardisation was explained and discussed, rather than simply taught as some rules to which the interviewers were required to adhere. From researchers who analyse standardisation rules -such as the ones Fowler and Mangione (1990) formulated- we know that in practice, interviewers often do not follow these rules but instead routinely use conversational resources for dealing with uncertainties during the interaction (see for an overview Maynard, Houtkoop-Steenstra, Schaeffer, & Van der Zouwen, 2002; Houtkoop-Steenstra, 2000; Van 't Hof, 2006). Therefore, in this research, the goals of standardisation

were extensively explained and discussed, before examples, strategies and some rules were introduced.

Following the same logic, the effects of suggestion were explained, the goals of all scripted questions were explained and discussed, all probing techniques and their functions were explained and discussed and the probing tactic to adhere to was explained, rather than instructed as a protocol. This was all done enthusiastically to encourage and motivate the interviewers.

The fourth choice that I believe ensured a good understanding of the intended interviewer behaviour was the intensity and diversity of the training. Participating in a very intensive crash course left students dreaming of interviewing. By keeping the didactics diverse, several modes of knowledge and skills could be addressed and trained. Some parts of the trainings were more lecture-like, while in other parts, group work, games, role-playing, videos, recorded practice interviews and group discussions were used. The various role-plays and practice interviews were especially useful.

Besides the training, the monitoring during the interviewing period was very useful for establishing interviewer quality, as the interviewers were directed towards the mistakes they made in the interviews.

8.4 Final warning

After reaching the conclusion that the inherent robustness of open interviews is accounted for through the feedback options for respondents answering open questions, a warning is essential. The comforting results for researchers who use open interviews may not lead to the conclusion that anything goes in open interviewing for the fact that “open interviews are so robust that interviewer behaviour is of no importance for the quality and the content of the received information”.

As stated in the previous section and throughout this thesis, the interviewers performed well. I believe that good interviewer behaviour is very important for ensuring robustness. First, if interviewers suggest answers, the robustness will deteriorate since the respondents will have less feedback options. When they make suggestions, interviewers directly close down possible answers (or categorisations).

Second, researchers need to formulate truly open questions beforehand, so they do not close down answering opportunities and therefore delimit the feedback options for the respondents.

Third, in addition to giving room for respondents' feedback options, interviewers should generate respondent feedback by adequately probing and then knowing when they can stop probing, by using something like the stop-criterion, used in this study.

The final warning, therefore, is that “anything goes” for the probing tactics, but only after the research is well prepared, and all interviewer behaviour is adequately performed.

Notes

¹ Kvale (1994) names seven more 'standard' objections, but I believe that this stray doll is rather grotesque.

Bibliography

- Abell, J., Locke, A., Condor, S., Gibson, S., & Stevenson, C. (2006). Trying Similarity, Doing Difference: The Role of Interviewer Self-Disclosure in Interview Talk with Young People. *Qualitative Research*, 6(2).
- Adams, R. G., & Allan, G. (1998). *Placing Friendship in Context*. Cambridge: Cambridge University Press.
- Adams, R. G., Blieszner, R., & De Vries, B. (2000). Definitions of Friendship in the Third Age: Age, Gender, and Study Location Effects. *Journal of Aging Studies*, 14(1), 117-133.
- Antaki, C., Barnes, R., & Leudar, I. (2005). Self-Disclosure as a Situated Interactional Practice. *British Journal of Social Psychology*, 44, 181-199.
- Ashmore, M., & Reed, D. (2000). Innocence and Nostalgia in Conversation Analysis: The Dynamic Relations of Tape and Transcript. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research [On-line Journal]*, 1(3).
- Atkinson, P., Coffey, A., & Delamont, S. (2003). *Key Themes in Qualitative Research: Continuities and Change*. Walnut Creek: Altamira Press.
- Atkinson, P., & Silverman, D. (1997). Kundera's Immortality: The Interview Society and the Invention of the Self. *Qualitative Inquiry*, 3(3), 304-325.
- Ayres, L. (2008). Semi-Structured Interviews. In L. M. Given (Ed.), *The Sage Encyclopedia of Qualitative Research Methods* (Vol. 1&2). Thousand Oaks: Sage Publications.
- Baarda, D. B., De Goede, M. P. M., & Van der Meer-Middelburg, A. G. E. (1996). *Open Interviews: Praktische Handleiding Voor Het Voorbereiden en Afnemen van Open Interviews*. Houten: Stenfert Kroese.
- Baker, C. D. (1997). Membership Categorization and Interview Accounts. In D. Silverman (Ed.), *Qualitative Research: Theory, Method and Practice*. (pp. 130-143). London: Sage Publications.
- Baker, C. D. (2004). Membership Categorization and Interview Accounts. In D. Silverman (Ed.), *Qualitative Research: Theory, Methods, Practice* (2nd ed.). London: Sage Publications.
- Bangerter, A. (2000). Self-Representation: Conversational Implementation of Self-Presentational Goals in Research Interviews. *Journal of Language and Social Psychology*, 19(4), 436-462.
- Barth, F. (1969). Introduction. In F. Barth (Ed.), *Ethnic Groups and Boundaries: The Social Organization of Cultural Difference*. Boston: Little, Brown and Company.
- Beatty, P. (1995). Understanding the Standardized/ Non-Standardized Interviewing Controversy. *Journal of Official Statistics*, 11(2), 147-160.

Probing Behaviour in Open Interviews

- Becker, H. (1954). A Note on Interviewing Tactics. *Human Organization*, 12(4), 31–32.
- Bethlehem, J., & Stoop, I. (2007). *Online Panels: A Paradigm Theft?* Paper presented at the The Challenges of a Changing World, Southampton.
- Billiet, J., & Loosveldt, G. (1988). Improvement of the Quality of Responses to Factual Survey Questions by Interviewer Training. *Public opinion quarterly*, 52(2), 190-211.
- Bird, C. M. (2005). How I Stopped Dreading and Learned to Love Transcription. *Qualitative Inquiry*, 11(2), 226-248.
- Boeije, H. R. (2002). A Purposeful Approach to the Constant Comparative Method in the Analysis of Qualitative Interviews. *Quality and Quantity*, 36(4), 391-409.
- Bravo-Moreno, A. (2003). Power Games between the Researcher and the Participant in the Social Inquiry. *The Qualitative Report*, 8(4), 624-639.
- Briggs, C. L. (1986). *Learning How to Ask: A Sociolinguistic Appraisal of the Role of the Interview in Social Science Research*. Cambridge: Cambridge University Press.
- Briggs, C. L. (2003). Interviewing, Power/Knowledge, and Social Inequality. In J. A. Holstein & J. F. Gubrium (Eds.), *Inside Interviewing: New Lenses, New Concerns*. Thousand Oaks: Sage Publications.
- Brinkmann, S. (2007). Could Interviews Be Epistemic?: An Alternative to Qualitative Opinion Polling. *Qualitative Inquiry*, 13(8), 1116-1138.
- Brubaker, R., Loveman, M., & Stamatov, P. (2004). Ethnicity as Cognition. *Theory and Society*, 33(1), 31-64.
- Carabain, C. (2007). *Taking Too Much for Granted? A Study on the Measurement of Social Attitudes*. Amsterdam: PhD-Thesis Vrije Universiteit Amsterdam.
- Cassell, C. (2005). Creating the Interviewer: Identity Work in the Management Research Process. *Qualitative Research*, 5(2), 167-179.
- Charmaz, K. (2006). *Constructing Grounded Theory. A Practical Guide through Qualitative Analysis*. Thousand Oaks: Sage Publications.
- Chirban, J. T. (1996). *Interviewing in Depth: The Interactive-Relational Approach*. Thousand Oaks: Sage Publications.
- Clark, H., & Brennan, S. (1991). Grounding in Communication. In L. B. Resnick, J. M. Levine & S. D. Teasley (Eds.), *Perspectives on Socially Shared Cognition* (pp. 127-149). Washington, DC: American Psychological Association.
- De Leeuw, E., & Van der Zouwen, J. (1988). Data Quality in Telephone and Face to Face Surveys: A Comparative Meta-Analysis. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II & J. Waksberg (Eds.), *Telephone Survey Methodology* (pp. 283-299).

- Dean, J. F., & Whyte, W. F. (1958). How Do You Know If Your Informant Is Telling the Truth? *Human Organization*, 17, 34-38.
- Denzin, N. K. (1970). *The Research Act in Sociology*. Chicago: Aldine.
- Dicicco-Bloom, B., & Crabtree, B. F. (2006). The Qualitative Research Interview. *Medical Education*, 40(4), 314-321.
- Dijkstra, W. (1983). *Beïnvloeding van Antwoorden in Survey-Interviews*. Utrecht: Elinkwijk.
- 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., Van der Veen, L., & Van der Zouwen, J. (1985). A Field Experiment on Interviewer-Respondent Interaction. In M. Brenner, J. Brown & D. Canter (Eds.), *The Research Interview; Uses and Approaches*. (pp. 37-63). London: Academic Press.
- Dijkstra, W., & Van der Zouwen, J. (1988). Types of Inadequate Interviewer Behaviour in Survey-Interviews. In W. E. Saris & I. N. Gallhofer (Eds.), *Data Collection and Scaling* (Vol. 1, pp. 24-35). London: Macmillan.
- Dohrenwend, B. (1965). Some Effects of Open and Closed Questions on Respondents' Answers. *Human Organization*, 24(2), 175-184.
- Edwards, D., Ashmore, M., & Potter, J. (1995). Death and Furniture: The Rhetoric, Politics and Theology of Bottom Line Arguments against Relativism. *History of the Human Sciences*, 8(2).
- Emans, B. (1990). *Interviewen: Theorie, Techniek en Training* (3e, herz. ed.). Groningen: Wolters-Noordhoff.
- Evers, J., & De Boer, F. (2007a). Het Ontwerpen van Een Individueel Interview. In J. Evers (Ed.), *Kwalitatief Interviewen: Kunst en Kunde*. Den Haag: Lemma.
- Evers, J., & De Boer, F. (2007b). Het Uitvoeren van Een Individueel Interview. In J. Evers (Ed.), *Kwalitatief Interviewen: Kunst en Kunde*. Den Haag: Lemma.
- Field, A. P. (2005). *Discovering Statistics Using Spss* (2nd ed.). London: Sage Publications.
- Foddy, W. (1998). An Empirical Evaluation of in-Depth Probes Used to Pretest Survey Questions. *Sociological Methods & Research*, 27(1), 103-133.
- Fontana, A., & Frey, J. H. (1998). Interviewing: The Art of Science. In N. K. Denzin & Y. S. Lincoln (Eds.), *Collecting and Interpreting Qualitative Data*. Thousand Oaks: Sage Publications.
- Fowler, F. J., & Mangione, T. W. (1990). *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. Newbury Park, California: Sage Publications.
- Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.

Probing Behaviour in Open Interviews

- Gladwell, M. (2008). *Outliers : The Story of Success*. New York: Little, Brown and Co.
- Glaser, B. G., & Strauss, A. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine.
- Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Garden City, NY: Doubleday Anchor.
- Gorden, R. L. (1956). Dimensions of the Depth Interview. *The American Journal of Sociology*, 62(2), 158-164.
- Gorden, R. L. (1969). *Interviewing: Strategy, Tactics, and Techniques*. (1st ed.). Chicago: The Dorsey Press.
- Gorden, R. L. (1980). *Interviewing: Strategy, Tactics, and Techniques*. (3rd ed.). Chicago: The Dorsey Press.
- Gorden, R. L. (1992). *Basic Interviewing Skills*. Prospect Heights: Waveland Press.
- Goudy, W. J., & Potter, H. R. (1975). Interviewer Rapport: Demise of a Concept. *Public opinion quarterly*, 39, 530-543.
- Guba, E. G., & Lincoln, Y. S. (1985). *Naturalistic Inquiry*. Beverly Hills, CA: Sage Publications.
- Hathaway, A. D., & Atkinson, M. (2003). Active Interview Tactics in Research on Public Deviants: Exploring the Two-Cop Personas. *Field Methods*, 15(2), 161-185.
- Hayes, A. F., & Krippendorff, K. (2007). Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures*, 1(1), 77-89.
- Hester, S., & Eglin, P. (1997). Membership Categorization Analysis: An Introduction. In S. Hester & P. Eglin (Eds.), *Culture in Action: Studies in Membership Categorization Analysis* (Vol. 4). Washington, D.C.: International Institute for Ethnomethodology and Conversation Analysis & University Press of America.
- Holstein, J. A., & Gubrium, J. F. (1995). *The Active Interview*. Thousand Oaks: Sage Publications.
- Housley, W., & Fitzgerald, R. (2002). The Reconsidered Model of Membership Categorization Analysis. *Qualitative Research*, 2(1), 59-83.
- Houtkoop-Steenstra, H. (1995). Meeting Both Ends: Between Standardisation and Recipient Design in Telephone Survey Interviews. In P. Ten Have & G. Psathas (Eds.), *Situated Order: Studies in the Social Organisation of Talk and Embodied Activities* (pp. 91-107). Washington: University Press of America.
- Houtkoop-Steenstra, H. (1996). Probing Behaviour of Interviewers in the Standardised Semi-Open Research Interview. *Quality & Quantity*, 30(2), 205-230.

- Houtkoop-Steenstra, H. (2000). *Interaction and the Standardized Survey Interview*. Cambridge: Cambridge University Press.
- Hutchby, I., & Wooffitt, R. (1998). *Conversation Analysis: Principles, Practices and Applications*. Oxford: Polity Press.
- Hyman, H. H., & Cobb, W. J. (1975 [1954]). *Interviewing in Social Research* (7th impr. ed.). Chicago: University of Chicago Press.
- Jayyusi, L. (1984). *Categorization and the Moral Order*. Boston: Routledge & Kegan Paul.
- Jefferson, G. (2004). Glossary of Transcript Symbols with an Introduction. In G. H. Lerner (Ed.), *Conversation Analysis: Studies from the First Generation* (pp. 13-31). Amsterdam/Philadelphia: John Benjamins.
- King, A. (2010). 'Membership Matters': Applying Membership Categorisation Analysis (Mca) to Qualitative Data Using Computer-Assisted Qualitative Data Analysis (Caqdas) Software. *International Journal of Social Research Methodology*, 13(1), 1 - 16.
- Kish, L. (1965). *Survey Sampling*. New York: Wiley.
- Krippendorff, K. (1980). *Content Analysis: An Introduction to Its Methodology* (Vol. 5). London: Sage Publications.
- Krippendorff, K. (2007, 01 June 2007). Computing Krippendorff's Alpha Reliability. ASC Working Paper of ScholarlyCommons@Penn. Retrieved 12 may 2008, from http://repository.upenn.edu/asc_papers/43/
- Kvale, S. (1996). *Interviews : An Introduction to Qualitative Research Interviewing*. Thousand Oaks, Ca.: Sage Publications.
- Kvale, S. (2006). Dominance through Interviews and Dialogues. *Qualitative Inquiry*, 12(3), 480-500.
- Kvale, S., & Brinkmann, S. (2008). *Interviews: An Introduction to Qualitative Research Interviewing*. London: Sage Publications.
- Lakoff, G. (1987). *Women, Fire, and Dangerous Things : What Categories Reveal About the Mind*. Chicago: University of Chicago Press.
- Lamb, M. E., Sternberg, K. J., Orbach, Y., Hershkowitz, I., Horowitz, D., & Esplin, P. W. (2002). The Effects of Intensive Training and Ongoing Supervision on the Quality of Investigative Interviews with Alleged Sex Abuse Victims. *Applied Developmental Science*, 6(3), 114 - 125.
- LeCompte, M., & Goetz, J. P. (1982). Problems of Reliability and Validity in Ethnographic Research. *Review of Educational Research*, 52, 31-60.

Probing Behaviour in Open Interviews

- Lee, R., & Fielding, N. (1995). Users' Experiences of Qualitative Data Analysis Software In U. Kelle (Ed.), *Computer-Aided Qualitative Data Analysis: Theory, Methods and Practice*. London: Sage Publications.
- Lee, Y.-J., & Roth, W.-M. (2004). *Making a Scientist: Discursive "Doing" Of Identity and Self-Presentation During Research Interviews*.
- Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability. *Human Communication Research*, 28(4), 587-604.
- Long, T., & Johnson, M. (2000). Rigour, Reliability and Validity in Qualitative Research. *Clinical Effectiveness in Nursing*, 4(1), 30-37.
- Loosveldt, G., & Sonck, N. (2008). An Evaluation of the Weighting Procedures for an Online Access Panel Survey. *Survey Research Methods*, 2(2), 93-105.
- Matthews, S. H. (1983). Definitions of Friendship and Their Consequences in Old Age. *Ageing and Society*, 3, 141-155.
- Maynard, D. W., Houtkoop-Steenstra, H., Schaeffer, N. C., & Van der Zouwen, J. (Eds.). (2002). *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview*. New York: Wiley.
- Mazeland, H. (2003). *Inleiding in de Conversatieanalyse*. Bussum: Coutinho.
- McLellan, E., MacQueen, K. M., & Neidig, J. L. (2003). Beyond the Qualitative Interview: Data Preparation and Transcription. *Field Methods*, 15(1), 63-84.
- Mishler, E. G. (1986). *Research Interviewing: Context and Narrative*. Cambridge, Mass.: Harvard University Press.
- Moerman, M. (1965). Ethnic Identification in a Complex Civilization: Who Are the Lue? *American Anthropologist*, 67(5), 1215-1230.
- Moore, R. J. (2004). Managing Troubles in Answering Survey Questions: Respondents' Uses of Projective Reporting. *Social Psychology Quarterly*, 67(1), 50-69.
- Nikander, P. (2000). "Old" Versus "Little Girl" a Discursive Approach to Age Categorization and Morality. *Journal of Aging Studies*, 14(4), 335-358.
- O+S (2004). *Naar Burgerschap in Amsterdam: Diversiteits- en Integratiemonitor 2004*. Amsterdam: Dienst Onderzoek en Statistiek, Gemeente Amsterdam.
- O+S (2005). *Amsterdam in Cijfers 2005*. Amsterdam: Dienst Onderzoek en Statistiek, Gemeente Amsterdam.
- O+S (2006). *Verkiezingen 2006 Tweede Kamer: Definitieve Uitslag Amsterdam*. Amsterdam: Dienst Onderzoek en Statistiek, Gemeente Amsterdam.

- Oakley, A. (1981). Interviewing Women: A Contradiction in Terms. In H. Roberts (Ed.), *Doing Feminist Research* (pp. 30 - 61). London: Routledge & Kegan Paul.
- Oksenberg, L., Cannell, C., & Kalton, G. (1991). New Strategies for Pretesting Survey Questions. *Journal of Official Statistics*, 7(3), 349-365.
- Ongena, Y. P. (2005). *Interviewer and Respondent Interaction in Survey Interviews*. Amsterdam: Vrije Universiteit Doctoral Dissertation
- Potter, J., & Hepburn, A. (2005). Qualitative Interviews in Psychology: Problems and Possibilities. *Qualitative Research in Psychology*, 2(4), 281-307.
- Potter, J., & Wetherell, M. (1987). *Discourse and Social Psychology Beyond Attitudes and Behaviour*. London: Sage Publications.
- Price, B. (2002). Laddered Questions and Qualitative Data Research Interviews. *Journal of Advanced Nursing*, 37(3), 273 - 281.
- Psathas, G. (1999). Studying the Organization in Action: Membership Categorization and Interaction Analysis. *Human Studies*, 22(2-4), 139-162.
- Puchta, C., & Potter, J. (1999). Asking Elaborate Questions: Focus Groups and the Management of Spontaneity. *Journal of Sociolinguistics*, 3, 314-335.
- Puchta, C., & Potter, J. (2004). *Focus Group Practice*. London: Sage Publications.
- Rapley, M., McCarthy, D., & McHoul, A. (2003). Mentality or Morality? Membership Categorization, Multiple Meanings and Mass Murder. *British Journal of Social Psychology*, 42, 427-444.
- Rapley, T. J. (2001). The Art(Fulness) of Open-Ended Interviewing: Some Considerations on Analysing Interviews. *Qualitative Research*, 1(3), 303-323.
- Rawlins, W. K. (1992). *Friendship Matters: Communication, Dialectics, and the Life Course*. New York: Aldine de Gruyter.
- Riessman, C. K. (2003). Analysis of Personal Narratives. In J. A. Holstein & J. F. Gubrium (Eds.), *Inside Interviewing: New Lenses, New Concerns*. Thousand Oaks: Sage Publications.
- Rogers, C. R. (1951). *Client-Centered Therapy*. London: Constable.
- Rosch, E., & Lloyd, B. B. (1978). *Cognition and Categorization*. Hillsdale, N.J. New York: L. Erlbaum Associates.
- Roulston, K. J. (2006). Close Encounters of the 'Ca' Kind: A Review of Literature Analysing Talk in Research Interviews. *Qualitative Research*, 6(4), 515-534.
- Roulston, K. J., Baker, C. D., & Liljestrom, A. (2001). Analyzing the Researcher's Work in Generating Data: The Case of Complaints. *Qualitative Inquiry*, 7(6), 745-772.

Probing Behaviour in Open Interviews

- Rubin, H. J., & Rubin, I. (2005). *Qualitative Interviewing : The Art of Hearing Data* (2nd ed.). Thousand Oaks: Sage Publications.
- Sacks, H. (1972). An Initial Investigation of the Usability of Conversational Data for Doing Sociology. In D. Sudnow (Ed.), *Studies in Social Interaction* (pp. 31-74). New York: The Free Press.
- Sacks, H. (1984). Notes on Methodology. In J. M. Atkinson & J. Heritage (Eds.), *Structures of Social Action: Studies in Conversation Analysis* (pp. 21-27). Cambridge: Cambridge University Press.
- Sacks, H. (1995). *Lectures on Conversation Volume I*. Oxford: Blackwell Publishing.
- Schaeffer, N. C., & Maynard, D. W. (2003). Standardization and Interaction in the Survey Interview. In J. A. Holstein & J. F. Gubrium (Eds.), *Inside Interviewing: New Lenses, New Concerns*. Thousand Oaks: Sage Publications.
- Schegloff, E. A. (1993). Reflections on Quantification in the Study of Conversation. *Research in Language and Social Interaction*, 26(1), 99-128.
- Schegloff, E. A. (2007). A Tutorial on Membership Categorization. *Journal of Pragmatics*, 39(3), 462-482.
- Seidman, I. (2006). *Interviewing as Qualitative Research. A Guide for Researchers in Education and Social Sciences*. (3rd ed.). New York: Teachers College Press.
- Sharrock, W. W. (1974). On Owning Knowledge. In R. Turner (Ed.), *Ethnomethodology, Selected Readings*. Harmondsworth: Penguin Books.
- Silverman, D. (1998). *Harvey Sacks: Social Science & Conversation Analysis*. New York: Oxford University Press.
- Silverman, D. (2006). *Interpreting Qualitative Data: Methods for Analysing Talk, Text and Interaction* (3rd ed.). London: Sage Publications.
- Snow, D. A., Zurcher, L. A., & Sjoberg, G. (1982). Interviewing by Comment: An Adjunct to the Direct Question. *Qualitative Sociology*, 5(4).
- Spradley, J. P. (1979). *The Ethnographic Interview*. New York: Holt, Rinehart & Winston.
- Springwood, C. F., & King, C. R. (2001). Unsettling Engagements: On the Ends of Rapport in Critical Ethnography. *Qualitative Inquiry*, 7(4), 403 - 417.
- Sturges, J. E., & Hanrahan, K. J. (2004). Comparing Telephone and Face-to-Face Qualitative Interviewing: A Research Note. *Qualitative Research*, 4(1), 107-118.
- Suchman, L., & Jordan, B. (1990). Interactional Troubles in Face-to-Face Survey Interviews. *Journal of the American Statistical Association*, 85(409), 232-241.

- Tanggaard, L. (2007). The Research Interview as Discourses Crossing Swords: The Researcher and Apprentice on Crossing Roads. *Qualitative Inquiry*, 13(1).
- Uhrenfeldt, L., Paterson, B., & Hall, E. O. C. (2007). Using Videorecording to Enhance the Development of Novice Researchers' Interviewing Skills. *International Journal of Qualitative Methods*, 6(1), 36-50.
- Van 't Hof, S. (2006). *From Text to Talk: Answers and Their Uptake in Standardised Survey Interviews*. Utrecht: LOT.
- Van den Berg, H. (1994). *On the Processing of Answers to Open Questions: A Contribution to the Methodology of Open Interviewing*. Paper presented at the Spiegel van de samenleving; Zesde sociaal wetenschappelijke Studiedagen: Siswo, Amsterdam.
- Van den Berg, H. (1996). Frame-Analysis of Open Interviews on Inter-Ethnic Relations. *Bulletin de Méthodologie Sociologique*, 53, 5-32.
- Van den Berg, H., & Carabain, C. L. (2003, May 15-18, 2003). *Assertions in Opinion Surveys: On the Limits of a Popular Question Format*. Paper presented at the The Annual Conference of the American Association for Public Opinion Research, Nashville, Tennessee.
- Van den Berg, H., Mehciz, M., Houtkoop-Steenstra, H., & Holleman, B. (2002). *Opiniemaken of Opiniemeten; De Rol van Stellingvragen in Markt- en Opinieonderzoek*. Amsterdam: Stichting voor Culturele Studies.
- Van den Berg, H., & Van der Veer, K. (1986). *Ideologie en Massamedia*. Amsterdam: Vrije Universiteit Doctoral Dissertation.
- Van den Berg, H., Wetherell, M., & Houtkoop-Steenstra, H. (Eds.). (2003). *Analyzing Race Talk: Multidisciplinary Approaches to the Interview*. Cambridge: Cambridge University Press.
- Van der Drift, J., & Derksen, W. (1985). Interviewstijlen en Interviewresultaten. *Mens en Maatschappij*, 60(2), 187-197.
- Van der Zouwen, J. (2002). Why Study Interaction in the Survey Interview? Response from a Survey Researcher. In D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer & J. Van der Zouwen (Eds.), *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview* (pp. 47-65). New York: Wiley.
- Van der Zouwen, J., Dijkstra, W., & Smit, J. H. (1991). Studying Respondent-Interviewer Interaction: The Relationship between Interviewing Style, Interviewer Behavior, and Response Behavior. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz & S. Sudman (Eds.), *Measurement Errors in Surveys*. (pp. 419-437). New York: Wiley.
- Van der Zouwen, J., & Smit, J. H. (2006). Effective Repair in Survey Interviews. *Quality & Quantity*, 40(2), 245-258.

Probing Behaviour in Open Interviews

- Viterna, J. S., & Maynard, D. W. (2002). How Uniform Is Standardization? Variations within and across Survey Research Centers Regarding Protocols for Interviewing. In D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer & J. Van der Zouwen (Eds.), *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview* (pp. 365-397). New York: Wiley.
- Watson, D. R. (1978). Categorization, Authorization and Blame-Negotiation in Conversation. *Sociology*, 12(1), 105-113.
- Webb, E., Campbell, D., Schwartz, R., & Sechrest, L. (1966). *Unobtrusive Measures: Nonreactive Research in the Social Sciences*. Chicago: Rand McNally & Company.
- Weiss, C. H. (1968). Validity of Welfare Mothers' Interview Responses. *Public opinion quarterly*, 32, 622-623.
- Weiss, C. H. (1970). Interaction in the Research Interview; the Effects of Rapport on Response. *Proceedings of the social statistics section of the American Statistical Association*, 17-20.
- Weiss, R. S. (1994). *Learning from Strangers: The Art and Method of Qualitative Interview Studies*. New York: The Free Press.
- Wengraf, T. (2001). *Qualitative Research Interviewing: Biographic, Narrative and Semi-Structured Methods*. London: Sage Publications.
- Wetherell, M. (2003). Racism and the Analysis of Cultural Resources in Interviews. In H. Van den Berg, M. Wetherell & H. Houtkoop-Steenstra (Eds.), *Analyzing Race Talk: Multidisciplinary Approaches to the Interview* (pp. 11-30). Cambridge: Cambridge University Press.
- Wowk, M. T., & Carlin, A. P. (2004). Depicting a Liminal Position in Ethnomethodology, Conversation Analysis and Membership Categorization Analysis: The Work of Rod Watson. *Human Studies*, 27(1), 69-89.

Summary in Dutch / Samenvatting

Doorvraagdgedrag in open interviews:

een veldexperiment over de effecten van doorvraagtactieken op de kwaliteit en inhoud van de verkregen informatie

Deze studie gaat over het gedrag van interviewers in open interviews en de mogelijke effecten van dat gedrag op de informatie die door middel van deze interviews wordt verzameld. In de ‘Interview samenleving’ (Atkinson & Silverman, 1997) waarin we leven, loopt iedereen wel op één of andere manier aan tegen interviews. We worden geïnterviewd, we houden zelf interviews en we kijken naar interviews op televisie.

Het interview komt niet alleen zeer vaak voor in het dagelijks leven, maar het wordt nog altijd op grote schaal gebruikt in sociaalwetenschappelijk onderzoek. In het hedendaagse kwalitatieve onderzoek, is het interview waarschijnlijk nog steeds de belangrijkste methode (Cassell, 2005).

Aangezien het interview op zo’n grote schaal wordt gebruikt, zou men verwachten dat er veel methodologisch onderzoek is naar effecten van interviewergedrag. Voor survey-interviews met vragen met gesloten antwoordmogelijkheden wordt aan deze verwachting ruimschoots voldaan. Voor het doen van open interviews, interviews met vragen zonder vooraf bepaalde antwoordmogelijkheden, is er opvallend weinig onderzoek gedaan naar effecten van interviewergedragingen. De vele introducties en handboeken voor open interviews zijn vooral gebaseerd op autobiografische anekdotes van de auteur. Het doel van mijn onderzoek is om inzicht te verkrijgen in de effecten van interviewergedrag in open interviews. De probleemstelling van dit onderzoek is toegespitst op het beoordelen van de effecten van verschillende interviewerstrategieën voor open interviews op de kwaliteit en inhoud van de verkregen informatie.

Het type interviews waar dit onderzoek zich op richt, zijn semigestructureerde interviews met voornamelijk open vragen, de zogenaamde *open interviews*. Een open interview kan worden gebruikt om te onderzoeken welke betekenissen een bevolkingsgroep toekent aan een bepaald concept. Dit kan als doel op zich, zoals vaak het geval is in kwalitatieve surveys, maar het open interview kan ook worden gebruikt in een vooronderzoek om de operationalisering van een grootschalige survey met gesloten vragen te informeren. Open interviews kunnen ook worden gebruikt na afloop van een enquête om de interpretatie van de antwoorden van de respondenten mogelijk te maken.

In dit onderzoek gaat het om ‘concept clarification interviews’. Aan volwassen autochtone Amsterdammers werden vragen gesteld over begrippen die te maken hebben met sociale categorisering, namelijk het begrip ‘Amsterdammer’, het begrip ‘vriend’ en het begrip ‘allochtoon’. Het gaat om drie zeer verschillende begrippen. ‘Amsterdammer’ is

een relatief neutraal topic, terwijl ‘vriend’ een persoonlijk/intiem en ‘allochtoon’ een maatschappelijk controversieel topic is. Tijdens de interviews probeerden de interviewers helder te krijgen hoe de respondenten deze begrippen definieerden en gebruikten.

Aangezien in open interviews de antwoordcategorieën niet vooraf zijn bepaald, kan het antwoord van de respondent alle kanten opgaan. De respondent kan onvolledig, te uitgebreid, te irrelevant en te onduidelijk antwoorden. Daarom zullen interviewers, vanaf het stellen van de initiële vraag, moeten improviseren en reageren op de al gegeven antwoorden. Dientengevolge is juist het doorvragen het belangrijkste interviewergedrag in een open interview. Doorvragen wordt hierbij gedefinieerd als het reageren op het antwoord van de respondent met als doel om zo veel mogelijk, zo specifiek mogelijk, zo uitgebreid mogelijk en zo diepgaand mogelijk relevante informatie te verzamelen. Goed doorvraagdgedrag van de interviewer is daarmee essentieel voor het bereiken van het algemene doel van een open interview: kwalitatieve informatie te verzamelen van zo hoog mogelijke kwaliteit. De interviewstrategie, waar in de probleemstelling over wordt gesproken, is in het onderzoek daarom toegespitst op het doorvraagdgedrag.

Zoals vrijwel al het interviewergedrag in open interviews is ook het doorvraagdgedrag in open interviews zelden systematisch bestudeerd en nog belangrijker, zeer zelden bestudeerd met behulp van een experimenteel design. Voor survey-interviews met gesloten vragen, is dikwijls vastgesteld dat de kwaliteit en de inhoud van de ontvangen informatie kwetsbaar is voor inadequaat doorvraagdgedrag. In survey-interviews neemt doorvraagdgedrag een relatief bescheiden plaats in omdat het doorvragen voornamelijk gericht op het repareren van ‘mismatches’ tussen de antwoordcategorieën en de antwoorden van de respondenten (Ongena, 2005). In open interviews moeten interviewers meestal doorvragen omdat antwoorden van de respondent op de initiële vraag zelden volledig zijn. Dit verschil leidt ertoe dat de resultaten van onderzoek naar interviewerstrategieën in survey-interviews met gesloten vragen niet toepasbaar zijn voor open interviews. Daarom is het zinnig om het doorvraagdgedrag in open interviews te onderzoeken.

Om de hiaten in kennis op het gebied van de effecten van doorvragen in open interviews op te vullen is de onderzoeksvraag hierop gericht: “Wat zijn de effecten van verschillende doorvraagtactieken in open interviews op de kwaliteit en de inhoud van de verkregen informatie?”

Om deze onderzoeksvraag te beantwoorden, is het begrip doorvraagtactiek gedefinieerd als een *houding* van de interviewer *ten opzichte van de antwoorden* van de respondent, en *de reactie als gevolg van* deze houding. Een doorvraagtactiek verschilt van een interviewstijl. Een interviewstijl wordt gebruikt gedurende het hele interview en is gericht op de persoon, terwijl de doorvraagtactiek gericht is op het antwoord en alleen wordt

gebruikt in reactie op antwoorden. Ik heb een onderscheid gemaakt tussen drie doorvraagtactieken, gebaseerd op noties die te vinden zijn in een breed scala van literatuur over interviewen:

De gereserveerde doorvraagtactiek. Hierin expliciteert de interviewer zijn of haar houding niet. Deze doorvraagtactiek komt voort uit de logica dat de interviewer zo min mogelijk de respondent moet beïnvloeden en zoveel mogelijk ruimte dient te bieden. Dit wordt soms beargumenteerd door survey-methodologen en dan is de achterliggende redenatie dat verstoringen van de vergelijkbaarheid voorkomen dienen te worden. Daarnaast wordt deze gereserveerdheid soms ook beargumenteerd vanuit narratieve interviews. In dat geval is de redenatie minder gericht op mogelijke verstoringen, maar meer om de geïnterviewde de vrijheid te bieden om het eigen verhaal te vertellen. Natuurlijk is bij beide methoden de uitvoering anders dan bij de gereserveerde doorvraagtactiek, maar het basisidee wordt dus gedeeld.

De meegaande doorvraagtactiek. Hierin neemt de interviewer het antwoord voor lief en laat dit expliciet merken in reacties. Deze tweede tactiek kent enige parallellen met wat in de kwalitatieve interviewliteratuur vaak wordt aangeduid met empathisch interviewen. Hoewel de beschrijving van dit empathische interviewen meestal nogal vaag is en vaak meer een morele stellingname dan een methodologische, kan het goed worden uitgelegd op een Rogeriaanse manier: de respondent moet altijd worden benaderd vanuit een onvoorwaardelijk positieve houding (Rogers, 1951). Pas door het meegaan in het verhaal kan er worden gezorgd voor het ‘openen’ van de respondent. Dit is dan ook de logica voor de meegaande doorvraagtactiek. Belangrijk verschil tussen Rogers onvoorwaardelijk positieve houding en de meegaande doorvraagtactiek is hierbij wel dat de doorvraagtactiek is gericht op de *informatie die de respondenten verschaffen* in plaats van op de persoon van de respondent *zelf*.

De prikkelende doorvraagtactiek. Hierin neemt de interviewer het antwoord niet zomaar voor lief en laat dit expliciet merken in reacties. Ook voor deze doorvraagtactiek zijn er parallellen te vinden in de interviewliteratuur. Bijvoorbeeld het antagonistische interview, waarbij een onvoorwaardelijk kritische houding wordt aangenomen. Het idee hierachter is dat door respondenten kritisch te bejegenen, zij ook kritischer op zichzelf worden en beter gaan antwoorden.

Doorvraagtactieken verschillen van doorvraagtechnieken. Doorvraagtechnieken zijn alle type doorvragen die een interviewer kan gebruiken in reactie op een antwoord van een respondent met als doel om meer, meer relevante, meer uitgebreide, meer specifieke of meer persoonlijke informatie te vergaren. In de literatuur over interviewen worden deze technieken op verschillende manieren geclassificeerd. In dit onderzoek heb ik ervoor gekozen om doorvraagtechnieken in eerste instantie te classificeren op basis van de vorm en het doel en pas in tweede instantie op basis van de inhoud. Ik maak onderscheid

tussen verschillende hoofdvormen: minimale technieken (zoals hummen en echoën), halve vragen of zinnen, herhalingen van de vraag (letterlijk of hergeformuleerd), verzoek om uitweiding, verzoek om specificatie (met verschillende subtypes op basis van de inhoud, bijvoorbeeld een verzoek om specificatie met behulp van een voorbeeld), een nieuwe doorvraag (met eveneens verschillende subtypes op basis van de inhoud, bijvoorbeeld een verzoek om uitleg van de redentatie of een verzoek om ervaringen), uitingen (van twijfel, begrip of gebrek aan begrip), parafrasering of samenvatting en een reflectie. In Box 1.2 op pagina 16 is te zien hoe deze technieken vanuit de interviewliteratuur specifiekere worden uitgewerkt.

In dit onderzoek ligt de focus dus op de effecten van de doorvraagtactieken op zowel de kwaliteit als de inhoud van de verkregen informatie. De eerste afhankelijke variabele, de *kwaliteit van de verkregen informatie*, is geoperationaliseerd met behulp van vijf indicatoren: *relevantie*, *diepte*, *hoeveelheid*, *reikwijdte* en *specificiteit*.

De eerste indicator, relevantie, is gebaseerd op de inverse van relevante informatie: off-topic informatie. Wanneer een interview relatief veel irrelevante informatie bevat, is de kwaliteit duidelijk lager dan wanneer het interview verhoudingsgewijs minder off-topic antwoorden bevat.

De tweede indicator, diepte van de informatie, wordt in de interviewliteratuur zeer veel genoemd als één van de belangrijkste kwaliteitscriteria van open interviews. Helaas is er weinig overeenstemming wat nu precies diepte is. Daarom heb ik diepte van de informatie geoperationaliseerd als de hoeveelheid respondentbeurten met nieuwe empirische gegevens over het persoonlijke leven of gevoelens van de respondent. De drie indicatoren, hoeveelheid, reikwijdte en specificiteit, zijn gebaseerd op codes die voortvloeien uit een aanpassing van het analytisch instrumentarium van ‘membership categorisation analysis’ (Sacks, 1995; Housley & Fitzgerald, 2002). Dit instrumentarium is opgebouwd rond het fenomeen dat mensen in interactie voortdurend categoriseren en dat de gesprekspartner deze categorisering als vanzelfsprekend begrijpen. Dergelijke alledaagse categorieënsystemen noemt Harvey Sacks ‘membership categorisation devices’: verzamelingen van common sense categorieën en de regels die mensen aanwenden om te interpreteren welke *categorieën* en eventuele *predicaten* van toepassing zijn. Enkele concepten en regels uit de ‘membership categorisation analysis’ zijn gebruikt tijdens het coderen van de interviewtranscripten om te duiden welke categorieën en bijbehorende predicaten respondenten gebruiken. Die coderingen vormen de basis voor het construeren van de indicatoren (hoeveelheid, reikwijdte en specificiteit van de informatie). De hoeveelheid informatie bijvoorbeeld is onder meer geoperationaliseerd als het aantal predicaten dat door een respondent wordt toegeschreven aan een bepaalde sociale categorie.

De tweede afhankelijke variabele, *de inhoud van de interviews*, wordt op twee manieren geoperationaliseerd. De eerste manier is *de inhoud van de antwoorden* op enkele specifieke vragen uit het interview zoals vragen naar het aantal genoemde vrienden in Amsterdam en het hebben van positieve en/of negatieve gevoelens ten aanzien van allochtonen. De tweede manier waarop de inhoud van de interviews is geoperationaliseerd is gebaseerd op een classificatie van type predicaten uit de literatuur over ‘membership categorisation analysis’. Deze predicaat types worden door respondenten gebruikt bij de betekenisgeving aan de begrippen ‘Amsterdammer’, ‘vriend’ en ‘allochtoon’. De predicaat types zijn zodanig abstract geformuleerd dat ze toepasbaar zijn voor de drie genoemde interview topics.

Het design dat is gebruikt om de onderzoeksvraag te beantwoorden is het experimenteel design. In een veldexperiment werden zes groepen van zes onervaren interviewers uitgebreid geïnstrueerd in één van de drie doorvraagacties. De training van de interviewers was zeer intensief en bestond uit een tweedaagse instructie, video's, oefeningen en rollenspelen. De twee dagen werden afgesloten met een op video opgenomen proefinterview en deze werd becommentarieerd door alle mede-interviewers en mij. Het was de bedoeling dat alle interviewers zes respondenten uit Amsterdam zouden interviewen. Twee interviewers hebben vijf respondenten geïnterviewd, wat resulteert in een totaal van 214 interviews. De interviewers ontvingen na twee interviews persoonlijke feedback over hun interviewgedrag.

Een aantal maatregelen is getroffen om de interne validiteit zo goed mogelijk te waarborgen. Ten eerste werden de interviewers willekeurig toegewezen aan één van de zes groepen. Ten tweede werd een aantal kenmerken van het interviewgedrag zoveel mogelijk gestandaardiseerd: de inleiding, de interviewstijl, de volgorde van de onderwerpen, een deel van de volgorde van de vragen, de formulering van de initiële vragen, de betreffende doorvraagactie, het ‘stoppen-met-doorvragen-criterium’ en zelfs de afronding van het interview. Alleen de keuze van de doorvraagtechniek was (binnen de kaders van de betreffende doorvraagactie) vrij. Ten derde werden de respondenten willekeurig toegewezen aan de interviewers en dus aan één van de drie experimentele groepen.

Om ervoor te zorgen dat ook de geldigheid in andere situaties zo hoog mogelijk is, werden verschillende maatregelen genomen. In de eerste plaats zijn alle drie de doorvraagacties mede gebaseerd op concepten uit een breed scala aan interviewliteratuur. Het zijn dus geen vreemde idiosyncratische manieren van doorvragen. In de tweede plaats is de keuze van de drie interview topics mede gebaseerd op de overweging dat het van belang is om na te gaan in hoeverre effecten van doorvraagacties topicafhankelijk zijn. In de derde plaats waren de interviewers zoals

deze geselecteerd werden voor dit onderzoek niet veel anders dan interviewers die werkzaam zijn in andere sociaalwetenschappelijke studies, dat wil zeggen jonge beginnende interviewers met een sociaalwetenschappelijke achtergrond. In de vierde plaats werden de respondenten geselecteerd op basis van een steekproef uit het online panelbestand van onderzoeksbureau O+S Amsterdam. Doel daarvan was het realiseren van voldoende variatie van respondenten naar criteria als leeftijd, sekse, opleidingsniveau en politieke voorkeur.

De audio-opnames van de interviews zijn zorgvuldig getranscribeerd en gecontroleerd. In totaal negen codeurs hebben ruim twee maanden intensief gecodeerd. Hierbij is in de eerste maand het interviewgedrag gecodeerd en werden vanaf de tweede maand de antwoorden van de respondent van codes voorzien. Ten behoeve van een toetsing van de inter-codeur-betrouwbaarheid, zijn voor het interviewgedrag drie interviewtranscripten en voor de antwoorden van de respondent vier transcripten door *alle* codeurs gecodeerd. De resultaten van die toetsen lieten over het algemeen zien dat het codeersysteem en de codering betrouwbaar is. Door daarnaast gebruik te maken van jurisprudentiemethodes door intensieve intervisie en supervisie is er ook sprake van een hoge validiteit en betrouwbaarheid van de coderingen bij de interviews die niet door alle codeurs zijn gecodeerd.

Uit de analyse van de codering van het interviewgedrag blijkt dat de interviewers zeer goed presteerden. Er was nauwelijks sprake van foutief interviewgedrag, zoals fouten met het overslaan van vragen, het zodanig herformuleren van vragen dat er sprake is van betekenisverandering of het impliciet of expliciet suggereren van een antwoord. Daarnaast behaalden de interviewers goede resultaten bij het volgens de instructies uitvoeren van het vraaggedrag en het doorvraaggedrag. De experimentele variabele doorvraagactiek, werd in alle groepen volgens instructie uitgevoerd.

De effecten die de doorvraagactiek hadden op de kwaliteit en de inhoud van de ontvangen gegevens werden geanalyseerd met behulp van voornamelijk variantie analyses en chi-kwadraat toetsen. Deze analyses tonen aan dat er bijna geen verschillen zijn tussen de drie doorvraagactieken voor open interviews in termen van de vijf dimensies van kwaliteit van de verkregen informatie: relevantie, diepte, hoeveelheid, uitgebreidheid en specificiteit. Dus ongeacht welke doorvraagactiek gebruikt werd, op elke dimensie en op elke indicator van kwaliteit scoorde de doorvraagactieken vrijwel gelijk.

Voor de inhoud van de informatie uit de interviews zijn er eveneens amper effecten zichtbaar als gevolg van de drie doorvraagactieken. Tussen de inhoud van de gecodeerde antwoorden op de vragen uit de prikkelende, meegaande en gereserveerde tactieken zijn namelijk nauwelijks verschillen die gerelateerd kunnen worden aan de doorvraagactieken. Ook uit de analyse van de predicaat types die respondenten toeschrijven aan de betreffende sociale categorieën blijkt dat er geen verschillen zijn tussen de drie

doorvraagactieken. Het gebrek aan significante afwijkingen geldt voor alle drie de onderwerpen van sociale categorisering, dus voor het neutrale, het persoonlijke en het controversiële onderwerp.

Deze resultaten zijn geruststellend voor vele interviewers en onderzoekers die gebruik maken van open interviews, omdat het gebruik van verschillende doorvraagactieken dus geen significante effecten heeft op de inhoud of de kwaliteit van de verkregen gegevens.

Het ontbreken van effecten van de doorvraagactieken is echter wel heel opvallend, mede gelet op de stereotype bezwaren ten aanzien van open interviews. Volgens die bezwaren is het open interview erg subjectief, staat het bloot aan heel veel bias en is het veel te afhankelijk van de persoon van de interviewer.

Deze resultaten zijn in eveneens in tegenspraak met wat in sommige onderzoeksliteratuur over interviewergedrag in open interviews wordt beweerd. Met name vanuit de conversatie-analyse wordt benadrukt dat in het beantwoorden van open vragen er onvermijdelijk sprake is van co-constructie (Baker, 2004; Rapley, 2001; Roulston, et al., 2001; Roulston, 2006). Dit wil zeggen dat bewust of onbewust de interviewer de antwoorden mede vormgeeft.

De resultaten worden nog meer intrigerend, wanneer men die in verband brengt met resultaten uit veel experimenten in survey-interviews met gesloten vragen. De algemene conclusie uit die experimenten is dat slechts geringe veranderingen in de vraagformulering, vraagvolgorde of het gedrag van de interviewer, grote effecten op de inhoud en kwaliteit van de antwoorden kunnen hebben.

De vraag rijst, hoe het mogelijk is dat er - ongeacht de aard van het vraagtopic - geen verschillen zijn tussen de drie doorvraagactieken wat betreft de kwaliteit en de inhoud van de verkregen informatie. De paradox tussen mijn bevindingen en de resultaten van methodologisch onderzoek naar survey-interviews zou verklaard kunnen worden door de hypothese dat respondenten in survey-interviews met gesloten vragen beperkt zijn in hun antwoordgedrag als gevolg van de vooraf bepaalde antwoordcategorieën. Bij de beantwoording van open vragen hebben respondenten de mogelijkheid om eerder geformuleerde (delen van) antwoorden te nuanceren, aan te vullen, ze in perspectief te plaatsen, of zelfs helemaal te herroepen. Het vermoeden is gegrond dat een open interview respondenten een zekere robuustheid biedt ten opzichte van mogelijke interveniërende factoren, zoals een specifieke doorvraagactiek. Het ontbreken van verschillen in de effecten die deze doorvraagactieken hebben op de kwaliteit en de inhoud van de informatie, suggereert dat de vorm van open interviews ertoe leidt dat respondenten zich vrijer uitspreken. Een gevolg daarvan is dat hun antwoorden daarmee meer geloofwaardig ('credible' volgens Guba en Lincoln (1985)) worden dan vaak wordt gedacht. Door gebruik te maken van hun eigen woorden om

open vragen te beantwoorden, zijn de respondenten beter in staat zich uit te drukken in de open interviews. Dit leidt tot de conclusie dat open vragen relatief ruime de feedbackmogelijkheden aan respondenten biedt om tijdens het doorvragen het eigen antwoord te completeren. En dit zorgt er vervolgens voor dat het open interview robuuster is dan gestandaardiseerde interviews die gebruik maken van gesloten vragen.

Dit betekent echter niet dat simpelweg het houden van een open interview voldoende is om die robuustheid te genereren. Een waarschuwing is dus op zijn plaats, want er zijn mijns inziens wel enkele noodzakelijke voorwaarden. De interviewer moet namelijk feedback kunnen genereren. Dit kan uitsluitend door goed interviewergedrag en dat kan door onderzoekers alleen worden afgedwongen door te letten op drie essentiële voorwaarden: een goede voorbereiding van het onderzoek, een goede training van de interviewers en een scherpe monitoring van de interviewers. Indien aan deze voorwaarden is voldaan en de interviewers hun werk goed hebben gedaan, durf ik de verwachting uit te spreken dat in alle interviews met een substantieel aantal open vragen, ongeacht hun specifieke onderwerp of uiteindelijke doel, verschillende doorvraagtactieken *geen* invloed hebben op de kwaliteit en de inhoud van de informatie.

Faculty of Social Sciences, VU University Amsterdam

DISSERTATION SERIES

- M.M. Meijer, (2004), *Does Success Breed Success? Effects of News and Advertising on Corporate Reputation*. Aksant Academic Publishers: Amsterdam.
- J.J. Woldendorp, (2005), *The Polder Model: From Disease to Miracle? Dutch Neo-corporatism 1965-2000*. Amsterdam.
- Y.P. Ongena, (2005), *Interviewer and Respondent. Interaction in Survey Interviews*, Yfke Ongena, Amsterdam.
- M.E.D. Lamboo, (2005), *Integriteitsbeleid van de Nederlandse Politie*. Eburon: Delft.
- T.E. Aalberts, (2006), *Politics of Sovereignty*. Vrije Universiteit Amsterdam.
- E. Zwart, (2006), *In Pursuit of Comfort. The Transnationalisation Process of Malaysian Chinese Small and Medium Enterprises*. Vrije Universiteit Amsterdam.
- H.Haenen, (2006), *Afrikaans Denken. Ontmoeting, Dialoog en Fricie. Een Filosofisch Onderzoek*. Buijten en Schipperheijn Motief, Amsterdam.
- Claver, (2006), *Commerce and Capital in Colonial Java. Trade Finance and Commercial Relations between Europeans and Chinese, 1820s-1942*. Vrije Universiteit Amsterdam.
- S. van der Pas, (2006), *Intergenerational Relationships of Older Adults. Family Structure, Contact and Norms*. Vrije Universiteit Amsterdam.
- P. Mika, (2007), *Social Networks and the Semantic Web*. SIKS dissertation series No. 2007-03.
- K.E. Knibbe, (2007), *Faith in the Familiar. Continuity and Change in Religious Practices and Moral Orientations of the South of Limburg, The Netherlands*. Vrije Universiteit Amsterdam.
- C.L. Carabain, (2007), *Taking Too Much for Granted? A Study on the Measure of Social Attitudes*. Gildeprint Drukkerijen B.V., Enschede.
- L.M. Cruz, (2007), *Repression and Rebellions in Southern Mexico. The Search for a Political Economy of Dignity*. Dutch University Press, Amsterdam.
- C.E. de Vries, (2007), *European Integration and National Elections. The Impact of EU Issue Voting on National Electoral Politics*. Vrije Universiteit Amsterdam.
- E. Kolthoff, (2007), *Ethics and New Public Management. Empirical Research into Effects of Businesslike Government on Ethics and Integrity*. Boom Juridische Uitgevers.
- R. Vliegthart, (2007), *Framing Immigration and Integration. Facts, Parliament, Media and Anti-Immigrant Party Support in the Netherlands*. Vrije Universiteit Amsterdam.
- L.L.J. van Noije, (2007), *The Democratic Deficit Closer to Home. Agenda Building Relations between Parliament and the Press, and the Impact of European Integration, in the United Kingdom, the Netherlands and France*. Vrije Universiteit Amsterdam.
- R. de Man, (2007), *Where to Draw the Line? Antecedents and Consequences of the Integration of Work in Home Life*. Vrije Universiteit Amsterdam.
- F. Krijtenburg, (2007), *Cultural Ideologies of Peace and Conflict: a Socio-Cognitive Analysis of Giriya Discourse (Kenya)*. Vrije Universiteit.
- R. van Steden, (2007), *Privatizing Policing. Describing and Explaining the Growth of Private Security*. BJU Legal Publishers.
- L. Reijn, (2008), *Identity at Work. Control and Commitment in Post Bureaucratic Organisations*. Vrije Universiteit Amsterdam.
- B. Vis, (2008), *Biting the Bullet or Steering Clear? Politics of (Not-) Unpopular Welfare State Reform in Advanced Capitalist Democracies*. Vrije Universiteit Amsterdam.
- M.J.M. de Koning (2008), *Zoeken naar een 'Zuivere' Islam. Geloofsbeleving en Identiteitsvorming van Jonge Marokkaans-Nederlandse Moslims*. Bert Bakker.
- P. Wiepking (2008), *For the Love of Mankind. A Sociological Study on Charitable Giving*. Labor Grafimedia.
- A. Supriyono (2008), *Buruh Pelabuan Semarang. Pemogokan-Pemogokan Pada Zaman Colonial Belanda Revolusi dan Depublic 1900-1965*. Vrije Universiteit Amsterdam.
- G. van Noort (2008), *Bricks versus Clicks. A Self-Regulation Perspective on Consumer Responses in Online and Conventional Shopping Environments*. Vrije Universiteit Amsterdam.
- W. Stam (2008), *Does Networking Work? A Study of Social Capital and Performance among New Ventures in the Dutch Open Source Software Industry*. Vrije Universiteit Amsterdam.

- A. Wigger (2008), *Competition for Competitiveness: The Politics of the Transformation of the EU Competition Regime*. Rozenberg Publishers.
- Z. van der Wal (2008), *Value Solidity. Differences, Similarities and Conflicts between the Organizational Values of Government and Business*. Vrije Universiteit Amsterdam.
- H.C. van Vugt (2008), *Embodied Agents from a User's Perspective*. Vrije Universiteit Amsterdam.
- J. ter Horst (2008), *Weaving into Cambodia. Trade and Identity Politics in the (post)-Colonial Cambodian Silk Weaving Industry*. Vrije Universiteit Amsterdam.
- A. Berhe (2008), *A Political History of the Tigray People's Liberation Front (1975-1991): Revolt, Ideology and Mobilisation in Ethiopia*. Vrije Universiteit Amsterdam.
- H. van den Heuvel (2008), *Between Optimism and Opportunism. Deconstructing 'African Management' Discourse in South Africa*. Vrije Universiteit Amsterdam.
- K. Lasthuizen (2008), *Leading to Integrity. Empirical Research into the Effects of Leadership on Ethics and Integrity*. Vrije Universiteit Amsterdam.
- I. Reci (2008), *Unemployed and Scarred for Life. Longitudinal Analyses of How Unemployment and Policy Changes Affect Re-employment Careers and Wages in the Netherlands, 1980-2000*. Vrije Universiteit Amsterdam.
- L.C. Horn (2009), *The Transformation of Corporate Governance Regulation in the European Union*. Vrije Universiteit Amsterdam.
- B.E. Büscher (2009), *Struggles over Consensus, Anti-Politics and Marketing*. Vrije Universiteit Amsterdam.
- H.M. Koolma (2009), *Verbalen en Prestaties, een Onderzoek naar het Gedrag van Woningcorporaties*. Vrije Universiteit Amsterdam.
- J.J. van Wijk (2009), *Moving beyond Heroes and Winners*. Vrije Universiteit Amsterdam.
- H.A. Binnema (2009), *How Parties Change. EU Integration and the Flexible Response of Political Parties*. Vrije Universiteit Amsterdam.
- E. Grassiani (2009), *Morality and Normalcy in Asymmetrical Conflict: Distancing, Denial and Moral Numbing among Israeli Conscripts in Everyday Practices of Occupation*. Vrije Universiteit Amsterdam.
- G. van Koningsbruggen (2009), *Processing and Acceptance of Threatening Health Information: the Effects of Self-Affirmation*. Vrije Universiteit Amsterdam.
- D. van Bergen (2009), *Suicidal Behaviour of Migrant Women in the Netherlands*. Vrije Universiteit Amsterdam.
- J. Perry (2009), *Goodwill Hunting: Accounting and the Global Regulation of Economic Ideas*. Vrije Universiteit Amsterdam.
- Linden (2009), *Besmet. Levenslopen en Motieven van Extreem-Rechtse Activisten in Nederland*. Vrije Universiteit Amsterdam.
- Schakel (2009). *A Postfunctionalist Theory of Regional Government. An Inquiry into Regional Authority and Regional Policy Provision*. Vrije Universiteit Amsterdam.
- Plaisier (2009). *Work and Mental Health: Studies on the Impact Characteristics, Social Roles and Gender*. Vrije Universiteit Amsterdam.
- J. Roeland (2009). *Selfation: Dutch Evangelical Youth between Subjectivization and Subjection*. Vrije Universiteit Amsterdam.
- Vliegthart (2009), *Transnational Actors and Corporate Governance Regulation in Postsocialist Europe*. Vrije Universiteit Amsterdam.
- S. Bhagavatula (2009). *Weaving Social Networks. Performance of Small Rural Firms in India as an Outcome of Entrepreneurs' Social and Human Capital*. Vrije Universiteit Amsterdam.
- van Essen (2009). *Seeking a Balance: The Emergence of New Public Management in Health Care Regulation*. Vrije Universiteit Amsterdam.
- J. Soons (2009). *Love Life and Happiness: A Study of Partner Relationships and Well-being in your Adulthood*. Vrije Universiteit Amsterdam.
- J. de Raadt (2009). *Contestable Constitutions: the Politics of Institutional Conflict in Post-Communist East Central Europe*. Vrije Universiteit Amsterdam.
- B. Tasew (2009). *Methaphors of Peace and Violence in the Folklore Discourses of South Western Ethiopia: A Comparative Study*. Vrije Universiteit Amsterdam.

EXPECTED

- D. Gberie (2010). *From Chains to Neckties: Liberated Africans and the Making of Modern Sierra Leone*. Vrije Universiteit Amsterdam.
- M. Guiaux (2010). *Social adjustment to widowhood: Changes in personal relationships and loneliness before and after partner loss*. Vrije Universiteit Amsterdam.
- W. Reedijk (2010). *Effecten van religieuze dialogen tussen Joden, Christenen en Moslims op eigen identiteit*. Vrije Universiteit Amsterdam.
- M. Hellendoorn (2010). *Opkomst en Neergang van het Kostwinnaarschapstelsel*. Vrije Universiteit Amsterdam.
- E. Ersanilli (2010). *Comparing Integration. Host culture adoption and ethnic retention among Turkish immigrants and their descendants in France, Germany and the Netherlands*. Vrije Universiteit Amsterdam.
- T.S. Salverda (2010). *Sugar, Sea and Power: How Franco-Mauritians balance continuity and creeping decline of their elite position*. Vrije Universiteit Amsterdam.

Probing Behaviour in Open Interviews

A Field Experiment on the Effects of Probing Tactics on Quality and Content of the Received Information

This research is focussed on the tactic the interviewer uses while probing. A probing tactic is an interviewer stance towards the answers of the respondent and the reaction following from this stance. In this research three different probing tactics are distinguished: the accommodating, the encouraging and the challenging.

A field experiment was held to determine whether these probing tactics have an effect on the quality and content of the received information. In order to tackle any possible topic-dependency of the effects of the probing tactics, the interviews that were held contained three different topics of social categorisation: 'Amsterdammers', friends and 'allochthons'.

The results of this study are remarkable in the light of what is known about effects of interviewer behaviour in closed interviews and what is often supposed in literature on open interviewing.

Gerben Moerman (1976) graduated in Sociology of Non-Western Societies from Leiden University in 1999. He conducted the present study at the Interuniversity Graduate School of Psychometrics and Sociometrics (IOPS) and the Faculty of Social Sciences at the VU University. He is currently employed as a lecturer at the department of Sociology and Anthropology at the University of Amsterdam.

ISBN: 9789086594436