Eurobarometer: measurement instruments for opinions in Europe

Saris, Willem E. (Ed.); Kaase, Max (Ed.)

Veröffentlichungsversion / Published Version
Sammelwerk / collection

GESIS

ZUMA

NACHRICHTEN

Spezial Band 2

Eurobarometer: Measurement Instruments for Opinions in Europe

Willem E. Saris & Max Kaase (Eds.)

ZUMA, Mannheim 1997

Zentrum für Umfragen, Methoden und Analysen (ZUMA)

ZUMA is a member of the "Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen e.V." (GESIS). GESIS is a member of the "Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz" (WGL).

Vorsitzender des Trägervereins ZUMA e.V.: Prof. Dr. Max Kaase

Direktor: Prof. Dr. Peter Ph. Mohler

Anschrift
B2,1
PO Box 12 21 55
D-68 072 Mannheim

Telefon	0621/1246-100	
Email	maka@medea.wz-berlin.de	
Internet	http://www.social-science-gesis.de/	*GESIS*
	http://www.zuma-mannheim.de/	*ZUMA*

ISBN 3-924220-12-3

Table of contents

Preface

For more than two decades the biannual representative sample survey studies conducted on behalf of the Directorate General X of the European Commission - the so-called Eurobarometer - have been a major data source for comparative research in the social sciences, especially political science, although they are mainly conducted for policy counseling reasons (for an assessment of these studies see Reif and Inglehart, 1991; the analytic potential of the Eurobarometers is documented in Kaase and Newton, 1995).

In order to speed up the policy information process, Karlheinz Reif, at that time responsible for the Eurobarometer operation at the European Commission, in 1993/1994 had proposed the normal Eurobarometers to be supplemented by a regular tracking study on European attitudes on a monthly basis. Since this objective could not be reached by face to face personal interviews for reasons of cost and speed of data gathering, the option had to be validated to gather the tracking material through personal telephone inerviews. An important question to be answered there was whether, given the uneven distribution of private telephone ownership in the member countries of the European Union, it was nevertheless possible to produce valid, reliable and representative data based on telephone ownership household samples.

Fortunately, this question could be answered, through the ingenuity and entrepreneurship of Karlheinz Reif, on experimental basis. For one, Reif assembled a group of methodologically proficient social scientists to advise him on this problem.

Second, through an investment by the Berlin-based FORSA institute and its director Manfred Güllner, as well as additional resources from the European Commission, an experimental setting could be constructed (for details see chapters 1 and 2 of this book) which was able to confront data from face to face and telephone interviews covering representative samples of the respective populations, and to add a telephone panel component in three countries (Belgium, France and Spain) to the regular April 1994 Eurobarometer (EB 41.0), which permitted tests of possible mode effects comparing face to face and telephone interviews and tests of the functional equivalence of questions across countries.

The following chapters present analyses of such problems and draw conclusions for future research. Given the fact that all over Europe especially in commercial market research the number of telephone interviews is drastically increasing and that of personal face to face interviews is decreasing, the findings from the experimental research described in this book reach far beyond the Eurobarometers. The editors thus very much hope that this study will contribute to improving the quality of social research in the commercial as well as in the academic realm.

As was mentioned before, this work owes a lot to Karlheinz Reif and to Manfred Güllner. Next to the editors, Roger Jowell, Hans-Dieter Klingemann, and Hermann Schmitt were the principal investigators in the advisory group. The editors also appreciate the decision by ZUMA to publish the manuscript in its special series on methodological problems in social research, and in particular the great effort by Suzanne Kabel of the University of Amsterdam and of Jolantha Müllner of ZUMA in producing this manuscript.

Willem E. Saris, Max Kaase
Amsterdam and Berlin, November 1997

CHAPTER 1

THE EUROBAROMETER
A TOOL FOR COMPARATIVE SURVEY RESEARCH

MAX KAASE AND WILLEM E. SARIS

1.1 Development of comparative survey research in political science

The fourties in the United States witnessed the development and implementation of techniques which were capable of reliably assessing social as well as political attitudes, beliefs and behaviours of large-scale populations through small samples of respondents. The application of probability theory in the development of sampling procedures in connection with the emergence of the standardised personal interview laid the ground for what has since long become a normal tool of social research: surveys with a limited number of respondents, say 2000, producing information that can be reliably generalised to the population from which the sample was drawn, almost independent of the size of that population.

Soon after the end of World War II this methodology spread from the United States to Western Europe where it quickly became the most frequently used instrument in market, social and political research. However, the expertise required to apply this tool intelligently, reasons of cost and not the least a certain conservatism in the academic world for quite a while made surveys a rather scarce phenomenon in sociology and political science. Only as the field of comparative government slowly transformed into comparative politics with its more pronounced interest in political processes and, as one central element in it, citizen orientations, survey data began to assume the kind of important place they now possess in this field (for electoral research as one example see the overview in Thomassen 1994). More than anything else it was the 'Civic Culture' comparative study of political orientations in the United States, the United Kingdom, (West) Germany, Italy and Mexico by Almond and Verba (1963) which revealed the analytical potential of this new methodology for political science.

There are many reasons why this variant of empirical research did not spread much more quickly after the initial success of the 'Civic Culture'. Some of the problems are spelled out succinctly in a book by Szalai and Petrella (1977) discussing five major East-West comparative survey studies which had been conducted in the 60s, partly under the auspices of the Vienna-based European Co-ordination Centre for Research and Documentation in the Social Sciences (Vienna Centre). The experiences from those studies, as summarised in the Szalai and Petrella book, give a good flavour of the theoretical and methodological challenges as well as the practical difficulties this kind of comparative survey research has had to confront (see also Verba, 1971). It is little wonder then that the Political Action study (Barnes et al., 1979)

conducted in seven European nations and in the United States in the 70s (with a replication in three of those countries around 1980; see Jennings et al., 1990) for a long time stood alone in university-based comparative survey research in political science. Since then, the 1981 and 1990 European Values Studies (Harding et al., 1986; Ester et al., 1993) as well as the Beliefs in Government project (Kaase and Newton, 1995) are further hallmarks of comparative survey research and its uses in political science.

What is still missing in this list is the seminal work of Ronald Inglehart on value change in the industrialised democracies of the West (Inglehart, 1977; 1990; 1997). These longitudinal analyses became only possible because Inglehart had been able to influence the content and the availability for secondary analyses of one source that the Commission of the European Communities in Brussels had established in the early 70s as a regular data collection under the supervision of Jacques-René Rabier: the Eurobarometer (Reif and Inglehart, 1991).

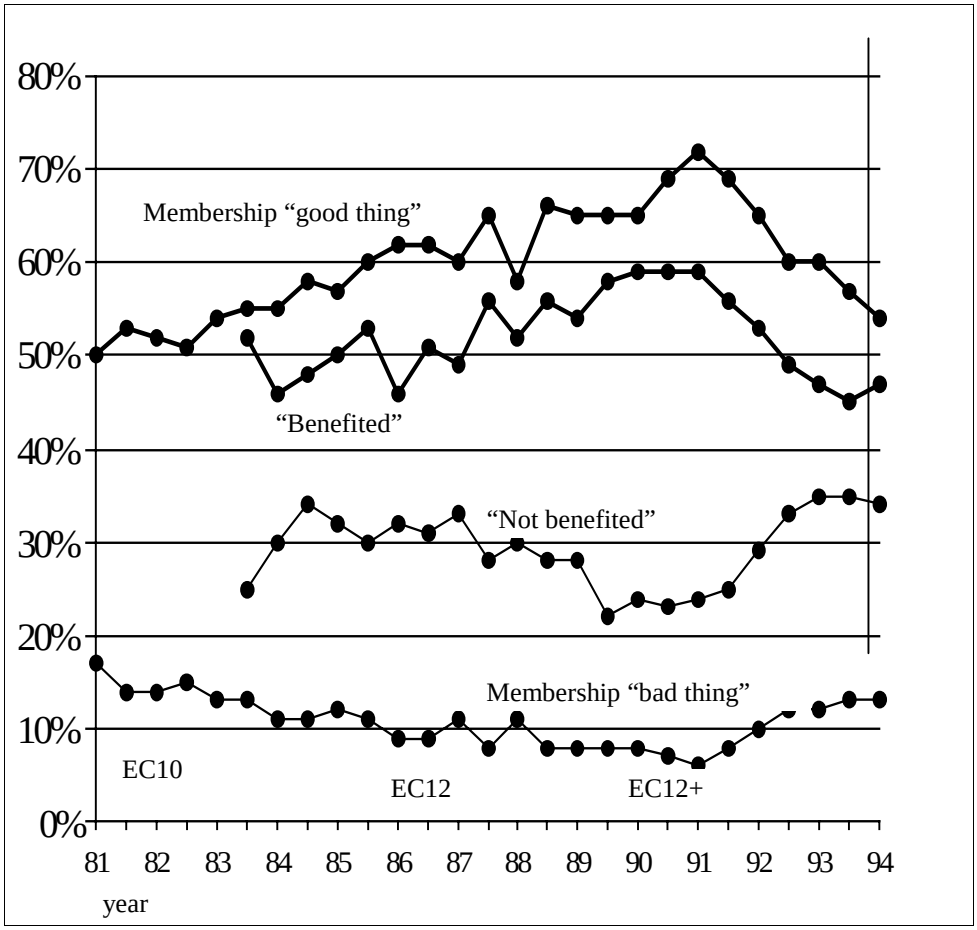
The Eurobarometers have been conducted twice a year since the autumn of 1973 in all member states of the European Union (formerly the European Community). As the EU grew in membership, so grew the number of countries in which the Eurobarometer was carried out (at the time of this writing in the spring of 1996 the survey is done in 15 EU member states). The most obvious advantage of these 24 years of regular surveys is its potential to look at change over time; without this data base the Inglehart study of value change would have been impossible, as is true for the analyses in the context of the Beliefs in Government project.

Obviously, the primary interest of the European Commission (as it was renamed after the 1993 Maastricht treaty) in funding these regular surveys was to gain information on the topic most dear to its heart: the process of European unification, as reflected in the attitudes and beliefs of the people in the EU member states (for the most recent, thorough analysis of this field see Niedermayer and Sinnott, 1995). Thus, other information regarded more pertinent for the analytical concerns of the academic social science community could be introduced into the Eurobarometers only in small doses and never covered more than a limited fraction of what one would have liked to find there. Still, there can be no question that by a far margin the Eurobarometers presently are the most valuable data source for comparative survey analysis on Western Europe in political science, a fact witnessed by the enormous number of scholarly publications using these data.

The data of the Eurobarometers can be used in three different ways. The first way is the use of the results in order to make comparisons across time. The second way is the use of the results in order to compare between the different countries. The third is a combination of the two. All three types of use will be illustrated for topics which have been in the Eurobarometer now for more than 20 years.

In figure 1.1 the responses of the people on the questions whether “the membership of the EU was a good thing for your country ?” and whether “your country has benefited of the membership of the EU?”. This figure gives the development of the support for the EU by the people in the EU countries from 1981 till 1994. It illustrates clearly that there are considerable changes through time with respect to support for the EU. For legitimacy of the EU policies such data are very relevant. The recent downwards movement gives the politicians of course a lot of concern.

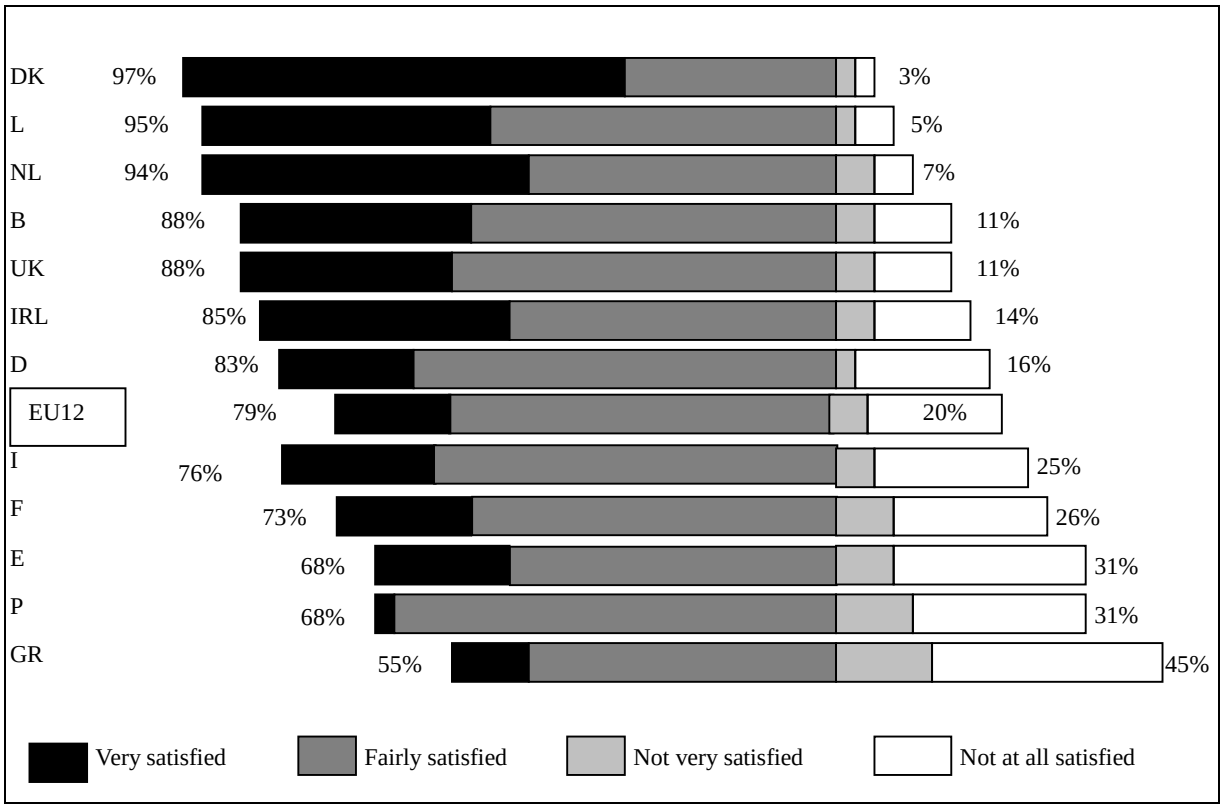
Figure 1.1 Support and perceived benefits of the EU membership from 1981-1994



For the illustration of the second type of use we have chosen a very different example namely the answer to the question: "How satisfied are you with your life in general ?" In figure 1.2 the results for this question in July 1994 are given. This figure illustrates nicely the large differences which exist between the different countries with respect to life satisfaction. Also this issue is politically interesting because it is hard to imagine that such differences (if they are real) can exist for a long time in the EU without leading to serious consequences like migration etc.

Finally the third way to use the Eurobarometer data is illustrated by the results on the question " Do you approve the introduction of a single currency in the EU? In figure 1.3 a comparison is made with respect to the answers of the populations of three countries and the EU as a whole for the period 1990 till 1994.

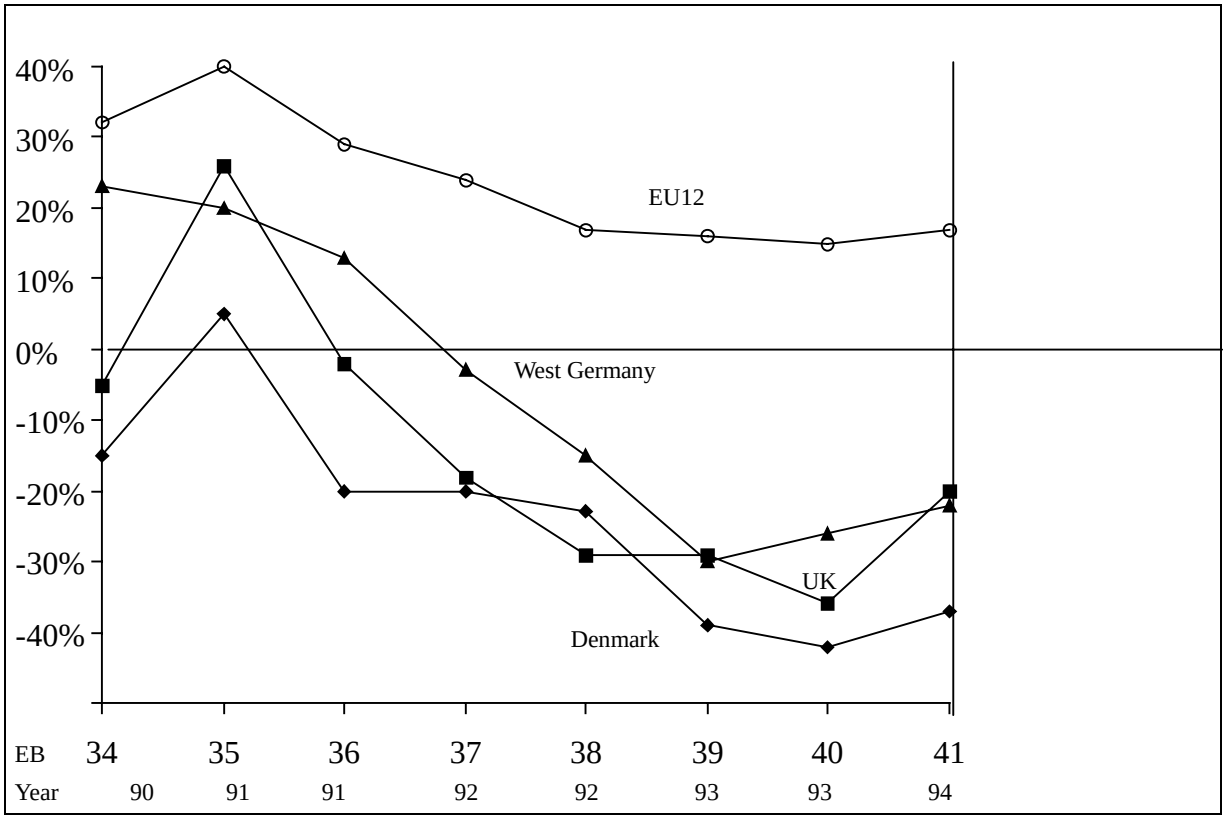
Figure 1.2 Satisfaction with life in general in different countries of the EU



This figure illustrates that all three countries have since 1992 a negative net approval while the EU as a whole is still at the positive side. In figures like these one can see the differences in development of the different countries through time.

As these examples demonstrate this kind of information is very valuable for the European Commission in order to see what the support is for different policies and the EU as a whole in the different countries through time. This information is even so important that the Eurobarometer is undergoing important changes. The information needs of the European Commission are changing in the direction of more short-time data. After having gone through the traumatic - because unanticipated - experience of the narrow margins in supportive post-Maastricht plebiscites in a variety of EU countries (in Denmark, even a second plebiscite became necessary after the rejection of the first), the European commission developed a thorough interest in obtaining information on the swing in public moods in the EU much more quickly than through the inflexible and time-consuming method of face to face interviewing which was used till that time for the Eurobarometer studies.

Figure 1.3 The "net approval " of the single European currency in Denmark, West Germany, UK and the EU as a whole



Therefore, the standard Eurobarometer has been extended by a device which reveals changes and new emphases in public moods vis a vis the EU much more quickly than the classical Eurobarometer was able to do. The resulting tracking instrument consists (in principle) of smaller monthly surveys of the people in the member states of the EU. Obviously, this purpose could not be achieved by the traditional face to face interview. Therefore, the method of choice to cope with both problems was to switch to computer assisted telephone interviews.

This choice became possible because - according to Eurobarometer 41 of spring 1994 - only East Germany with about 50 percent, Portugal with about 68 percent and Ireland with about 70 percent are below the 80 percent telephone density mark. In addition, computer-assisted telephone interviewing (CATI) has been perfected to a point where its advantages in being able to do a large number of interviews quickly, to have the collected data immediately ready for analysis and to exert full control over the interviewing process through supervision, are so obvious that this procedure is used in 29% of all quantitative studies in Europe (see ESOMAR, 1996:15).

While the payoffs of using telephone interviewing are apparent, they nevertheless need to be balanced against the requirement to obtain and their ability to produce valid and reliable data

on the populations in question. It is for example questionable if the data remain comparable through time (figure 1.1 and 1.3) moving from one data collection mode to the other. It is also questionable if the results remain comparable across countries (figure 1.2 and 1.3) given that the samples will differ with respect to ownership of telephone.

These are methodological problems which can only be satisfactorily answered by some kind of an experimental research. Therefore, the offer by the Berlin-based FORSA institute under the direction of Manfred Guellner to do a representative telephone survey in the 12 EU member countries in 1994 at the same time the usual face to face Eurobarometer (41) was in the field to test the potential for future telephone interviewing for purposes of the European Commission, was gratefully accepted. On this basis, a complex research design was developed by the scholars contributing to this book to test a set of important methodological issues related to the questions under which circumstances one can expect equivalent and valid information both from face to face and from telephone interviews and how to improve the overall quality of the Eurobarometers.

1.2 Design of the methodological study

Given the relevance of the Eurobarometer data for all people interested in developments in Europe and for social science research, the changes which were expected to occur in the data collection of the Eurobarometer data were sufficient reasons for a number of survey researchers[1] to suggest that methodological research should be done to evaluate the consequences of the change of mode of data collection. The argument for such research was based on the existing knowledge on mode effects. For example Groves (1989) gave 9 possible reasons why one can expect differences between the face to face data collection and the telephone procedures. The reasons given can be condensed as follows:

1. The coverage of the population will be different for face to face interviews from telephone interviews, since those people who do not have a telephone will not be part of the general population from which the sample should be drawn. In Europe, this difference can be substantial because in some areas and countries household telephone density is close to 100% while in other areas and countries the coverage is closer to 50%. In general, it has been found that this bias leads to considerable differences in responses on several dimensions (Groves and Kahn, 1979; Cannel et al., 1987).
2. The field work of the organisations doing the surveys can and usually will be quite different with respect to the interviewers used, their training and supervision, the number of times that a respondent is contacted, and the rules by which a refusal is accepted. Differences in these management aspects will lead to differences in responses and nonresponse and consequently to differences in findings.
3. The mode of data collection itself can also lead to different results. It is possible that people react differently to the same question in a telephone interview than in a face to face

1 This idea was brought forward by Roger Jowell, Max Kaase, Hans-Dieter Klingenman, Willem E. Saris and Hermann Schmidt.

interview. For instance, it has been found that open-ended questions result in more elaborated answers in face to face interviews than in telephone interviews. Also, more acquiescence and an extremeness bias might be expected (Groves, 1989). However, the general picture is that these mode effects, after correcting for all other factors, are rather small (De Leeuw and van der Zouwen, 1988).

Also, mode-connected effects are possible, that is effects which might occur due to the fact that changes in the approach are necessary depending on the mode of interview, and that these changes will matter. For example, the use of show cards is not possible in telephone interviews, and as a consequence the procedure for complex questions has to be adjusted. Commonly used in telephone surveys to cope with this problem is a two-step approach where first a small number of crude categories is presented which are later split up into more differentiated ones. The idea here is to obtain the same kind of precision by telephone as in personal interviews where show cards are used to present ten or so categories at the same time. These mode-related changes in the questionnaire can lead to substantial discrepancies in the results, as has been shown by Groves and Kahn (1979), Miller (1984), and Monsees and Massey (1979).

This brief overview indicates that there are three major reasons why a change from face to face to telephone interviews will most likely lead to different results. As indicated above, one can expect that the total difference (T) between face to face and telephone interview responses in percentages or in mean score will be equal to the difference due to coverage (C) plus the difference due to difference in nonresponse (N) plus the difference due to the mode of data collection (M):

$$T = C + N + M \qquad (1)$$

This equality should hold exactly for each question separately if populations would be studied. Because samples are used we can not expect a perfect equality. Even if the differences due to coverage, nonresponse and mode would be zero the samples can provide different response percentages because of the randomness of the two samples involved. This difference due to sampling fluctuation can not be studied separately and will therefore be integrated in the effect of the field organisation (M) but later we will test if the differences deviate significantly from zero in order to take this point into account.

Given the possible confusion due to discrepancies in results between the different studies, it has been suggested by those co-operating in writing this book that a methodological study of this issue should be done in order to be able to assess the errors in both procedures more precisely and to find ways to adjust the findings in such a way that the results become comparable. In this chapter, the design will be described which has been chosen by the research group to study the biases in the different approaches and to develop procedures to correct for these biases.

Beforehand, one point should be mentioned: Due to the fact that the Eurobarometer by now has been in use for more than 20 years, one might assume that its findings are valid and reliable, and that deviations from these results in telephone interviews are to be blamed on that method. This is, however, a naive view. Personal interviews have their own errors which can also cause deviations from the population parameters. Of course, this point will be given

some attention in this study, too, but the main emphasis is on the differences in the findings between the two methods.

1.3 The research design

There are two principally different ways to study mode effects. The first approach is experimental and keeps all characteristics of the data collection under control in order to merely study the mode effect (see Ekman, 1965; Mekrabian, 1968; Champness, 1972; Williams, 1974; 1977). Such experimental studies must face the challenge that they are a far distance away from the reality of daily survey research. As Groves (1989) suggests, this methodological approach is mainly aimed at assessing communication effects which is not the primary aim of the present study.

The completely opposite approach, called maximum telephone/maximum personal design, was used by Woltman et al. (1980) and Hochstim (1967); it suggests that one should compare two data collection procedures for the same population and herewith maximise the possibilities for measuring the impact of the two different modes. This can be done with independent random samples or with panel data.

In the present study, the last design could be used because the Berlin-based FORSA research institute offered to collect data for a limited set of questions in all countries through telephone interviewing, while at about the same time the INRA organisation conducted the standard Eurobarometer 41 face to face. This design is a good simulation of the future situation in Europe when two studies will be done about the same time by two different organisations each using a different data collection mode. This design gives the opportunity to estimate the total difference in responses for two specific organisations (T) for all questions present in both studies.

This design, however, has as a major weak point that too many differences will exist between the various approaches, and that one cannot determine which factor possibly causes these differences. Besides, there is the difficulty that all the field work-connected errors are specific for the survey organisations which are involved in the given study. Therefore, should these organisations be exchanged later on, other differences in the results should be expected. Given this lack of strength in the design, also a third approach, a panel element, has been included in this study.

Here, the respondents were first confronted with the normal face to face Eurobarometer questionnaire. In addition, they were asked whether they had a telephone. Those who had a phone were called back after about a week and were asked to respond to a small number of questions already put to them before in the Eurobarometer. This panel design offers better insights into the effects of the two different sources of error, as follows.

First, when the telephone owners and non-owners are compared, an estimate can be obtained of the effect of telephone ownership on the responses to the relevant variables. In this comparison no other variables intervene because the same people are studied and all questions are presented in a face to face interview. So the only possible explanation for

differences is telephone ownership, and thus a good estimate of the coverage error which will occur, is provided (C).

A second effect one can study with this design relates to the mode of data collection since one can compare the answers of the respondents to the same questions in the personal interview and in the telephone interview. This evidence is not so strong as in the case of the comparison of telephone owners and non-owners because there are other factors besides the mode effects which can come into play. One concern has to do with the repeated observations. It is possible that the people want to be consistent by trying to reproduce their answer from a week ago. However, Van Meurs and Saris (1989) have demonstrated that the respondents cannot remember their responses anymore after about 20 minutes of being exposed to questions. In the present study there was at least a gap of one week between the two interviews so that one can expect that memory effects will only play a minimal role, if any.

Furthermore, people may have changed their opinion or behaviour in the time between the first and the second interview. However, for the topics which are asked in the Eurobarometers, such changes are unlikely in a period of one to two weeks unless dramatic events happened but that was not the case in the research period.

A third problem is that the sample of those respondents co-operating in the telephone interview is not the same as the one participating in a normal telephone interview. Selective loss of respondents thus might cause a different group to refuse co-operation after a face to face interview than in a normal telephone interview.

Although these factors all need to be kept in mind, it is nevertheless quite unlikely they will have a strong effect on the existing mode effects. If this assumption is valid, which we think is very likely, then the study of the responses of the same persons in the personal interview and in the telephone interview will provide good estimates of possible mode effects (M).

This design does not allow for an independent estimate of the effects of the fieldwork organisation on the nonresponses (N). However, one could estimate this effect when a combination of a maximum telephone maximum personal design and a panel design is used. The direct comparison of personal interviews with telephone interviews gives an estimate of T. Using the panel design, C and M can be estimated. Using the combination of the two designs, the effect of the difference in nonresponse due to different organisational procedures will be:

$$N = T - C - M \qquad (2)$$

It should be mentioned here that the coverage error (C) is an estimate which is for the largest part independent of the organisation that did the research because the effect is determined by the difference between owners and non-owners in the population. This difference will only minimally be influenced by the specific procedure used for data collection, as long as this procedure is not completely flawed.

The same point can be made for the estimate of the mode effect M as was argued above. On the other hand, the estimates of N and T are clearly determined by the organisations which perform the studies. The total difference varies directly with the difference due to

nonresponse which is produced by the two organisations in question. So general statements are difficult to make about these two components although they can be properly assessed for a specific case.

Furthermore, the estimates of the coverage error and of the mode effect can also vary with the questions being used. Telephone non-owners can differ in their opinions on certain questions, and this will lead to differential effects although for other questions the differences can again be very small. In the literature, some questions have been mentioned to be more effected than others, like open-ended questions, questions requiring a heavy cognitive burden such as long questions or questions with a large number of categories, and also questions which are normally asked with a show card, a procedure presently not available in telephone interviewing. Given the effects of the type of question asked, it will be necessary to discuss this problem in the present mode effect study, too.

1.4 The questions used

The questions which have been asked in the different data collections of this study can be subdivided into five groups. Below, an overview is given of the formulation of the questions of the Eurobarometer which have also been chosen for the telephone interviews. Adjustments and small differences in the questions for the telephone interviews are not documented here.

Group 1: The standard Eurobarometer questions

The first group consists of core standard Eurobarometer questions. Obviously, the results from these questions should be equivalent across the different studies. The following questions were chosen from the pool of all possible questions as the most important ones.

Life satisfaction
On the whole, are you very satisfied, fairly satisfied, not very satisfied or not at all satisfied with the life you lead? Would you say you are (READ OUT)

very satisfied
fairly satisfied
not very satisfied
not at all satisfied
———
DK/No answer

Satisfaction with democracy
On the whole, are you very satisfied, fairly satisfied, not very satisfied or not at all satisfied with the way democracy works in (your country)? Would you say you are (READ OUT)

very satisfied
fairly satisfied
not very satisfied
not at all satisfied
———
DK/No answer

Frequency of political discussions
When you get together with friends, would you say you discuss political matters frequently, occasionally, or never?

frequently
occasionally
never
———
DK/No answer

Motivation to persuade others
When you hold a strong opinion, do you ever find yourself persuading your friends, relatives or fellow workers to share your views? Does this happen (READ OUT)

often
from time to time
rarely
never
———
DK/No answer

Media exposure

About how often do you (SHOW CARD)
a) watch the news on television
b) read the news in daily papers
c) listen to the news on the radio

	every day	several times a week	once or twice a week	less often	never	DK/ No answer
News						
on TV	1	2	3	4	5	6
in daily papers	1	2	3	4	5	6
on the radio	1	2	3	4	5	6

Interest in EU politics
To what extent would you say you are interested in European politics, that is to say matters related to the (EC/EU): a great deal, to some extent , not much, or not at all?

a great deal
to some extent
not much
not at all
———
DK/No answer

Subjective sense of political information
All things considered, how well informed do you feel you are about the (EC/EU), its politics, its institutions? (READ OUT)

very well
quite well

not very well
not at all well
———
DK/No answer

Evaluation of EU membership for respondent's country
Generally speaking, do you think that (our country's) membership of the (EC/EU) is (READ OUT)

a good thing
a bad thing
neither good nor bad
———
DK/No answer

Benefit of EU membership for respondent's country
Taking everything into consideration, would you say that (our country) has on balance benefited or not from being a member of the (EC/EU)?

benefited
not benefited
———
DK/No answer

These questions have been asked in many Eurobarometers and will also be asked in the tracking studies. It is therefore necessary to know whether differences in responses will occur and whether they can be corrected.

Group 2: Different types of simple questions

There is also a number of questions which have not been asked literally in the same way in each Eurobarometer and will not be asked in each tracking study, but will continue to be asked occasionally in the future. These questions concern opinions on specific problems and knowledge about certain political issues. The content of these questions can be changed, but their format will usually not be changed. Therefore it makes sense to study the effect of these formats in the different modes. A distinction has been made between simple questions which are presented in this section, and complex questions which are covered in the next section.

Opinions
What is your opinion on each of the following proposals? Please tell me for each proposal, whether you are for it or against it. (READ OUT IN ROTATING ORDER)

a) There should be a European Monetary Union with one single currency replacing by 1999 the (national currency) and all other national currencies of the member states of the (EC/EU).
b) The (EC/EU) member states should work towards a common defence policy.
c) Any citizen of another (EC/EU) country who resides in (your country) should have the right to vote in local elections.

d) Any citizen of another (EC/EU) country who resides in (your country)should have the right to vote in European elections.
e) Any citizen of another (EC/EU) country who resides in (your country) should have the right to be a candidate in local elections.
f) Any citizen of another (EC/EU) country who resides in (your country) should have the right to be a candidate in European elections.
g) The (EC/EU) should be responsible only for matters that cannot be effectively handled by national, regional and local governments.

Knowledge of White Paper
Have you ever heard about the "White Paper" by the European Commission in Brussels about growth, competitiveness and employment in Europe?

Yes
No
———
DK/No answer

Information on European Parliament
Have you recently seen or heard, in the papers, on the radio or on TV, anything about the European Parliament?

Yes
No
———
DK/No answer

Knowledge about date of the next European elections
Do you know the date on which the next European election will take place in (your country), or not? (IF YES) On which date?

Yes, and the correct date
Yes, but the date mentioned is not correct
No, does not know the date

Only the last of these questions was included in the FORSA study. This means that for most of the questions the coverage errors and the mode effects can just be studied by using the answers of the respondents in Eurobarometer 41 and its panel component.

Group 3: Complex questions

Three more complex questions have been chosen for further study because such questions are likely candidates for mode effects (Groves 1989). These questions concern vote probabilities, reasons why people may not vote, and left-right self-placement.

The question on the probability to vote has a very difficult formulation which can lead to problems in telephone interviews.

The questions on the reasons for non voting can lead to problems not because of the complex formulation of the question but because of the complex response categories which are normally presented on a show card and have to be read out in a telephone interview.

Vote probability
In June 1994, the citizens of countries belonging to the (EC/EU), including the (nationality of the respondent), will be asked to vote to elect members of the European Parliament. If there were such a "European election" tomorrow (for respondents under 18 years old add: "and you had a vote") would you certainly go and vote, probably go and vote, probably not go and vote, or certainly not go and vote? (IF VOTING IS COMPULSORY IN THE COUNTRY; ADD) "if the vote was not compulsory in our country"

Will certainly go and vote
Will probably go and vote
Will probably not vote
Will certainly not vote
Other answer (SPONTANEOUS)
———
DK/No answer

Reason for non-participation (only asked of those respondents who are not certain to vote)
What is the main reason why you might not go and vote at the next European elections in June 1994? (SHOW CARD - ONE ANSWER ONLY)

I am not interested in politics or elections
I am not interested in European elections
I lost interest in European matters
I have always been against Europe
I am against even more Europe
Not well enough informed to vote in European elections
I think the result is a foregone conclusion
Other reasons (SPONTANEOUS)
———
DK/No answer

Left - right position
This question in face to face interviews is often asked on a ten point scale ranging from extreme left to extreme right, the respondents being provided with a show card with the ten possibilities. For telephone interviews it has been suggested that the use of a two step question would be functionally equivalent (Miller, 1984; Monsees and Massey, 1979). This formulation, however, has also its risks. Sykes and Hoinville (1985) did not find differences, but Groves and Kahn (1979), when looking at other questions, did find that differences can occur.

We have introduced the two-step procedure in a split ballot experiment in the face to face study. If differences were detected already in the identical mode, one would not have to try the same question in a telephone survey because then the question formulation were the cause of deviations and not the mode of data collection. To test this effect, in the face to face

interview one of the following questions has been given to two randomly selected sub-samples.

Split Ballot A
When people talk about politics, the terms " left" and "right" are always used. We would very much like to ask you, whether you put yourself as rather "left" or rather "right"?

Rather "left"
Middle/neither nor (SPONTANEOUS)
Rather "right"
Refusal
———
DK/No answer

When rather 'left'
Please imagine for a moment a scale, from 1 to 5, where 5 means 'very left' and 1 'not very left'. Where would you put yourself?

not very left				very left
1	2	3	4	5

When rather 'right'
Please imagine for a moment a scale, from 1 to 5, where 5 means 'very right' and 1 'not very right'. Where would you put yourself?

not very right				very right
1	2	3	4	5

Split Ballot B
In political matters people talk of "the left" and "the right".
How would you place your views on a scale from one to ten, where 1 means very much to the left, and 10 very much to the right? You may, of course, use the numbers in between 1 and 10 in order to shade your opinion. (SHOW CARD. DO NOT PROMPT. IF CONTACT HESITATES, ASK TO TRY AGAIN)

left									right
1	2	3	4	5	6	7	8	9	10

No statement
DK/No answer

Group 4: Open-ended questions

Open-ended questions were introduced in the tracking study because they provide information about developments in public opinion without cues by the researchers through the use of response categories.

For the methodological study this type of question is interesting because it has been found that people provide less information in telephone interviews than in face to face interviews on open-ended questions (Groves and Kahn, 1979; Kormendi, 1988). The explanation given is

that telephone conversations are expected to be shorter. The consequence, however, can be that the responses differ in the different studies. Since these questions are of great concern for the tracking study, they have been asked in the two waves of the Eurobarometer 41 panel as well as in the telephone study of FORSA, thereby enabling comparisons between the two data collection modes.

Most important problem in our own country
Generally speaking, what is the most important problem facing (our country) today?

Most important problem in EU
Generally speaking, what is the most important problem facing (EC/EU) today?

Group 5: Background variables

Obviously, also a number of background variables have been asked in all three data collections. Here we mention only those questions which are comparable.

Age
How old are you (YEARS OF AGE)
(FORSA classification: 18-29; 30-44; 45-59; 60 and older)

Household goods
Do you or anyone else in your household own? (READ OUT)

Colour tv
Video recorder
Video camera*
Radio clock*
PC/ home computer
Still camera*
Electric grill*
Electrical deep fat fryer*
2 or more cars
Second home or a holiday home/flat
———
* not asked in both telephone studies

Social class
If you were asked to choose one of five names for your social class, which would you say you belong to? (SHOW CARD)

Middle class
Lower middle class
Working class
Upper class
Upper middle class
Refuse to be classified
Other
———
DK/No answer

After the questions have been specified, the existing options for analysis can be summarised in Table 1.1. This table shows that for the most questions all three kinds biases defined above can be studied. Only for the group 2 questions (simple questions) the analysis is limited to the coverage error and the mode effect because these questions have not been asked in the FORSA study. This table indicates that the chosen design allows for a large number of analyses.

1.5 Plan of the book

The book has started with an introductory chapter discussing the relevance of the topic and the design of the study. In the next four chapters the research design of the face to face and telephone study are evaluated, and the quality of these designs, the differences in sampling and their effects are analysed.

The second chapter will give a description of the research design for both studies. The basis of this chapter are the descriptions given by INRA and FORSA, the organisations which organised the data collections, about their procedures. This chapter provides basic information for anybody who uses the standard Eurobarometer or the tracking studies with respect to the way the data is collected. This description is more detailed than can be found anywhere else in the literature

The third chapter shows that both sets of samples differ considerably from the sets of populations they want to describe, on the demographic variables for which population information is available. Furthermore it is indicated how the different samples can be reweighed using optimal weights. Also, the efficiency of the various procedures and the quality of this reweighing for substantive variables is demonstrated. This chapter provides users of the different EB's with information how the unavoidable deviations in the background variables can be corrected and what the consequences of reweighing on other variables are.

Many people think that face to face studies can not be compared with studies using telephone interviewing. The fourth chapter shows how different the telephone owners and non owners really are with respect to a number of background variables and some substantive variables. The chapter indicates that there are indeed considerable differences. However, these differences will be placed in perspective in the next chapter.

In the fifth chapter of the first part of the book the estimates of the total difference between the two methods are decomposed in the following components: the effects of the coverage error, the nonresponse differences and the mode differences. These components have been estimated for three different countries for which the data are available: Belgium, France and Spain. Other than general opinion has it, the coverage error (due to the differences between owners and non-owners of telephones) is rather small while the differences due to the use of a different mode of data collection and of the differences in fieldwork of the two companies conducting the surveys is rather large. This result is of more general importance than only for this specific study.

Table 1.1 The possibilities for analysis in the study

Questions	EB 41.0	EB 41. Panel	FORSA	Possibilities for analysis
Standard questions				
Life satisfaction	+	+	+	T,C,M,N
Satisfaction with democracy	+	+	+	T,C,M,N
Political discussion	+	+	+	T,C,M,N
Persuade others	+	+	+	T,C,M,N
Political news (3x)	+	+	+	T,C,M,N
Interest in EU politics	+	+	+	T,C,M,N
Informed about politics	+	+	+	T,C,M,N
EU membership	+	+	+	T,C,M,N
EU benefit	+	+	+	T,C,M,N
Simple questions				
Opinions (7x)	+	+	-	C,M
Knowledge of white paper	+	+	-	C,M
Knowledge of parliament	+	+	-	C,M
Knowledge of elections	+	+	-	C,M
Complex questions				
Vote probability	+	+	+	T,C,M,N
Reasons for non-participation	+	+	-	C,M
Left-right 10 point scale	+	+	+	T,C,N
Left right 2 step scale	+	+	-	C
Open ended questions				
National problem	+	+	+	T,C,M,N
EU problem	+	+	+	T,C,M,N
Background questions				
Age	+	+	+	T,C,M,N
Household goods (4x)	+	+	+	T,C,M,N
Social class	+	+	+	T,C,M,N

In the second part of this book the attention goes to the comparability of the results across modes of data collection and across countries.

In the sixth chapter the latent class response model will be introduced, and a test will be done in order to evaluate how the mode of data collection effects the standard barometer questions. This is a very general issue in the literature. In some studies large differences have been found, in others not. Thus, it seems that much here depends on the questions. Therefore, the user has to know whether such effects occur when comparisons are made across studies using different modes of data collection.

The seventh chapter discusses the mode effect for a specific question: the open ended question on the most important problem facing one's own country and the EU. For this analysis the development of a general codebook for different countries has to be discussed first. Then a description of the results will be given comparing the results of the three studies in order to detect if the specification of the different problems is different for the different modes.

In the eighth chapter the possible differences between a 10 point scale for left-right orientation in a face to face and in a telephone interview will be discussed. Next, the consequences of the two step procedure for such questions will be considered on the basis of the available experimental data. This approach has been suggested being most adequate for telephone interviewing and is used in recent studies. In this chapter we show that this was a right or a wrong decision.

In the ninth chapter analyses have been done with respect to the comparability of the results across modes of data collection and across countries. It was found that for some questions comparisons across modes and countries can be made without any problem. For other questions it is found that the results can be compared across countries if the same mode of data collection is used. For other questions, like the frequently used satisfaction questions, it was found that and is explained why comparison across modes and across countries is not possible. Obviously, this result is highly relevant for researchers using these data.

In the last chapters ways are proposed which help to make the results obtained with the different designs as comparable as possible. This has to be done by two separate corrections: one for the sample differences, and one for the response mode effects.

In the tenth chapter procedures are presented to correct for sample differences and mode effects which have been discussed before. This is also illustrated by applying them to the results of the INRA and FORSA surveys.

Finally, in chapter 11, a summary of the results and practical recommendations will be given for users of the questions and data of the standard Eurobarometers and the tracking study with respect to the way in which they should use the data. Furthermore, recommendations will be formulated for correction of some of the questions in order to make them more comparable across modes of data collection and/or across countries.

CHAPTER 2

SAMPLE DESIGN AND CONSEQUENCES

PETER SCHUBERT AND ANGELIKA GREIL

2.1 Introduction

In chapter 1 it was indicated that in this book the results of two different approaches to the collection of survey data will be compared. One study, the standard Eurobarometer 41.0, used face to face interviews in 12 EU member states and was carried out by the research company INRA which normally collects the data for the Eurobarometers. The second study used as data collection method telephone interviews and has been carried out by FORSA in the same 12 EU member states. The purpose of this study was to assess the feasibility of telephone interviewing in the EU member states. By comparing the results of these two studies one can see how large the differences will be between the results of a face to face and a telephone survey in Europe. For methodological purposes the standard Eurobarometer study was augmented by a panel component in form of a telephone study also done by INRA in three countries in order to study the mode effects of the different approaches on a series of questions, the countries being Belgium, Spain and France.

Since the differences in sampling design and fieldwork in the three studies can have caused differences in the results, these differences in design and fieldwork will be described. After that the weights to correct for differences from the populations will be discussed, and finally a comparison of the response rates in the studies will be given in order to look at some fieldwork-related differences.

2.2 Fieldwork

Fieldwork of all three surveys took place in the spring of 1994. The populations to be studied in the face to face and in the telephone study were identical: the populations of the 12 EU member states aged 15 years and older. The target sample of the INRA surveys was 1000 interviews per country, and 500 interviews per country with the FORSA surveys. The INRA face to face - telephone panel study aimed only at a sample of 3 countries because of its special purpose.

2.2.1 The face to face study: The standard Eurobarometer 41.0

The Eurobarometer 41.0 (EB41.0) survey was carried out by specialised polling firms co-ordinated by INRA Europe, Brussels. From April 4th until May 6th, 1994, the respondents were contacted in their private homes and questioned in face to face interviews.

In order to select the respondents a stratified multi-stage random sample routine was used. The sampling frame for EB41 were the smallest enumeration units of the census in each country. Of these units a random sample was drawn stratified on the basis of the regions (Eurostat NUTS II level) and degree of urbanisation. In this way at least 100 primary sampling units were selected in each country. Within each sampling unit a starting address was drawn in a random fashion from an official address book or otherwise. In most countries more than 100 starting addresses were chosen. A randomly chosen increment determined the 10 more address per sampling unit in order to obtain a sample of approximately 1000 households. Within a household, the actual respondent was selected among those aged 15 or older by the criterion of who in the household had the next birthday or the Kish-grid.

The training of the interviewers was the responsibility of the national member of INRA. There are no agreed-upon supervision procedures between the national representatives of INRA except the general rule to visit each household only one more time if at the first time no interview could be done. After two failures the household was registered as nonresponse. Respondents who refused to co-operate have not been contacted once more for conversion purpose. If at a household no interview could be done, this address was substituted by a random walk procedure.

Table 2.1 Completion rate EB41.0

EB41.0	B[2]	DK	W-G	E-G	GR	ESP	F
Gross sample	2318	2739	2358	2208	2007	2354	2190
Net sample	1087	1005	1064	1058	1010	1003	1034
Completion rate (%)	47	37	45	48	50	43	47

	IRL	NIRL	I	Lux	NL	P	GB	EU12
Gross sample	2025	944	2636	1447	2632	1975	2155	29988
Net sample	1068	306	1058	625	1015	1002	1067	13402
Completion rate (%)	53	32	40	43	39	51	50	44.7

Table 2.1. displays the completion rate of the Eurobarometer 41.0 in the 12 EU countries (plus a separate listing for West Germany and East Germany). The gross sample is the number of addresses drawn according to the sampling frame. The net sample is the number of successfully completed interviews. The completion rate (in percentages) is the net sample's share of the gross sample. The table shows that the completion rate of the EB41.0 varies

2 Country abbreviations are: B = Belgium; DK = Denmark; W-G = West Germany; E-G = East Germany; GR = Greece; ESP = Spain; F = France; IRL = Ireland; NIRL = Northern Ireland; I = Italy; Lux = Luxembourg; NL = The Netherlands; P = Portugal; GB = Great Britain; EU12 = European Union (B, DK, W-G, GR, ESP, F, IRL, I, Lux, NL, P, GB)

between 37% and 51% which is not different from the results obtained normally in the Eurobarometer studies. For more detailed information about the procedure the methodological reports of INRA in the written Eurobarometer reports can be consulted.

2.1.2 The FORSA telephone survey

From April 28 until June 3, 1994, 500 computer assisted telephone interviews (CATI) were carried out in each country of the EU which were administered centrally from the FORSA offices in Berlin and Dortmund. CATI guides the interviewer step by step through the questionnaire and stores keyed-in responses such that they are ready for immediate computer analysis.

The sampling basis for the study were paper telephone directories. No random dialling technique was used. Each country was divided into a number of regions, and the telephone numbers in each region were counted. The number of telephone addresses to be drawn from each region/telephone directory was computed in proportion to the number of inhabitants in the region. In this way an effort was made to correct for the differences in telephone ownership in the different areas. The columns or pages in the telephone directories were selected in a systematic way with a frequency calculated by the above mentioned procedure.

Final telephone addresses of potential respondents were drawn in a random fashion. Addresses which were clearly identified to be corporate lines were replaced with another random telephone number on the same column/page. The entire random choice procedure was computerised. Obviously, not every telephone address chosen that way belongs to a private household. The non-private households are removed from the list and substituted.

Training and supervision of interviewers is obviously facilitated by a centralised telephone interviewing operation as the one reported on here. After an oral briefing on the questionnaire the FORSA interviewers had to conduct at least three test interviews. One supervisor per (approximately) ten interviewers was present through the entire interviewing period. The specific FORSA-CATI-System of computerised dialling ensured that call-backs were made at the agreed-upon time, that busy numbers were re-dialled after a predetermined delay and that 'no answer' or 'not at home' coded addresses were re-dialled after a longer delay.

Because of the difference in alphabet the interviews in Greece were not done by CATI but by paper and pencil. Therefore it was not possible to use the usual nonresponse analysis. The number of recalls was fixed on 12 while the respondents who refused co-operation were not contacted any more. Telephone numbers which did not lead to a contact have been substituted by a random chosen number. Given this procedure for the field work the completion rate presented in table 2.2 was obtained.

Table 2.2 Completion rate FORSA feasibility study

	B	DK	G-W	G-E	ESP	F	IRL
Gross sample	1209	1001	1023	1013	1153	1254	959
Net sample	500	500	500	500	500	501	500
Completion rate (%)	41	50	49	49	43	40	52

	NIRL	I	Lux	NL	P	GB	EC without GR
Gross sample	372	1088	1388	1137	1124	1456	14177
Net sample	150	501	500	500	500	500	6152
Completion rate (%)	41	46	36	44	45	34	43.4

Details about the FORSA fieldwork can be found in a special report (FORSA, 1994). When tables 2.1 and 2.2 are compared, one can see that the completion rates of the INRA face to face study and the FORSA telephone study are at about the same level.

2.2.3 The telephone panel: A continuation of the EB41.0

The telephone panel study was based on the face to face interviews, such that only respondents who had participated in the face to face EB41.0 study could be selected for the telephone panel. The fieldwork was done from April 5 to April 30, 1994. It was centralised in Brussels, and all interviews in the three countries of Belgium, France and Spain were conducted from a central call-centre. The survey was run by the Company MARKETING UNIT - the Belgian INRA member Company- using the BELLVIEW CATI (Computer Assisted Telephone Interviewing) software.

The respondents were told that due to some technical problems their answers from the face to face interview had been lost and thus were kindly asked to answer some of the questions one more time. In the face to face interviewing phase an effort was made to secure the telephone address of respondents in Belgium, France and Spain. Not every interviewee had a telephone or was willing to reveal his or her telephone number, and therefore it was not possible to approach everyone of the 3124 Belgian, French and Spanish EB41.0 respondents for a re-interview. On the whole, 2352 first-wave respondents could be approached once more.

The Brussels agency conducting the interviews was also in charge of interviewer supervision and training. Call-backs were limited to a maximum of eight attempts. Four questions helped to screen the interviewees in order to make sure that identical respondents were re-interviewed: age, sex, occupation and subjective social class. Respondents who refused to co-operate were not contacted again.

Using this approach for the fieldwork, 884 respondents could be successfully re-interviewed. The following table gives the country-specific information on the completion rates.

Table 2.3 Completion rate EB41.Panel

EB41.Panel	B	ESP	F	Total
Gross sample EB41.0 = (respondents whose telephone number was known)	767	731	854	2352
Net sample (respondents who were sucessfully re-interviewed)	234	309	341	884
Completion rate (%)	31	42	40	37.6

It is clear from this table that the willingness to co-operate in the panel study was considerably less than in the original study. Also, one has to keep in mind that table 2.3 represents an additional selection stage since the face to face study already represents a selection from the population studied.

Details about this experiment can be found in a special fieldwork report (ZEUS, 1994). Table 2.4. summarises the different sampling methods used in the three studies. More detailed information from the national institutes co-operating in the INRA chain would have been desirable with regard to, for example, interviewer training and supervision, but this information was not available.

Although the overall completion rates are not so different, the table clearly points to differences in the procedures. Apparently given these differences, one can also expect considerable differences with respect to the background variables describing the populations in the various countries. Normally, such differences are corrected using unequal weights for the different respondents. Therefore, in the next section the weights specified in these studies will be discussed.

2.3 The weights of the data sets

The weighting variable corrects the national samples in such a way that the samples are brought as closely as possible in accordance with known distributions of the national populations with respect to socio-demographic characteristics. For each of the two studies such weights were estimated. Below the weights estimated by INRA for the face to face studies and by FORSA for the telephone studies are presented. The weights for the panel study need not be addressed because the weighting is less relevant.

2.3.1 Weights of the INRA face to face study

INRA generates the weights on the basis of target tables with the joint distributions of age by sex, and the distribution of the population with respect to region (Eurostat NUTS II), occupation, size of locality and size of household. The population data have been taken from Eurostat for all countries.Table 2.5 displays the means, minimum and maximum as well as the standard deviation of the weighting variable WSAMPLE used in the Eurobarometer EB41.0.

Table 2.4 Summary of the sampling methods

	EB41.0	**EB41.Panel**	**FORSA**
Type	face to face	telephone	telephone
Fieldwork	April 4th - May 6th	April 5th - April 30th	April 28th - June 3rd
Countries	all 12 EU member states	France, Belgium, Spain	all 12 EU member states
Completion rate	EU: 44,7%	37,6% (% of eligible households.)	EU:43,4%
Sample frame	- Census enumeration units (or otherwise)	- respondents of EB41.0 with a telephone who have given their number	Telephone directories
Selection method	- more than 100 sampling units per country are randomly chosen as start address - a random increment provides up to 10 addresses - one person/per household selected by next birthday or Kish method or an other	- all possible respondents are contacted - controlled by Age, Sex, Occupation and Subjective Social Class	- From 10 to 22 'provinces' per country samples are drawn according to the size of the province's population - one person/per household selected by next birthday method
Interviewers' testing and supervision	INRA's national associates are responsible	INRA central - computerised dialling	FORSA central - tests in advance - computerised dialling
Call backs	2 revisits	8 call-backs	12 call-backs
Refusals	no refusal reversion	no refusal reversion	no refusal reversion
Substitution	random walk	no substitution	by random number

Table 2.5 The weights estimated by INRA

WSAMPLE	**Mean**	**Minimum**	**Maximum**	**Std. Dev.**
France	.97	.43	2.06	.26
Belgium	.92	.35	2.62	.30
The Netherlands	1.00	.12	4.93	.53
West Germany	.97	.26	3.54	.47
Italy	.95	.12	3.06	.39
Luxemburg	1.00	.35	3.25	.47
Denmark	1.00	.35	2.42	.35
Ireland	1.00	.45	3.70	.33
Great Britain	.95	.24	2.49	.26
Northern Ireland	.98	.88	1.10	.06
Greece	1.00	.38	2.33	.21
Spain	1.00	.19	1.97	.32
Portugal	1.00	.41	2.85	.32
East Germany	.95	.36	2.14	.29

2.3.2 Weights of the FORSA telephone survey

FORSA has corrected for the differences of the sample size and population size in each region by providing different inclusion probabilities to respondents of different regions. Furthermore FORSA constructs weights to approximate the joint distribution of age by sex. Table 2.6 displays means, minimum and maximum as well as standard deviation of the weighting variable WFSAMPLE used in the FORSA survey:

Table 2.6 The weights estimated by FORSA

WFSAMPLE	Mean	Minimum	Maximum	Std. Dev.
France	1.00	.68	2.35	.35
Belgium	1.00	1.00	1.00	.00
The Netherlands	1.00	.62	4.10	.46
West Germany	1.00	.71	1.49	.19
Italy	1.00	.47	1.81	.36
Luxemburg	1.00	.70	2.84	.34
Denmark	1.00	.72	2.34	.33
Ireland	1.00	.73	1.78	.27
Great Britain	1.00	.70	2.21	.33
Northern Ireland	1.01	.70	2.21	.34
Greece	1.00	.64	2.39	.38
Spain	1.00	.52	2.99	.40
Portugal	1.00	.63	2.28	.37
East Germany	1.00	.57	3.37	.35

The difference between these two sets of weights will be due to the different target table and the differences between the samples. Whereas the FORSA weights approximate a simple age-by-sex distribution, INRA specifies a more complex target table that incorporates additional variables. Besides that INRA uses as population statistics Eurostat data while FORSA has used the statistical information from the statistical offices of the different countries.

Due to this difference in approach different characteristics of the national samples can be expected. In order to make the samples comparable, one should use the same population figures and the same variables. This will be done in the next chapter where also the results for the substantive variables after weighing will be compared.

2.4 Comparison of item nonresponse rates

In this last section the face to face study and the telephone study will be compared with respect to item nonresponse because it presents another indication of the differences in the field work of the different organisations. A number of often used closed-ended questions have been selected to investigate possible differential nonresponse rates. In table 2.7, the combination of 'don't know', 'no answer' and 'refused to answer' codes between the face to face (INRA) study and the telephone (FORSA) study are compared.

With respect to item nonresponse no remarkable differences could be observed between the two studies on European level. There are questions that produce more refusals than others, but those differences appear both in the INRA face to face and the FORSA telephone survey.

Table 2.7 The percentage of 'don't know', 'no answer' and 'refused to answer' codes in the face to face EB41.0 and telephone (FORSA) study

Variable	EB41.0	FORSA
Satisfaction with life	0.6	0.5
Satisfaction with democracy	3.9	4.8
Frequency of political discussion	0.7	0.5
Persuade friends to share opinion	1.7	1.5
Watching news on television	0.2	0.0
Reading news in daily papers	0.3	0.1
Listen news on the radio	0.3	0.1
Interest in European politics	1.1	0.9
Informed about European politics	1.9	2.3
Is membership in EU good-bad thing	5.0	6.8
Has country benefited from EC membership	17.0	18.6
Will R vote in EP election	6.0	3.9

National differences were also studied but are not reported here. This detailed analysis shows that there is national variation in certain questions: some nations display more refusals with regard to these questions than others. These national differences, however, also appear in both surveys. As a result, it can be stated that there are no systematic differences between the institutes (FORSA vs. INRA) with respect to item nonresponse.

2.5 Conclusion

It was the purpose of this chapter to discuss the differences in sampling design, fieldwork and to look at the consequences for the total non response, the necessary weights and the item non response.

The findings indicate that:
- the total nonresponse looked very similar,
- the item nonresponse was also quite similar, and
- the weights differed substantially.

The first two findings do not imply that the samples are equivalent; it only indicates that the co-operation of the sampled respondents was similar. Due to the differences in the sample designs the originally drawn samples were already quite different. For example we expect that the ownership of a telephone could make a difference between households. This problem will be elaborated in chapter 4. Due to differences in the fieldwork, especially the number of recalls and the substitution of non co-operating households, the final samples certainly were even more different. For these reasons the necessary weights were also quite different. These points will be further evaluated in the next chapters.

CHAPTER 3

DEVIATIONS FROM THE POPULATION AND OPTIMAL WEIGHTS

SABINE HÄDER AND SIEGFRIED GABLER

3.1 Introduction

Nonresponse has become an important problem in the empirical social sciences.[3] This problem plays a role in face to face studies as well as in telephone and mail surveys. Presently, the percentage of nonresponse frequently ranges from 30-50%.[4] These losses due to nonresponse may lead to systematic biases in the samples which result in biased estimates. The application of weighting procedures is a usual way to compensate for this bias.

The weighting process changes the relative importance of the respondents. Technically speaking, weighting means the attachment of numbers to elements of the responding population. As a consequence, after weighting the sample profile of a variable of interest should be more similar to the population profile than without weighting. But for this logic to apply, it should be taken for granted that within each weighting class, the profiles of the survey variables are very similar for the responding and nonresponding parts of the sample (Elliot, 1991: 5).[5] In this way it is possible "...to make the sample data we collect more representative of some population data we are trying to measure or estimate" (INRA, 1994: 26).

In general, the adaptation cannot be achieved for all variables at the same time. Therefore, some variables for which the population profiles (e.g. gender, age, household size) are known are usually selected to act as so-called active variables. Testing the effect of weighting is possible only for a few additional variables not used for weighting for which we also know the distributions from official statistics, like income, marital or professional status. But the key assumption for successful weighting is that the bias of the other (passive) variables like attitudes or behaviours is reduced by the weighting procedure as well. However, resulting

3 For a classification of nonresponse see Kish, 1965: 532-534.

4 Some examples to illustrate this statement: The ALLBUS 1994 - a face to face study in Germany - had a nonresponse rate of 46 percent (Koch et al., 1994: 82). In the National Readership Survey (UK) a nonresponse rate of 39 percent was stated. Also, in American surveys the nonresponse rates reach about 40 percent nowadays (Bradburn, 1992: 392). The average rate only of refusals and hangups of the FORSA European Telephone Survey 1994 was about 40 percent (FORSA, 1994: 7-13) of the Net Sampe Pool.

5 Of course this condition cannot be tested from the survey data. So we have to assume that the nonrespondents are a random subsample in each weighting class.

changes in passive variables due to weighting cannot be controlled by comparisons with the distributions of the total population because we do not have population profiles from sources other than from empirical surveys which may be biased themselves.

Because of these difficulties the weighting procedure is discussed controversially by the empirical social scientists (for a description of different points of view in the matter see Gabler et al., 1994).

By showing the impact of weighting factors in the concrete framework of the Eurobarometer Experiment 1994, we want to contribute to this discussion. First of all, it will be tested how the samples represent the population in a comparison of different national studies. Therefore, we will show the results of a comparison of the profiles of some selected variables from the national samples and the populations. Following this, we will describe the statistical characteristics of the weights we have used. Finally, we will pursue the question whether there are differences in the distributions of the passive variables before and after weighting.

As a result of our analysis it should therefore be possible to show typical patterns of the impact of weighting in the framework of these European studies.

Table 3.1 Sample sizes of selected studies of the Eurobarometer Experiment

Nation	EB41.0	EB41.Panel	FORSA
Belgium	1087	234	501
France	1034	341	500
Spain	1003	731	500
East Germany	1058	-	500
West Germany	1064	-	500

For reasons of parsimony and clarity, we shall select only a few countries for our investigation: Belgium, France and Spain have been chosen because they are also used in the panel study.[6] We have included West Germany and East Germany in our investigation because of the large differences between these two parts of Germany concerning the availability of telephones in private households. In 1994, more than 9 out of 10 households in West Germany had a telephone, while the level of telephone availability in private households had reached only about 50% in East Germany till then (Drews, 1994; Häder, 1994). This low level of telephone penetration in East Germany contains a great risk of a major sampling error because not all of the households of the target population have a known positive chance of being included in the sample. Besides the nonresponse error, this fact can also negatively affect the accuracy of the estimates. Therefore, the differences in the representation of the target population between the face to face and the telephone survey in East Germany should be particularly observed.

6 We do not analyse the samples of the panel studies because the number of realized valid cases is too small for our investigation.

3.2 Comparison of the distributions of selected active variables of the national samples with statistical data on the target populations

A comparison of the distributions of selected demographic variables (age, gender, household size) from the different samples with the reference data based on the national statistical yearbooks shows how large the differences in all variables and nations are (see Table 3.2, 3.3 and 3.4).

Table 3.2 Percentages of males in different nations and studies as well as in the target populations

Nation	EB41.0	FORSA	Reference data
Belgium	49.0	48.0	48.3
France	48.6	39.9	48.1
Spain	48.8	37.8	48.4
East Germany	49.4	44.2	47.4
West Germany	51.0	42.6	48.1

Table 3.3 χ^2 - statistics for the distributions of age groups[7] in EB41.0, in FORSA and reference data based on the statistical yearbooks

Nation	χ^2- Statistics for Eurobarometer-Reference	χ^2- Statistics for Telephone-Reference
Belgium	1.27	4.39
France	9.61	1.34
Spain	1.49	19.63
East Germany	10.90	19.57
West Germany	11.11	7.08

Table 3.4 χ^2 - statistics for the distributions of household sizes[8] in EB41.0, in FORSA and reference data based on the statistical yearbooks

Nation	χ^2 - Statistics for Eurobarometer-Reference	χ^2 -Statistics for Telephone-Reference
Belgium	56.35	41.95
France	57.05	38.33
Spain	40.25	29.13
East Germany	51.82	51.87
West Germany	45.63	70.68

7 For the age groups we chose the following categories: 1.: 15-29 years; 2.: 30-44 years; 3.: 45-59 years; 4.: 60 years and older.

8 For the household size we chose the following categories: 1.: one person; 2.: two persons; 3.: three persons; 4.: four and more persons.

In the two Tables 3.3 and 3.4 we show the χ^2 - statistics for the test on similarity of the expected and observed marginal distributions for the variables age and household size. The critical value for an α-level of 0.05 and 3 degrees of freedom is 7.81.

For all countries - in part large (household size) - differences between the sample distributions and those of the target populations (adults of 15 years and older) have to be stated as the computed statistics show. Altogether in 15 out of 20 tested cases the differences between the distributions are significant. This means that all samples are biased demographically. Therefore, biases in other variables due to nonresponse or (other) sampling errors can be expected as well. The usual way in social and market research to deal with this problem is - as stated above -

1. to assume a fairly high correlation between demographic and all other survey variables, and
2. therefore, to adjust some of the demographic variables to their known distributions of the target population.

In the following we want to pursue this approach and - after that - discuss the results we have obtained with this method.

3.3 Effects of weighting

Before presenting our special weighting procedure we will offer some remarks on weighting in general. We distinguish two kinds of weighting:

a) *Design weights* (also called preweights): Sometimes the sampling design requires weighting of the observations to avoid bias. If the probabilities of selecting the units are not equal, we should weight the units with the inverse of the inclusion probabilities. This leads to the well-known Horvitz-Thompson estimator which is unbiased for the population total (or mean). This way of weighting is useful provided that the sample design has been realized precisely and the inclusion probabilities are exactly known.

However, analyses of the fieldwork of face to face studies as well as telephone surveys have shown that in reality this is usually not the case. Also, in the studies of the Eurobarometer Experiment considerable rates of refusals existed. For example, for the telephone surveys among the five countries discussed in this contribution refusal rates range from 19.6% of the Net Sample Pool in East Germany to 49.9% in Belgium (FORSA, 1994: 7, 13).

In several investigations it was found that there is no guarantee for improving the estimator by using design weights.[9] Anyway, as a practical problem for our investigation of the Eurobarometer Experiment we have to state that the design weights to correct for the unequal chance for households of different sizes to be selected, cannot be determined because the necessary information for the calculation is not available from the FORSA telephone surveys. As a consequence, we will do our analyses without this mode of weighting.

9 Rothe stated as a result of his investigations, „daß bei personenbezogenen Variablen die Hochrechnung ohne Verwendung einer Gewichtung eher bessere Werte liefert als die theoretisch korrekte Gewichtung mit der reduzierten Haushaltsgröße“. (see Rothe, 1994: 71)

b) *Cell weights* (also called postweights): After collecting the data, the sample profiles of some characteristics, i.e. gender or age, may differ from the corresponding population profiles which are known from external sources. Poststratification is the usual way to include this information into the estimation. If the multivariate distribution for these characteristics is unknown but some marginal distributions are known, we can try to find weights which yield representative estimators in the sense of (Hájek, 1981). This means that after weighting the sample profiles and the population profiles are identical for each of these characteristics. The solution of this problem is not unique. Usually the number of solutions is infinite. A requirement with regard to the weights should be that they are as close as possible to one. One popular solution is the **I**terative **P**roportional **F**itting solution also known as raking. The algorithm has its origin in the paper of Deming and Stephan (1940).

The weights can be found by minimizing the objective function (similar to the discrimination information or Kullback-Leibler-information) which represents the distance between the weights w_i and 1, the unweighted case,

$$\sum m_i (w_i \ln(w_i) - w_i + 1)$$

where m_i is the number of elements in cell i and w_i is the weight which we attach to all units of cell i. The constraints under which the objective function has to be minimized are that some marginal distributions of the weighted units are representative. These constraints can be represented by a matrix equation of the form $A\underline{m}_w = \underline{n}$ with known restriction matrix A , the vector $\underline{m}_w=(m_1w_1, m_2w_2, ...)'$ and the vector $\underline{n}=(n_1, n_2, ...)'$ of marginal cell frequencies.

Raking estimates are not maximum likelihood estimates of the cell proportions when the observed data are a random sample from the target population, but they are consistent and best asymptotically normal as Ireland and Kullback (1968) show.

If we do not choose the logarithmic function as an objective function which we want to minimize but the quadratic function (**L**east **S**quares, minimum variance criterion)

$$\sum m_i (w_i - 1)^2$$

we get weights with the highest efficiency[10]. The solution is given by

$$\underline{m}_{WL} = A(AA')^{-1}\underline{n}$$

which is the Moore-Penrose-Inverse of $\underline{n}$. It may be that the solution in the quadratic case is no longer nonnegative. Usually, weights less than a positive minimal number g_0 or higher than g_1 are truncated to g_0 or g_1, respectively.[11]

3.4 Restrictions for the construction of the weighting factors in the framework of the Eurobarometer Experiment

According to our research interest we wanted to construct weights in such a way that in all countries, for both the face to face and the telephone survey, an adjustment to the same variables is performed. By doing this we would be able to compare the statistical characteristics of the weights for the two samples in each country. Furthermore, it would be useful to analyse the impact of the weighting procedures on the profiles of selected passive variables.

First of all we had to solve some practical problems concerning the active variables:

- The telephone survey done by FORSA contains only 500 cases per nation. For reasons of consistency empty cells should be avoided in a weighting procedure. Thus we had to decide whether we wanted to use fewer, but more finely subdivided active variables or broader categories for more variables.
- In the questionnaire of the telephone survey the demographic questions are only partly replicated from the Eurobarometer study. For example, questions on the marital status of the respondents and on the number of children living in the household are missing. Besides, in some other eligible variables different categories have been used in the questions of the two surveys we want to compare. These facts have reduced the number of possible active variables for our analysis.
- The basis for reference statistics were the annual Statistical Yearbooks of the different countries and the German Statistical Yearbook 1994 for Foreign Countries (Statistisches Bundesamt, 1994), respectively. The tables in the different National Yearbooks are not standardized but vary from country to country. This led to problems in finding comparable distributions of eligible active variables. Therefore, we decided to take the reference data for the joint distributions of age and gender out of the statistical yearbooks of the different countries. As data source for the marginal distribution of the household sizes we used the Statistical Yearbook 1994 for Foreign Countries where the presentation was the same for all five countries.

---

10 For the definition of efficiency see chapter 3.6.2.

11 An overview of the construction of weights can be found in Alexander (1987), Deville et al. (1992, 1993), Gabler (1994), Zaslavsky (1988). A comparison of the various solutions is given by Little and Wu (1991).

3.5 Realization of the weighting procedure

Considering the problems mentioned above, we selected the following active variables to be adapted in the weighting procedure:

Age (4 groups) * Gender (2 groups) * Household size (4 groups) = 32 cells for France, Belgium, Spain, East Germany and West Germany (both samples).

As we have shown in the Tables 3.2, 3.3 and 3.4, the distributions of these variables in the national samples (face to face study and telephone survey) differ considerably from those of the reference data. Therefore, an adjustment of these selected variables seemed to be useful.

Moreover, the selection of age, gender and household size as active variables is frequently done in social as well as in market research[12]. Sometimes weighting by these factors is interpreted as a cure for biased samples. Therefore, there is a need to observe the impact of weighting on the passive variables carefully. Besides we wanted to compensate for the missing of design weights by including the household size into the group of active variables.

Because we had no joint distribution for the variables age, gender and household size for each nation, which would be the condition for simple cell weighting, we used the two different ways as already referred to for determining the weighting factors. We therefore obtain two solutions, the IPF-Solution and the LS-Solution.

3.6 Results of the analysis of the weighting factors

3.6.1 Ranges of the weighting factors

In the following we want to describe the factors we have obtained as a result of the two routines to be used for the weighting procedures. In Table 3.5 and Table 3.6 we present the ranges of the weighting factors for each country.

Table 3.5 Ranges of the cell weights for EB41.0 and for FORSA for the selected nations (IPF-Solution)

Nation	EB41.0	FORSA
Belgium	0.580 - 1.919	0.472 - 2.276
France	0.637 - 1.901	0.381 - 2.159
Spain	0.696 - 2.166	0.939 - 3.006
East Germany	0.579 - 1.901	0.268 - 5.492
West Germany	0.770 - 1.618	0.475 - 2.901

12 These active variables are for instance an important part of the weighting procedure for the „Media-Analyse“ which is well known in market research as a highly reliable study (see Rothe and Wiedenbeck, 1994).

Table 3.6 Ranges of the cell weights for EB41.0 and for FORSA for the selected nations (LS-Solution)

Nation	EB41.0	FORSA
Belgium	0.466 - 1.767	0.253 - 2.018
France	0.599 - 1.741	0.085 - 1.873
Spain	0.639 - 1.995	0.289 - 2.455
East Germany	0.506 - 1.762	- 0.453 - 3.390
West Germany	0.758 - 1.579	0.300 - 2.409

For both solutions, in all nations the ranges of the weighting factors for the telephone survey are larger than the ranges of the factors for the Eurobarometer study. In particular, it should be noted that some negative weighting factors for the telephone survey in East Germany resulted from the LS-Solution.

As previously mentioned, it is a requirement that the weights should be as close to one as possible. The Eurobarometer study is closer to the achievement of this goal in all nations. Furthermore, the sizes of the weights for the telephone study are mostly larger. So we have a first indication concerning the comparison of the quality of the two samples.

3.6.2 The efficiency of weighting procedures

"A useful measure of the effect of unequal probability sampling on precision is provided by the 'effective sample size' or ESS. It measures the size of an equal probability sample that would produce the same precision as the unequal probability design under consideration." (Elliot, 1991: 8)[13].

The efficiency (i.e. ESS as a proportion of the actual sample size) shows what proportion of the original sample size an unweighted random sample with the same variance as the weighted random sample would have. Efficiency also can be used for the comparison of the quality of two samples. In that case different efficiencies mean that the samples represent the population with different precision.[14]

In this sense we want to interpret the efficiencies in Table 3.7 and 3.8 for the Eurobarometer study and the telephone survey for our selected nations and both weighting solutions as control criteria for the quality of the different samples.

13 "The main assumption under which the formula is derived is that the true population variances are equal in the groups having different weights, although it also assumes independent simple random sampling in the different groups. In many situations theses assumptions may be reasonable enough." (Elliot, 1991: 8)

14 "Man kann das Effektivitätsmaß natürlich auch dazu verwenden, die Abbildungsgüte zweier Stichproben miteinander zu vergleichen. In diesem Fall zeigt ggf. unterschiedliche Effektivität, daß die ungewichteten Stichproben die Grundgesamtheit unterschiedlich genau abbilden." (Von der Heyde, 1994: 150)

Table 3.7 Efficiencies of the IPF-Solution (in %)

Nation	EB41.0	FORSA
Belgium	92.8	88.4
France	92.6	87.5
Spain	93.9	84.4
East Germany	92.5	69.8
West Germany	94.2	81.9

Table 3.8 Efficiencies of the LS-Solution (in %)

Nation	EB41.0	FORSA
Belgium	92.9	88.7
France	92.7	87.9
Spain	94.0	85.1
East Germany	92.6	73.2
West Germany	94.2	82.4

Table 3.7 and Table 3.8 show higher efficiencies for the Eurobarometer study in all nations with both weighting procedures. We can interpret this result as a sign that the representation of the target population is better achieved in the Eurobarometer study than in the telephone survey. The most similar efficiencies are shown by the two Belgian studies. The largest differences in efficiency between the Eurobarometer study and the telephone survey exist in East Germany. Of course, the weights generated as a result of the LS procedure cause higher efficiencies than the IPF weights for all countries, except for the Eurobarometer study in West Germany were the efficiencies are the same for both solutions.

3.7 Results of the analysis of the impact of the weighting procedures on selected passive variables

In the following section we want to describe the similarities and dissimilarities of selected weighted and unweighted passive variables. For this analysis we used the following indicators (see Chapter 1, for the wording of the questions):

* Satisfaction with life
* Satisfaction with democracy
* Subjective social class identification

These indicators are frequently used in the social sciences.

It is obvious that weighting procedures do not have an impact on the distribution of a passive variable if its relative frequencies are similar for each cell.

For the Eurobarometer study in West Germany Figure 3.1 shows the histograms for the passive variable "Satisfaction with life" with respect to the 32 cells defined by the active variables. The cells are lexicographically ordered with respect to gender, age and size of household. For example, the second picture from the left in the second row is based on the 76

male respondents in West Germany of age 45-59 living in households of size 2. Since most of the histograms are of similar shape we are not surprised that the weighting procedures do not essentially change the distribution of the passive variable. One can also show that the dependency and thus the correlation between our active variables and "Satisfaction with life" is very weak and that therefore the correction is minimal.

It can also be clearly seen in Table 3.9 that weighting has almost no effect on the marginal distributions of the passive variables. To save space we will present only the case of "Satisfaction with life". For the other passive variables mentioned above we obtained similar results.

The profiles are approximately the same before and after weighting but in most cases very different for the Eurobarometer study and the telephone survey.

In order to summarise the similarities and dissimilarities between the weighted and the unweighted data of the two surveys in each nation, we discuss the results of a correspondence analysis shown in Figure 3.2.

The correspondence analysis is a multivariate method for the graphical representation of the rows and columns of a contingency table. The four categories of "Satisfaction with life" serve as rows of the input data matrix. The columns consist of the relative frequencies of the four row categories in the case of the weighted (IPF and LS) and unweighted studies in the different nations.

The circles in the map represent the rows, the squares and triangles the columns. The first letter denotes the nation[15]. P and T, respectively, are abbreviations of Personal Eurobarometer study and Telephone survey, respectively. The squares are used as symbol of the Eurobarometer data, the triangles represent the columns of the telephone survey. The last two letters denote the weighting procedure[16]. The nearer two profiles are the nearer are the corresponding points in the map. Since the quality of the two-dimensional plot is 96.4% we have an excellent representation of the data in the plot. The map shows very clearly that the three points belonging to the same nation and the same survey type are close together. That means the PU and PW as well as the TU and TW of one nation lie side by side, partly one on top of the other. This is an indication that the weighting procedures do not have an impact on the results of the indicator "Satisfaction with life". The reason for it is the similar distribution of this passive variable according to the cells defined by the active variables.

Our map also shows a clear distinction between the Eurobarometer study and the telephone survey. The first axis in the map can be interpreted as the "satisfaction axis". The more to the right the projections of the points onto the first axis are, the more satisfied the respondents on average were. Since the telephone survey points lie always to the right of the corresponding Eurobarometer points, the FORSA respondents in contrast to the Eurobarometer respondents are more satisfied. The difference in Spain is especially striking. In general, this may be an influence of the survey method. To investigate this question further we added two supplementary columns which separate the telephone owners from the respondents without

15 B=Belgium, F=France, S=Spain, O=East Germany, W=West Germany
16 U=unweighted, W=weighted, I=IPF, L=LS

telephone in East Germany. We choose East Germany as an example because of its relative low telephone penetration. The squares labelled by O_PU_TO and O_PU_NTO, respectively, are an indication to the fact that at least in East Germany the answers depend on the telephone ownership. The point O_PU_TO is closer to O_TU. If we proceed in the same way for Spain we obtain a similar effect, although it is not so drastic. The conclusion is that telephone ownership cannot explain the immense difference between the Eurobarometer points and the FORSA points for Spain.

Figure 3.1 Histograms for "Satisfaction with Life" in 32 cells

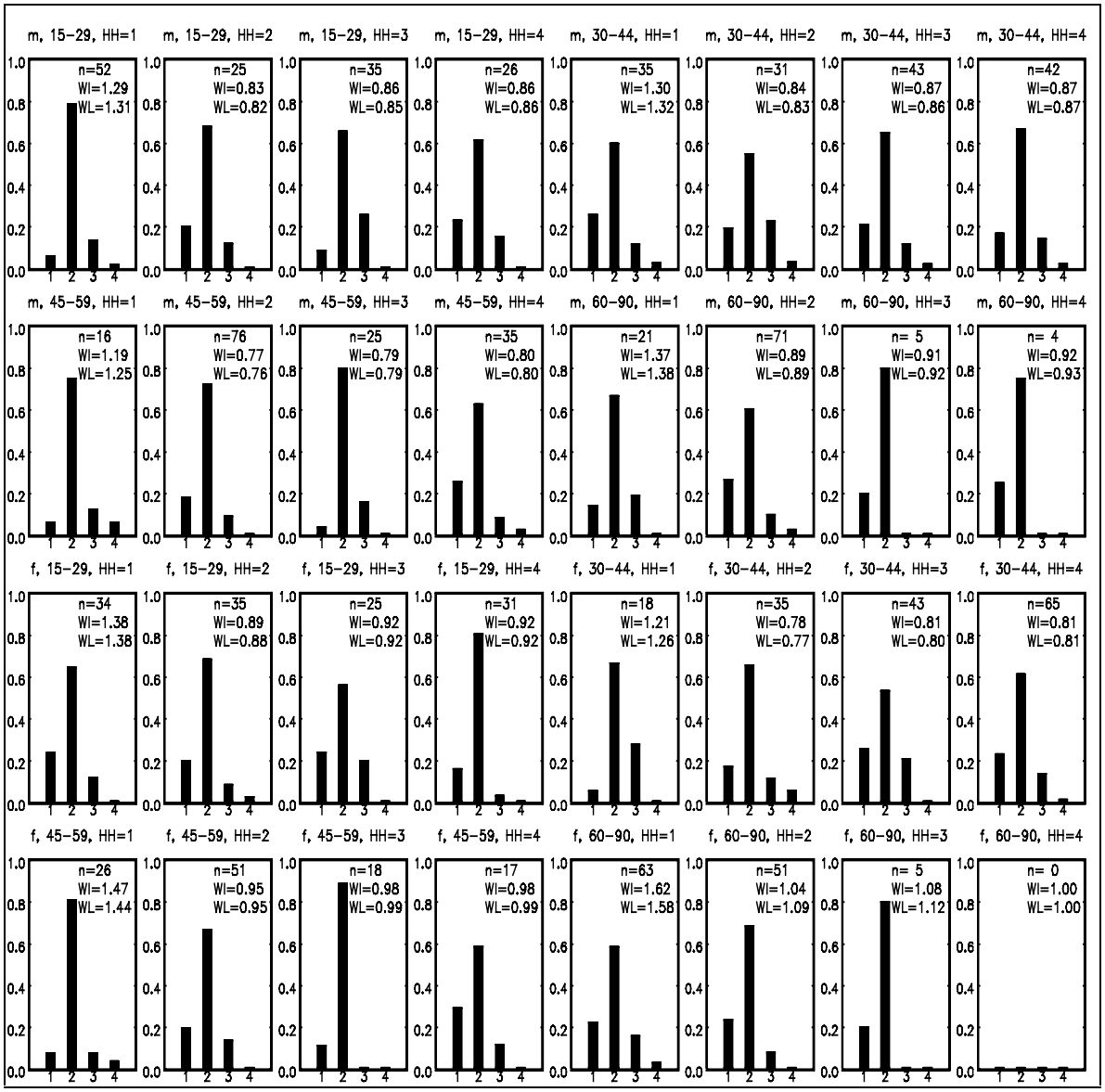


Table 3.9 Satisfaction with life: Marginal distributions for the unweighted and weighted data in five European countries (P=EB41.0, T=FORSA, U=Unweighted, W=Weighted, I=IPF-solution, L=LS-solution)

Satisfaction in:						
Spain	**PU**	**PWI**	**PWL**	**TU**	**TWI**	**TWL**
very satisfied	13.9	13.6	13.6	35.8	36.0	35.9
fairly satisfied	53.8	54.0	54.0	42.8	43.1	43.1
not very satisfied	25.2	25.3	25.3	18.0	17.6	17.7
not at all satisfied	7.1	7.1	7.1	3.4	3.3	3.3
Belgium	**PU**	**PWI**	**PWL**	**TU**	**TWI**	**TWL**
very satisfied	32.5	31.4	31.4	32.7	32.2	32.2
fairly satisfied	56.0	56.3	56.4	60.2	60.0	60.0
not very satisfied	9.0	9.7	9.6	5.3	5.8	5.9
not at all satisfied	2.5	2.6	2.6	1.8	2.0	1.9
France	**PU**	**PWI**	**PWL**	**TU**	**TWI**	**TWL**
very satisfied	13.7	13.3	13.3	21.2	21.9	21.5
fairly satisfied	58.5	58.2	58.3	61.7	61.2	61.5
not very satisfied	21.3	21.5	21.4	11.7	10.7	10.8
not at all satisfied	6.5	7.0	7.0	5.4	6.2	6.2
East Germany	**PU**	**PWI**	**PWL**	**TU**	**TWI**	**TWL**
very satisfied	7.3	7.7	7.7	13.9	13.9	14.0
fairly satisfied	64.5	63.7	63.7	64.1	63.3	64.0
not very satisfied	21.7	21.8	21.8	18.2	18.0	17.2
not at all satisfied	6.5	6.8	6.8	3.8	4.8	4.8
West Germany	**PU**	**PWI**	**PWL**	**TU**	**TWI**	**TWL**
very satisfied	19.1	18.8	18.8	26.7	27.4	27.5
fairly satisfied	66.5	66.6	66.6	61.5	60.1	59.7
not very satisfied	12.9	13.0	13.0	9.4	9.6	9.9
not at all satisfied	1.5	1.6	1.6	2.4	2.9	2.9

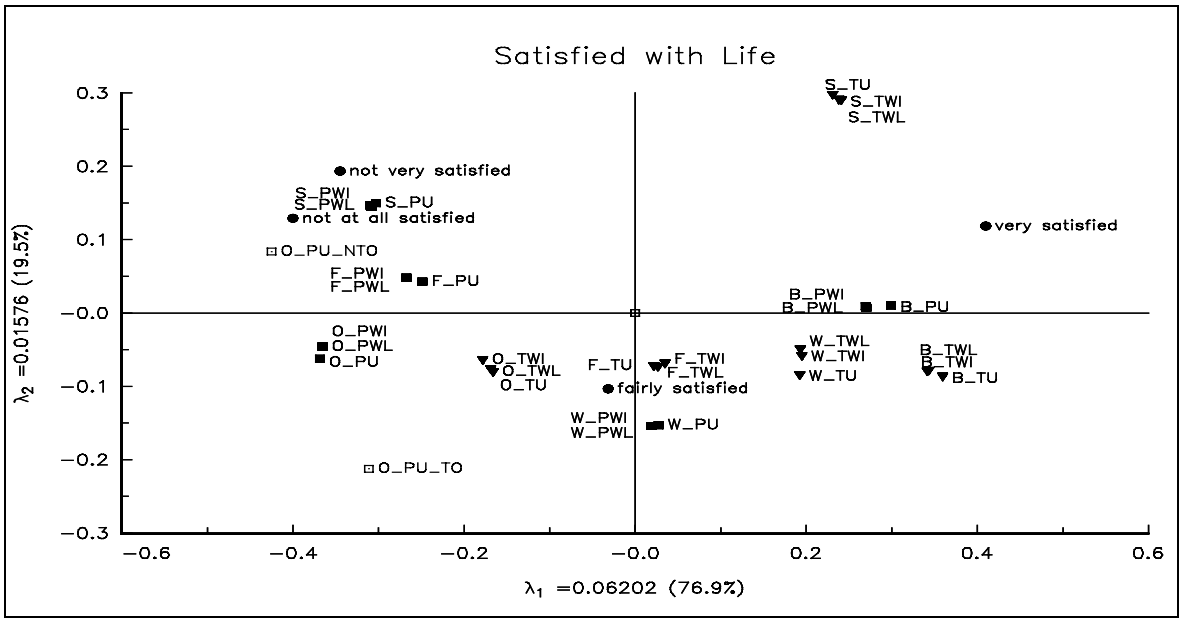
3.8 Conclusion

The results of our analysis lead to the conclusion that both the face to face as well as the telephone sample deviate quite a bit from the different populations with respect to the reference variables. We also found that the weighted and the unweighted passive variables have nearly the same distributions. This means that existing differences between the Eurobarometer study and the telephone survey are not reduced by weighting. This result is valid for all nations, all analysed variables and both weighting routines.

Furthermore, our analysis is in agreement with the findings of Schnell (1993), who looked at the impact of weighting by the active variables gender, age and region on attitude variables of the ALLBUS study. As a summary he stated that it is not possible to derive the unbiasedness

of the passive variables from the unbiasedness of demographic variables in a sample.[17] The reason he gives for this quite consequential conclusion is the only indirect and low impact which the active variables have on the passive variables.

Figure 3.2 Correspondence analytic map of "Satisfaction with Life" for the two modes in the five selected countries



Our analysis also results in the message that cell weighting with the selected demographic variables does not adjust the marginal distributions of the passive variables in both surveys. This means that for our investigation nonresponse bias or sample design bias could not be corrected by the weighting procedures applied.

The crucial problem for the application of weighting procedures seems to be the selection of the active variables. If there exist no relations between the active variables and the survey variables of interest, respectively, weighting does not help. Therefore, a mechanic application of weighting routines without checking these relations does not make much sense.

Furthermore, the results of our analysis show that the Eurobarometer study seems to represent the target population better than the FORSA telephone study does. This could be clearly demonstrated for some demographic variables. There are several indications that this is valid for the selected passive variables, too. However, we cannot generalise this conclusion for all telephone studies since the quality of a survey depends on many factors of which the mode of data collection is only one. Further investigations are necessary to throw light on this problem.

17 Schnell stated "daß aus der Unverzerrtheit 'demographischer Variablen' in den Stichproben nichts über die Unverzerrtheit anderer Variablen folgt" (Schnell, 1993: 29).

CHAPTER 4

TELEPHONE OWNERSHIP A CAUSE OF SAMPLING BIAS IN EUROPE ?

JÜRGEN LASS

4.1 Introduction

The coverage error used to be a central argument in the past against telephone surveys for getting information about an entire population. The former label of a "quick, but dirty approach" points (Kreiselmaier and Porst, 1989:7) in that direction. In an ideal case a survey sample is just a homomorphically smaller copy of the population. That means there is no substantial difference between the target population to be studied and the "frame population", to which the sampling scheme is applied. In reality there is a discrepancy between the calculations on the target population and the calculations on the frame population, which is called coverage error. Groves (1989:83) pointed out: "Coverage error arises from the failure to give some units in the target population any chance of being included in the survey...". For drawing methodologically legitimate conclusions about the target population which are only footed on the smaller copy of a survey one of two essential conditions should be met in telephone surveys. First, all national private households (in the following no distinction between a sample of households and a sample of persons) do have telephones. In this case every household has an equal chance to enter the sample. Or second, if in a country the telephone coverage is not 100%, the characteristics of households with and without a telephone do not differ. This applies not only to qualities on which the researcher is focused but to qualities in a broader sense which may indirectly be related to the qualities under research.

Which condition is more important for practical work? Concerning the single countries of the EU the first quantitative condition is not fulfilled. Some data on this problem are shown in detail below. It can be argued that a full telephone coverage is not necessary because alternative instruments like face to face-surveys do not work better although this instrument can reach all household theoretically. Based on well known face to face surveys in international social science programs the rate of completed interviews ranges from 62% to 77% (Porst, 1993)[18]. Interviews via mail get usually much lower rates of completed interviews (Sosdian and Sharp, 1980) unless the efforts are increased by using the Total

18 Completed interviews in ISSP on average: West Germany 62.3, Great Britain 66.9, The Netherlands 76.5. Completed interviews in General Social Surveys in the United States on average 77%.

Design Method (Dillman, 1978). Since the traditional survey instruments fail to cover more than 20% of the intended target group it could be argued that a telephone density of more than 80% or 90% would be sufficient for practical reasons. The main problem would not be full coverage but nonresponse.

The assumption the higher the telephone density the less important becomes the problem of coverage could be theoretically challenged. In general, the absence of the first quantitative condition would not be particularly problematic, if the second qualitative condition is met. It is theoretically possible that a relatively low telephone density in a country goes together with a structural equality or similarity between owners and nonowners of a telephone. In this case conclusions drawn from a telephone survey can be generalised across the entire population. On the other hand, countries with high telephone supplies might theoretically show the following feature: the smaller the group of nonowners the more different this group will be in its demographic and attitudinal composition.

This phenomenon could be described as a process of concentration of nonownership in the course of ongoing technological modernisation. In the first phase people are more or less equal, only pioneers own a telephone. In the final phase all are more or less equal too, but only a small group of „dropouts“ does not own a telephone. This could result in a biased sample which could in this case not be the basis for drawing correct conclusions about the entire population.

Thus the qualitative question of differences is much more important than the quantitative, country-dependent question of telephone density and the question whether there is a qualitative problem can only be answered empirically.

Some results about the composition of nonowners have been published. These results are not surprising for experts in survey research, but they should shortly be repeated to mention the reasons for the scepticism towards telephone surveys. In the USA, where more than 93% of the 75 million households had a telephone already at the end of the seventies, households without a phone tended to be lower-income households, including retired persons, minority families, and single parent families (Backstrom and Hursh-Cesar, 1981:114). Groves and Kahn (1979) systematically compared personal interview surveying and telephone surveying in the seventies. According to their results, single adult households, less educated and poorer people, minorities, and non-professional and nonmanagerial workers are more likely to have no telephone. West Germany also achieved a high telephone coverage at the end of the eighties (more than 90%, Euler, 1989:314). Jung (1990) found in his West German study 1987/88 in which he compared telephone surveys with face to face-interviews that better educated persons are clearly overrepresented in telephone surveys. Frey et al. (1990:15) did not conduct an empirical study themselves but made some suppositions. They ask the question: who does not have a telephone in West Germany? They proposed the following hypothesis:

- socially weak persons (lower social strata, workers, people with low incomes, people living in the rural areas, homeless people),

- socially „insecure“ people (alone living, older than 70 years), who are not necessarily socially weak persons,

- foreigners,

- people “sine nobiles“, who try to protect their private homes.

Nonownership by „snobbism“ may be characterised as an indicator of disintegration. In Germany and not only here a social norm is established that paying a visit without prior announcement by phone is impolite. Therefore nonownership may be an indicator of nonparticipation in social life or of disintegration (Lange and Zernick, 1990:103). If the last point is put aside the reported results can be labelled as a „stratum bias“ in telephone surveys. Like other higher value consumer goods telephones do not spread in the socioeconomic pyramid simultaneously, but over the time from the top to the bottom. Nonownership is to a great extent a function of the availability of material resources (income, purchasing power). This is where simple economic rules play a role: the price of a good has an impact on the demand for it. Costs of telephone charges differ throughout Europe. According to some (unfortunately) older data the cost of maintaining a telephone over a year was two times higher in Germany than in Denmark and three times higher than in Luxembourg. For getting a telephone the relationship of costs had been even more unfavourable in Germany compared to the other two countries (Gölz, 1983). As shown below, the ranking of these countries as regards telephone noncoverage is similar.

The „stratum bias“ could lead to serious problems. In social research people of lower strata are often important subgroups which may differ considerably in their attitudinal profile from the rest of the population. Special institutional forms of integration into society, cultural reasons or feelings of deprivation, dissatisfaction, and apathy have an impact on this special attitudinal profile. Although Groves and Kahn (1979) concluded in their study that telephone nonowners in the USA do not differ greatly from owners as regards many aspects of attitude, one should be cautious about drawing parallels with Europeans countries without having gathered empirical evidence (De Leeuw and van der Zouwen, 1988).

In practice a second aspect of the coverage problem could be become virulent. The use of telephone books as sample frame for surveys of population excludes not only the nonowners but also people with secret numbers. The proportion of secret numbers can differ heavily from country to country (Frey et al., 1990:67). The solution to this problem consists of using forms of random digit dialling. In most cases the whole telephone number is not created by chance but only the last digits. This approach is more efficient than the creation of complete random numbers because it leads to more telephone numbers which are in use. Thus, this aspect of a possible coverage error does not necessarily play a role.

This chapter deals empirically with the comparison of telephone and nontelephone households on the level of 14 countries of the EU (West and East Germany are separated). A wide set of variables is included which corresponds to the findings in the cited studies. The compositional structure of the group of telephone nonowners is examined for each country of the EU individually in order to answer the question, which parts of the population are overrepresented in telephone surveys. In a second step a closer look is taken on possible interactional effects between variables.

A crossnational design is a special feature of this study. Countries with a high telephone density can be compared with countries with a lower density. Questions like, “Is the ‘causal’

pattern everywhere the same?", "Or do the countries differ?", can be answered. This would be consequential for efforts to correct for under-coverage bias in each country, if necessary.

4.2 Method

The data base consists of conventional face to face-surveys of the Eurobarometer in the EU member states, excluding the newer members Austria, Sweden and Finland. In the Eurobarometers 38, 38.1, 39, 39.1, 40, and 41, carried out between Autumn 1992 and Spring 1994, information about households with and without a telephone was collected. According to the study reports these surveys are nationally representative. These datasets were cumulated into samples with more than 6,000 respondents for each country (except Luxembourg and Northern Ireland). There were more than 3,000 respondents in Luxembourg and more than 1,800 in Northern Ireland.

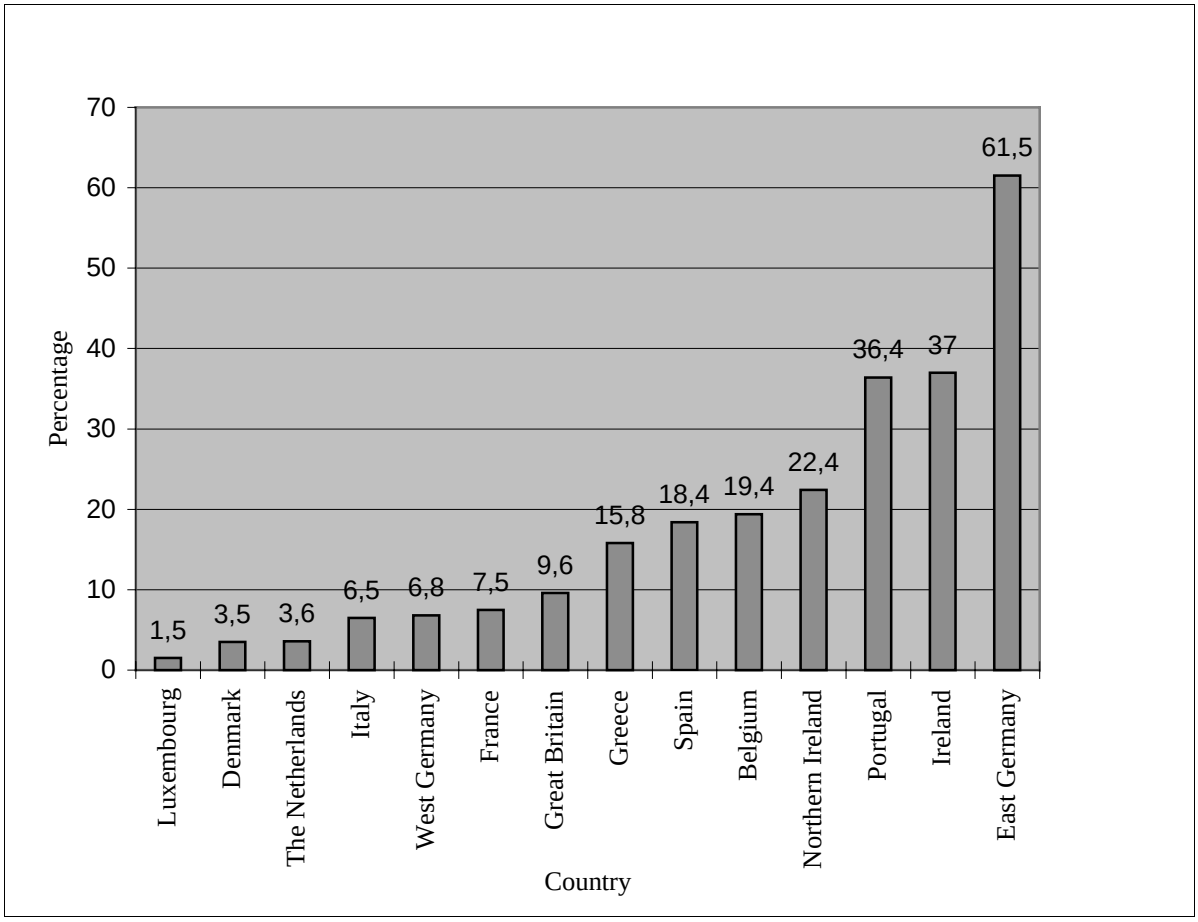
Table 4.1 N of cases in the cumulated Eurobarometer dataset (38, 38.1, 39, 39.1, 40, 41)

Country	N
Luxembourg	3140
Denmark	6005
The Netherlands	6049
Italy	6206
West Germany	6205
France	6094
Great Britain	6391
Greece	6024
Spain	6050
Belgium	6229
Northern Ireland	1832
Portugal	6002
Ireland	6085
East Germany	6336
Total	78648

In each Eurobarometer the interviewers have been asked to record whether the interviewed household has got a telephone or not. The respondents have not been asked explicitly. Figure 1 shows the national rate of nonownership.

Three groups of countries can be identified. The first group of countries with a high telephone density (more than 90%) includes Luxembourg, Denmark, the Netherlands, Italy, West Germany, France, Great Britain. A second, middle-ranged group (below 90% but above 70%) consists of Greece, Spain, Belgium, and Northern Ireland. The third group (below 70%) comprises countries with an extremely low density (Portugal, Ireland, East Germany). In the last group of countries, the situation is changing rapidly. Particularly in East Germany the change is almost dramatic. A 10% increase in ownership (from 39% to 49%) occurred between Fall 1993 survey and Spring 1994 survey and an increase of 9% (from 49% to 58%) between the surveys conducted in spring and in fall of 1994 (this survey is not further considered due to a lack of some data).

Figure 1: Percentages without a telephone in household



The data themselves were gathered in a sampling procedure, which is not necessarily free of errors. There is some likelihood that not all intended interviews were completed, because the fieldwork was finished after a target count of interviews had been realised. In addition, these results concerning telephone ownership are merely based on interviewer observations, which may be systematically biased. For example, the interviewer may not have noticed the telephone if the interview took place in a room other than where the telephone was located. For some countries telephone density data are available from the Statistical Office[19]. These data are used for checking. Nevertheless, the sample distributions closely correspond to the statements from the official data source. The correlation between both sources is really high (r = .86).

The used variables and the categories are presented in table 4.2 To evaluate the differences between owners and nonowners, nine demographic variables are examined. In the last part of the chapter some standard attitude variables are included for demonstrating attitudinal differences between owner groups.

19 The data (number of telephones per 100 inhabitants) exclude the owners of mobile telephones.

Table 4.2 Variables used

Variable	Categories
Gross household income	quartiles
Purchasing power	index based on responses to questions regarding the possession of 10 durable goods with four categories: low, medium low, medium high, high
Social class	middle class, lower middle class, working class, upper class, upper middle class, refuses to be classified, other
Occupation of the respondent	looking after household, student, unemployed, retired, unskilled manual worker, skilled manual worker, supervisor, salesman or driver, farmer or fisherman, shop-owner or craftsman, deskworker, employed professional, professional, middle management, general management
Education	9 categories from 14 years up to 22 years, category 'still studying' is according to the respondent's age distributed over the other categories
Age	recoded into groups: up to 34 years, 35 to 49 years, 50 to 64 years, 65 and older
Sex	female, male
Size of household	number of members, 5 and more put together
Subjective size of community	rural area or village, small or middle size town, big city
Left-right-self placement	harmonised into 3 categories: left, centre, right
Frequency of political discussions	frequently, occasionally, never
Index of 3 media Involvement variables: TV, radio, newspapers	5-point-scales have been added and divided by 3 and then recoded into 5 scale categories

Some remarks concerning the comparability of the items are necessary. A preferable objective size of the community variable was skipped because the categories differ between countries. Therefore the subjective variable has been selected. The income variable is a very difficult one due to response behaviour and technical procedure of harmonisation throughout all the countries. The DK/No answer - refusal-category of the income variable is with 25% still high. A deeper examination shows that people with higher incomes probably tend to refuse an answer. The harmonisation of categories throughout the countries is not exact because the scope of original categories varies from country to country. Thus, a purchasing power variable yields more complete information. The proportion of missing data on the purchasing power variable is slightly more than 1%, and therefore still low. The Eurobarometer contains information about ownership of ten durable goods. The answers are recoded into an index reflecting the purchasing power. Purchasing power as operationalised as it is done here provides information about the extent to which a household's needs are satisfied. Of course, income and purchasing power are attached to the same dimension (the correlation coefficient of both variables r=.5).

An additional methodological problem should be mentioned. The possession of a telephone is a characteristic of a household. To explore ownership of telephones data referring to the whole household should be used. Some data examined refer to the individual level. Thus there is a gap between what is intended to study and what the data really cover in some cases. Income, size of household, purchasing power do not cause a problem. These variables refer to the household. In other cases like class or basic political orientation a relative homogeneity between all the adult members of a household could be assumed. Such clearly individual variables like sex or age constitute a problem. This problem of a gap between household and

individual remains unsolved and can only be mitigated by a careful interpretation of the results aware of the „noise“ in the data.

4.3.1 The composition of the group of telephone-nonowners

What are the demographic differences between people with and without a telephone? This question shall be discussed with an special attention to the quantitative dimension of telephone coverage. This is useful because of the possible consequences. If there are also and perhaps stronger differences in countries with a high coverage it may be necessary to correct the sample in each country by a different procedure.

A graphical way of presentation has been chosen to describe the very rich data material[20]. The following figure show percentage differences. The proportion of people, who belonged to a special category in the group of owners, is subtracted from the proportion of those, who belonged to a special category in the group of nonowners. Thus, a positive sign indicates an overrepresentation of a specific subgroup and a negative sign indicates an underrepresentation of a specific subgroup within the group of nonowners[21].

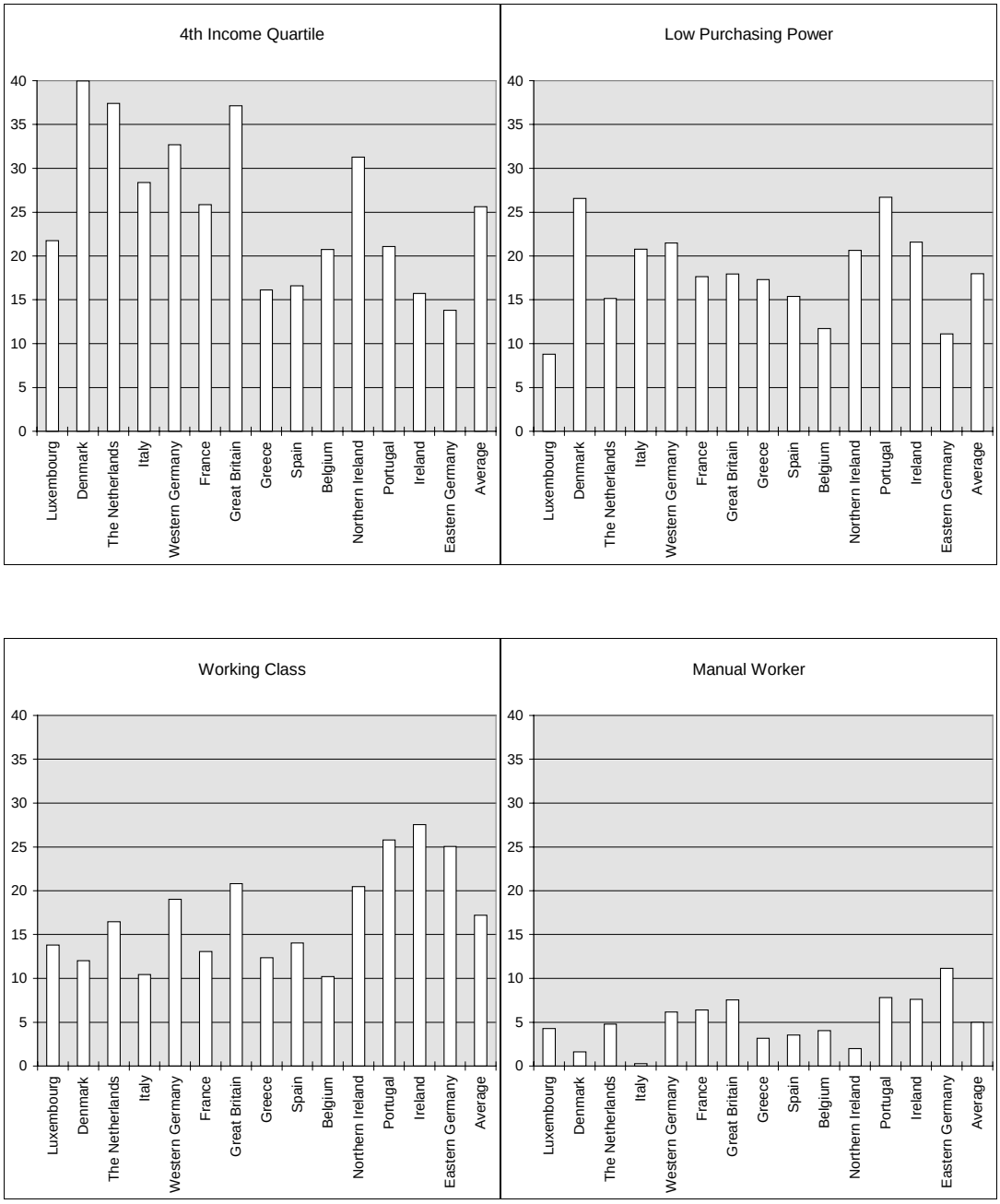
Figure 2 shows the results for some variables which refer more or less to socio-economic status. The variables are family income: falling into the lowest income quartile; low purchasing power: having only two out of ten durable goods; occupation: being a manual worker; social class: belonging to the working class. The percentage differences deliver a simple message. The expectation that lower status groups tend to have no telephone is confirmed. In the group of nonowners people with a low income and low purchasing power, members of the working class, and manual workers are overrepresented. The amount of differences varies: it is large on the income variable and it declines on the other variables with manual workers being the lowest.

The data indicate a relationship between telephone density in a country and an income effect. Especially in the group of countries with a very high telephone coverage the deviation is enormous. In Denmark, the Netherlands and Great Britain almost 40% more households with low income can be found in the group of nonowners than in the group of owners. In countries with a lower telephone coverage like Ireland or East Germany the income effect is not that strong. This relationship seems to indicate a process of concentration of nonownership in some special groups as the process of technological modernisation goes further.

20 The presentation is limited to categories only, where „important“ and more or less systematic differences can be observed. Missing values are excluded. The significant categories of all variables are shown later in the CHAID-analyses. There is a rank order between the countries: Luxembourg with the highest telephone density is always presented first and East Germany with the lowest always last.

21 For example: does the characteristic of being a manual worker play a role for telephone ownership or not? The structural composition of the two groups of owners and nonowners of a telephone is compared. If being a manual worker does not play a role, the proportion of manual workers should be the same in both groups.

Figure 2: The social characteristics of telephone-nonownership I



As far as purchasing power is concerned the distribution across the countries is more equal. Only two countries, Denmark and Portugal - both with a different telephone profile of coverage, are above average. In summing up, the deviation on this variable between the two ownership groups is remarkable. This is also true for members of the working class. On average nonowners belong 17% more to the working class than owners. The difference is sharper in countries with a low than in countries with a high density. There seems to be an inverse relationship in comparison to income. This impression is strengthened by the next

variable. Since class membership reflects the subjective side of self description occupation corresponds more to the objective one. Being a manual worker, whether skilled or unskilled, has an impact on ownership as well. The effect is much smaller than in the case of the subjective variable, but the relationship across countries seems to be the same. Thus all the data taken together give evidence of a status effect, even in the countries with high telephone coverage this effect exists, but the aspect of status which does have an impact varies between the countries.

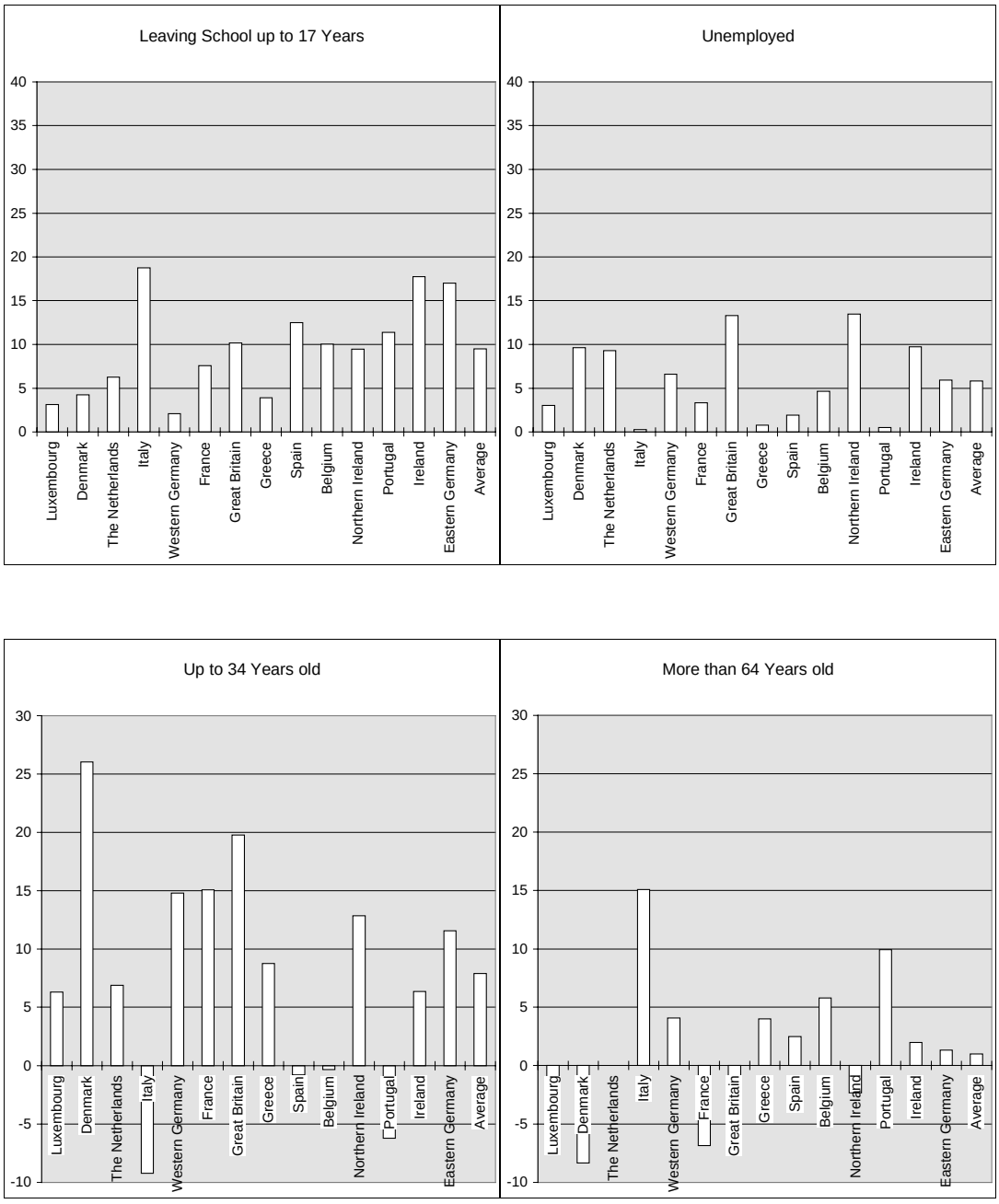
Figure 3 offers information about other (in part) status related variables: education (school leaving up to 17 years) and unemployment. People who left school early and the unemployed are overrepresented in the group of nonowners. No country makes an exception here. Concerning education the country-specific pattern is the same, if the outlier Italy is ignored, as is the case at working class and at manual worker. This is not surprising because these variables are correlated. The variable "unemployment" can reflect a tendency to be poor and may also reflect status, because the risk to become unemployed is unequally distributed throughout the different status groups. But a country specific relationship could not be found.

The age group variable is not directly linked up to status and it may reflect life style. In addition, it should be repeated that age is an individual characteristic. In a case of a family persons of different age groups could belong to a household which owns a telephone or not.

However, age may be connected to a person's material resources that determine to what degree that particular person is able to participate in consumption. There are typical risk pattern of material scarcity in the life cycle. The explanations for this phenomenon were documented long ago. Younger people at the beginning of their professional careers may earn too little to fulfil all their expectations as consumers or to satisfy all the needs of their families.

Retired people may not be capable of compensating the loss of their regular income by public pensions and so on (Kohl, 1992). The data confirm this idea at least in part. Especially younger respondents are overrepresented in the group of nonowners. There also seems to be a slight country-related trend suggesting the higher the telephone density the more are the younger overrepresented. Italy, Spain, Belgium and Portugal do not fit into that pattern. Belonging to a younger age group may also reflect a way of life differing from the life style of older groups. Concerning the age group of older respondents the picture looks different. There is a tendency that the higher telephone coverage in a country the more are the older people underrepresented in the group of nonowners. Italy is an outlier both as regards education and the younger generations variable. Again the data give evidence that remarkable differences do exist between ownership groups.

Figure 3: The social characteristics of telephone-nonownership II



The last series of figures deals with some basic demographic variables: size of household, sex, and size of community. In Europe lacking material resources vary with the size of household. One person households and households with five or six and more persons are more likely to be poor (Institut für Sozialstudien, 1990:45). Thus nonownership is more likely in these groups than in others. All over the examined European countries the one person

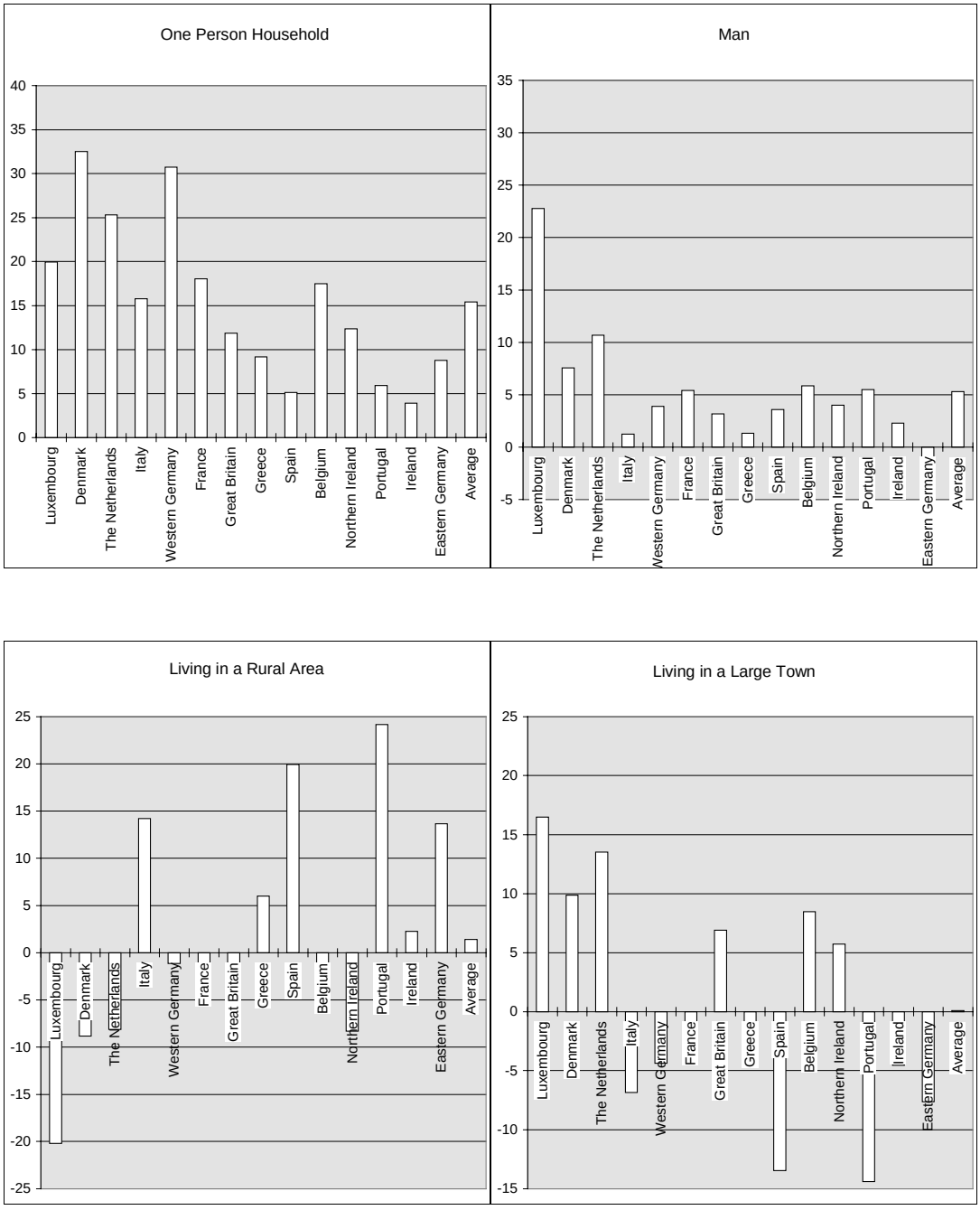
households are overrepresented with 15% at average in the group of nonowners. This difference is one of the highest found in one of the categories of the nine variables considered. In addition, there is a strong country and coverage relationship: the higher the telephone coverage the higher the overrepresentation in the group of nonowners. The five and more person households are not overrepresented. It can be argued that the proportion of one person households is an indicator of individualisation with a North-South difference. The proportion of one person households is in Denmark (35%) and in West Germany (35%) higher than in Spain (11.2%) or Portugal (13.4%)[22]. Family structures are more effective in the South (Hradil, 1992:65). Regarding the higher overrepresentation in the North the process of individualisation obviously goes together with some forms of disintegration or separation from the outside world, although one may think that people living alone should be particularly interested in establishing contacts with the outside world.

The respondent's sex also makes a difference. If a man lives in a household there is a slightly higher probability that the household does not own a telephone. Australian empirical studies show that there is a "pervasive feminine culture of telephone" (Moyal, 1990:196). The usage of telephones refers to the social role of men and women. There is a tendency that men use telephones for instrumental reasons (appointments, making arrangement, purchasing, seeking information) while women use them for intrinsic calls (personal exchange and communication, counselling) (Noble, 1990). It may be that this more a family and neighbourhood contact organising behaviour of the often homemaking women leads more to telephone ownership than the more instrumental orientation of men, who have more favourable opportunity structures for maintaining contacts at their workplace. There is also a very slight relationship with country specific telephone coverage.

Beyond individual variables there may also be a simple structural impact on nonownership. It may be that in some countries of the EU the communication infrastructure is underdeveloped in rural areas compared to towns (Garnham, 1988). Effects of the community size variable on ownership may indicate differences in development of infrastructure. Especially in the South, the rural districts are overrepresented: in Portugal, but also in Spain, in Italy and in Greece, although the latter three countries have a higher coverage than the first. But this can be interpreted as an indicator that developing infrastructure in towns has been given priority. Again, there also exists a relationship with coverage. Rural areas are particularly well served in countries with a high coverage (Luxembourg, Denmark, and the Netherlands). A reflection of this finding can be found if the population in „large" towns is examined. This country-related pattern also backs the supposition that infrastructure does indeed have an impact on telephone ownership.

22 See Statistisches Bundesamt 1994.

Figure 4: The social characteristics of telephone-nonownership III



4.3.2 Interaction: which variables interplay?

The composition of the nonowners throughout the European countries are shown and some expectations are confirmed. So far the description has been univariate or bivariate. Now the description changes to the multivariate level. What about interaction effects? It may be that

some variables influence nonownership only in combination with others. It may also be that some variables strengthen each other mutually in their relationship with nonownership. As a third possibility, it may occur that some effects detected on the univariate level may only reflect an underlying dimension like status and that they go down in effects of other variables. As a consequence, effects of low education level, unemployment or being a one person household may disappear in a multivariate analysis. Therefore a simple CHAID approach[23] is used. Compared to the previous description all categories of variables are examined including the missing data category. In addition, information concerning the size of a group is supplied. CHAID is an explorative approach and a problem of exploration exists. For example, if ten random variables get correlated at least two significant coefficients appear. But in this case some criteria exist to decide whether a prediction is plausible or not. For the sake of readability the CHAID trees are omitted. Instead of this only the most deviant groups are presented in table 4.3.

In the CHAID-analysis, the following criteria for the detection of extremely deviant groups have been used: significance $p < 0.5$, no group should be smaller than 2 % of the population, group should be 50 % or more (East Germany: 25%) above average of nonownership (normal type), 75 % or more (East Germany: 35%) above average of nonownership (italic), 100 % or more (East Germany: 45%) above average of nonownership (bold). Under the given condition that no segment with a high proportion of nonowners should be smaller than 2% of the population older than 14 years and the deviation should be extreme (at least 50% above average except East Germany) in most cases three variables are sufficient to „predict“ nonownership. Almost in every country subgroups could be identified where the proportion of nonowners is twice as high as among the average population (in East Germany a ceiling effect is working because of a proportion of more than 60% of nonowners). Especially in countries with a high density such subgroups could be found. For evaluating this deviation the size of the group should be taken in consideration. The higher the proportion of nonownership and the larger the group the more serious the deviation. In the table these important cases are marked with a *. It can be seen that only in half of the cases the group with the highest proportion of nonownership is also the most important one.

23 It identifies those groups which are most likely to have no telephone. CHAID segmented the respondents into groups which differ with respect to ownership as a dependent variable. It sorts the „predictors“ according to their significance (chi square). That means the „best predictor“ is that variable where the observed relationship between an independent and dependent variable is most likely. Or in other words, the differences in the proportion of the dependent variable are highly significant in one or more categories of the independent variable. In the next step the categories of the „best predictor“ are split up into smaller subgroups of the second „best predictor“ and so on (Kass, 1980; DuToit et al., 1986).

Table 4.3 Causes of nonownership: most deviant groups (results of a CHAID-analysis)

Country	Description of the Groups	% population (A)	% non-owners (B)
Luxembourg	3rd or 4th income quartile and small or large town and man	7.8	7.0
Denmark	4th income quartile and up to 49 years old and low purchasing power	3.8	22.2
	4th income quartile and up to 49 years old and not low purchasing power +uk*	9.0	10.8
	3rd income quartile and low purchasing power +uk	2.9	8.0
The Netherlands	4th income quartile and man and one person household +uk*	3.5	18.4
	3rd income quartile +uk and one person household and lower middle class or working class or other +uk	2.8	13.7
	4th income quartile and man and more than one person household	4.8	7.9
	4th income quartile and woman and up to 64 years old*	8.5	7.6
Italy	low purchasing power +uk and one person household*	4.7	25.5
	low purchasing power +uk and two or three persons household and lower middle class or working class +uk	6.3	15.7
	medium low purchasing power and lower middle class or working class or refused or other + uk and education up to 15 years +uk	6.3	14.5
	medium low purchasing power and middle class or upper middle class or upper class and rural and man	3.4	11.3
West Germany	4th income quartile and low purchasing power +uk and up to 49 years old	2.5	34.4
	4th income quartile and not low purchasing power and man and one person household	3.2	21.5
	4th income quartile and low purchasing power +uk and older than 49 years*	4.9	18.4
	3rd income quartile and low or medium low purchasing power and up to 34 years old	3.1	15.4
	4th income quartile and not low purchasing power and woman and working class or refused +uk	2.7	14.9
France	low purchasing power +uk and up to 64 years old and man*	4.7	24.4
	low purchasing power +uk and up to 64 years old and woman and rural or small city	2.8	23.4
	medium low purchasing power and man and up to 34 years old	5.5	16.2
	low purchasing power +uk and up to 39 years old and large city	3.6	15.7
Great Britain	4th income quartile and up to 49 years old and man	2.7	47.7
	4th income quartile and up to 49 years old and woman*	4.1	35.9
	4th income quartile and older than 49 years and low purchasing power +uk	4.2	22.6
	unknown income quartile and medium low purchasing power and up to 34 years old	2.6	21.9
	3rd income quartile and up to 34 years old	5.0	21.2
	unknown income quartile and low purchasing power	2.8	20.6
Greece	unknown purchasing power*	4.6	34.7
	low purchasing power and one person household	4.1	34.4
	low purchasing power and more than one person household and looking after household or student or unemployed or employed	4.1	33.3

	professional or management or employed position (working mainly at a desk) or manual worker		
	medium low purchasing power and up to 34 years old and up to two persons household	2.6	29.6
	low purchasing power and more than two persons household and up to 34 years old and 2nd or 3rd or 4th income quartile	4.2	27.3
Spain	low purchasing power +uk and rural*	10.2	36.9
	medium low purchasing power and small city and education up to 18 years +uk and man	4.1	29.6
	medium low purchasing power and rural	11.3	28.8
Belgium	low purchasing power +uk and man	4.6	43.2
	medium low purchasing power and one person household +uk*	5.6	38.6
	low purchasing power +uk and woman	6.0	34.8
Northern Ireland	low purchasing power +uk and up to 34 years old	5.1	62.8
	low purchasing power +uk and older than 34 years and 4th income quartile	5.3	53.6
	medium high purchasing power and up to 34 years old*	9.7	42.1
Portugal	unknown purchasing power*	6.0	80.3
	low purchasing power and working class +uk and unemployed or manual workers or supervisor or farmer/fisherman or employed position (working mainly at a desk) or professional or management	4.4	74.9
	low purchasing power and working class +uk and looking after home or student or retired or shopowner/craftman or salesman/driver and up to two persons household	5.3	63.2
	medium low purchasing power and working class +uk and up to 34 years old	2.5	56.9
	low purchasing power and lower middle class and rural	4.1	55.4
Ireland	low purchasing power +uk and working class +uk and up to 34 years old*	3.9	77.0
	low purchasing power +uk and working class +uk and between 35 and 64 years old and man	2.7	71.6
	medium low purchasing power and unemployed	2.7	68.7
	low purchasing power +uk and working class +uk and between 35 and 64 years old and woman	2.6	60.5
	medium low purchasing power and manual worker	4.7	57.7
	low purchasing power +uk and working class +uk and older than 64 years	5.1	57.7
	low purchasing power +uk and lower middle class or refused or other class	3.4	56.5
East Germany	working class and small or large city +uk and up to 29 years old and up to two persons household	2.5	91.7
	working class and rural and low purchasing power +uk	3.5	90.1
	working class and rural and medium low purchasing power	6.9	84.1
	working class and rural and higher purchasing power and man	2.9	81.1
	working class and small or large city +uk and 2nd or 3rd income quartile and up to 34 years old	4.5	79.0
	working class and small or large city +uk and 4th income quartile*	7.7	77.8

* indicates „Importance“ defined as the product of A and B

As indicated in the first description the impact of variables varies with telephone coverage in a country. In the country group of high telephone density, with the exception of Italy and France, income is the „best predictor“. In the two other groups purchasing power is the dominant factor. But in the group, containing Portugal, Ireland and East Germany working class is a overriding feature. In East Germany it is even the „best predictor“. It is not surprising that the status-related variables are significant, but the really important result is that the pattern depends on the country. Of course, the general pattern needs some qualification. Working class plays also a little role in some cases like the Netherlands, Italy and West Germany. But in this group it tends to be more a phenomenon of working class and lower middle class. In addition, in some cases not only the lowest purchasing power and income group is involved. But the general pattern remains untouched by this qualifications.

Second, a further important result is that the „predictive power“ of less education, being manual worker and unemployment evaporates in most cases, although on the univariate level evidence of a substantial effect has been found. That means in most cases that the effect is dissolved by other variables like income, purchasing power, and belonging to the working class. In part the same seems to be true for household size. In Denmark, where the largest effect of household size was found on the univariate level, the household variable loses its “predictive power”, but on the other hand in other countries like for instance Luxembourg, France, Great Britain. In the Netherlands the impact of household size remains remarkably strong.

Third, although the effects of sex have been rather slight on the univariate level they are not neutralised altogether. In eleven of 14 countries being a man plays a role.

Forth, age has some impact. In all the countries investigated it could be generalised that nonownership is more the matter of all age groups up to 64 years old than of the older people. In 26 interaction terms involving age only in four cases the age group is older than 64 years. But this must not be interpreted as follows: even poor older people try to arrange to have a telephone because their action radius is limited and they want to maintain a communication device. In four cases being older has an automous effect.

Fifth, effects of infrastructure are hard to detect. This would be the case if in a country also in any other than in a low status group the overrepresentation of rural areas would be visible. This could be the case in Spain.

Sixth, in two cases a deeper cause of nonownership is visible. There is an apparent inconsistency in Denmark and in West Germany each in the second group. In both cases the group is characterised by the forth income quartile and by medium low or higher purchasing power. This inconsistency can be interpreted as an indicator of indebtedness of the respective household. This indebtedness is the deeper cause of nonownership.

Seventh, Italy tends to be exceptional. It does not really fit into the country-pattern. Although it falls into the group of high telephone coverage the variable of purchasing power and not of income has an impact. Working class membership and low education levels play a role here. In Italy the last group is characterised by an possible contradiction. Although the respondents claim to be members of at least the middle class they seem to be ascetic.

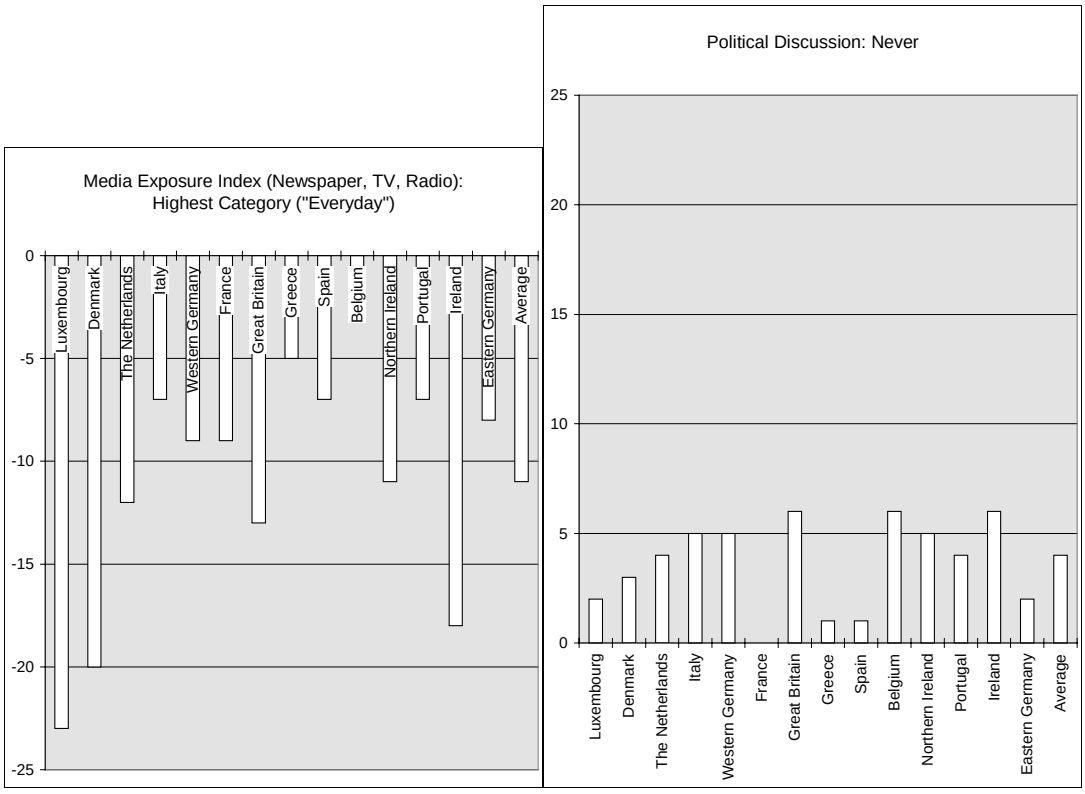
In general, the results of the CHAID-analysis give a structured picture as it is presented in the seven points above. The interactional structure between the variables is explored. The possibility that some variables influence nonownership only in combination with others can be practically neglected. Regarding the interactions the main result is a clearing up of the findings of the univariate level. Some effects (low education level, unemployment, manual worker, one person household) detected on the univariate level reflect in most cases only an underlying dimension like status or something else like poverty, although some effects have been remarkable. Now they go down due to effects of other variables and they disappear. There are some variables that strengthen each other mutually in their relationship with nonownership. There are also some other surprising interactions like between low income and low purchasing power, but these effects remains exceptional. Sex is a candidate with an autonomous effect despite the problem that it is as individual feature and not a characteristic of a household.

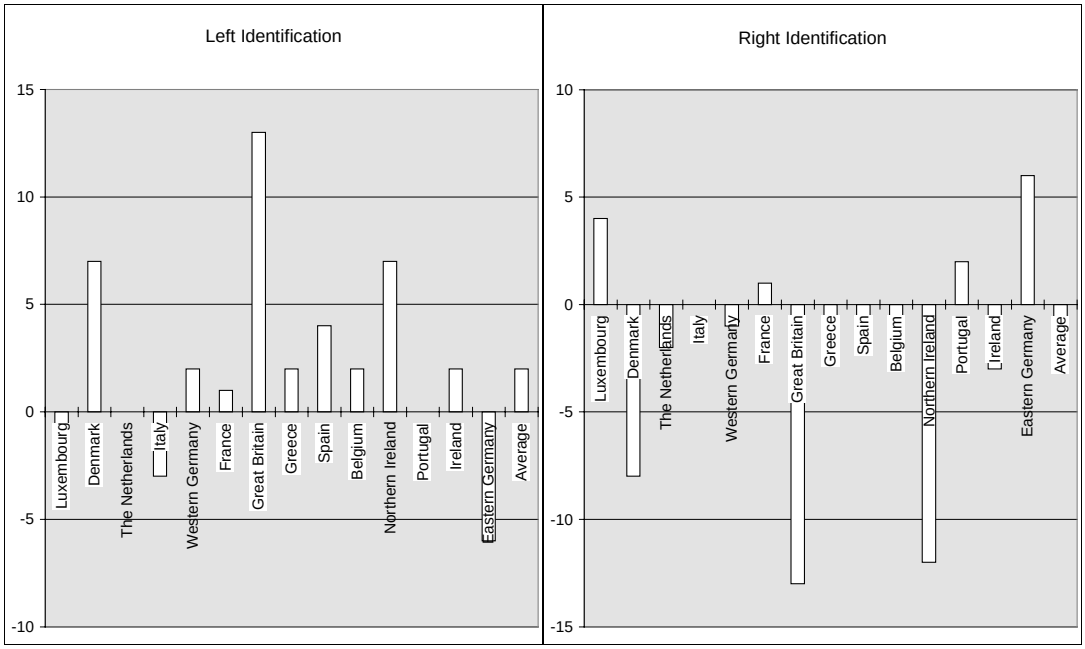
4.4 Effects on attitudes

In the normal case of survey research attitudes are the central variables. If socioeconomic differences exist, it does not necessarily mean that these cause substantial differences at the level of attitudes. Considering the literature, which argues that there is an ongoing process towards stronger individualisation in Western societies (Beck, 1986), the relationship between social structure and attitudes may become weaker or may eventually disappear (Schnell and Kohler, 1995). In addition, a developed, nation-wide system of mass communication can compensate original differences. Thus, what are the attitudinal differences between people with and without a telephone? If the status hypothesis reflects reality, attitudinal effects should also be supported by the data. There may also be behavioural effects. It could be expected that nonowners are less involved into the political information flow. They may also be less inclined to discuss politics. This is primarily due to a lack of motivation. Cognitive effects may be visible in the way how politics is structured or at least evaluated with the left-right scheme. In the group of nonowners a tendency to use the left ideological label should be found.

Figure 5 presents the results. The variables show the expected pattern. Nonowners tend to be less involved into the information flow. Thus, they can be less mobilised, they are less affected by new ideas. Consequentially they participate less in political discussions. In some countries (Denmark, Great Britain, Northern Ireland), there is also a tendency to think more left in the group of nonowners.

Figure 5: Attitudinal characteristics of telephone-nonownership





Again, differences have been found. It is possible that these differences do not matter, because the group of nonowners is very small in most of the countries. The proportion of nonowners in the sample must be taken into consideration. That is the quantitative question which will be addressed in the next chapter. Here we have at least seen that there can be large differences between owners and nonowners of telephones.

4.5 Conclusion

In quantitative perspective differences in telephone coverage exist. But in the not fully covered countries the situation changes rapidly. And looking at market and opinion research in West Europe one can find a simple rule of thumb: the higher the telephone density the higher the proportion of telephone interviews (ESOMAR, 1995)[24]. Telephone interviews will increase even further. Nevertheless a „stratum bias“ in telephone surveys even in countries with a high coverage still exists. Especially in countries with a high density extreme deviant subgroups could be found.

To summarise the qualitative results: There are substantial differences between owners and nonowners. The groups differ at the socio-demographic level. People, who earn less than average, who have a low purchasing power, who describe themselves as belonging to the working class tend to be among those candidates, who do not have a telephone. On some variables like income the differences are particularly large for countries with a high coverage. The most remarkable result is that the appearance of this bias changes between country groups. In the countries with high density income is the key variable and in the countries with low density working class gets more important. In the more modernised countries class stratification is weakened (Clark and Lipset, 1991). In the countries with low density value systems and lifestyles related to class stratification seems to be influential. Not having a telephone is part of this lifestyle.

But other variables play a role too. In the Netherlands one person households would be underrepresented. It is more unlikely that older people do not have a telephone than younger. In addition, households where a man lives are undercovered. There is a general tendency in telephone surveys that women are overrepresented because housewives are easier to contact by phone than outside the home working men. The underrepresentation of men in telephone surveys would be strengthened by the overrepresentation of man in the group of nonowners. The found demographic differences seem to have an effect on attitudes. It should not be forgotten that the basis of the presented calculations may also be biased. The problems of representativity of classic face to face-surveys are well known (see chapter 3). That means that the real bias in telephone surveys may be more serious than reflected in this study. Here we saw large differences between owners and nonowners. How large the bias in responses is due to coverage will be an issue in the next chapter.

---

24 The European Society for Opinion and Marketing Research (ESOMAR) has published findings according to which, for example, in Denmark 53% of all quantitative data collection efforts were conducted by telephone in 1994 (Luxembourg 75%, Germany 29%, Greece 26%, United Kingdom 18%, Spain 32%, Portugal 12%, Ireland 12%).

CHAPTER 5

SIZES OF THE DIFFERENT EFFECTS: COVERAGE, MODE AND NONRESPONSE

JÜRGEN LASS, WILLEM E. SARIS AND MAX KAASE

5.1 Introduction

In chapter 1 it was mentioned that in this study attention will be given to the three main sources of differences in surveys: coverage differences, mode differences, and nonresponse or organisational differences. These differences which occur in any survey research can explain why results obtained for the Eurobarometer 41 by INRA using the face to face mode of data collection differ from those from a telephone survey carried out at about the same time by FORSA. The sum of these differences - the total differences - has been described in chapter 3 for a number of questions. In chapter 3 it was argued that the samples of the two studies were really different and that weighing on the basis of background variables did not improve the results for the variables of interest. The problem of coverage errors has been discussed in chapter 4. Coverage errors are due to a systematic error in the sample design caused by differences in telephone ownership; in the telephone books nonowners of telephones are not present. Therefore the use of telephone books as a sampling frame for surveys of general populations will lead to a systematic bias in the sample. Chapter 4 shows that the owners of telephones differ considerably in several aspects from the nonowners of telephones. Therefore, a systematic bias in the samples can be expected due to this factor.

A mode difference is the difference in response distributions produced by the special features of two different observation techniques. Chapters 6 to 9 will give lots of attention to this issue when comparing face to face and telephone interviews.

The nonresponse differences result from differences in sample designs and the fieldwork by survey organisations for instance in establishing contact with a household or in dealing with a refusal to co-operate in an interview on the part of the household. Each organisation has its own procedures of going about such problems which leads to specific nonresponse errors. These differences between organisations obviously can also lead to differences in results.

The purpose of the study reported in this chapter is to quantify the coverage difference (C), the mode difference (M), and the nonresponse (N) or organisational difference for different questions. The total difference (T) between face to face and telephone interviews can be decomposed into the three mentioned differences.

As shown in chapter 1, the corresponding equation is rather simple:

$$T = C + M + N \quad (1)$$

The total difference (T), the coverage difference (C) and the mode difference (M) can be estimated independently of each other. The remaining forth component, the nonresponse effect (N), can only be derived from them. The purpose of this chapter is to provide an estimate of these different components.

It should be clear from the outset that only differences are discussed and not biases or errors, let alone the bias of a specific mode or organisation. The data does not allow for such analyses except in the case of differences due to coverage errors. In face to face interviews people can be asked whether they have a telephone. If this is not the case, these people will drop out in telephone interviews. By comparing a full sample with the sample of telephone owners, one can study the bias caused by this coverage error. For the other two components only differences can be detected. Nonresponse differences are due to differences in fieldwork. Face to face interviewing may lead to more co-operation, but telephone interviewing has the advantage that the fieldwork can be better controlled and attempts of contacting a household can be cheaply repeated. Unfortunately, we cannot derive a measure of the quality of the fieldwork (N) for each research organisation separately. Only the differences in results can be presented.

The same holds for the mode effects: differences can be detected, but they permit no conclusion as to which mode is better. The consequence from the above conclusion is that the differences between the methods of data collection will be shown and these differences will be decomposed in the three mentioned components. This analysis will reveal which component is larger, but we will not be able to say which measure performs better.

In this study the estimation of the different components is restricted to those countries where panel data are available: France with a high telephone density, Belgium and Spain with a medium-sized density. 15 Variables with non-political and political, national and international references are examined. Variables of factual political knowledge (knowledge of date of European Parliament-Election) or behaviour (probability to vote) are skipped. This is because effects of learning or mobilisation during an election campaign can not be excluded which may affect the panel data. Besides the attitudinal variables, four demographic variables (possession of durable goods) are included. For an overview of all variables we refer to table 5.3 and for the formulation of these questions in the different studies we refer to Appendix 1.

The analysis starts with the estimation of the total differences, continues with the coverage errors, goes to the mode differences and finally derives the organisational differences. The presentation follows a simple format. To demonstrate the procedure of the calculation the “benefit” variable is selected because of its simple structure and its relevance in the Eurobarometer. In table 5.1 the percentages of each response category are shown for each country. In a second step (see table 5.2) the differences for each category between the face to face and the FORSA telephone survey are shown. From these differences, in table 5.3 an overall measure for the differences is presented. This measure is calculated for all variables analysed in this chapter. In the same way to the other analysis are presented.

5.2 Total difference

The estimation of the total differences between face to face and telephone data is based on the comparison between the Eurobarometer 41 and the FORSA telephone sample. Table 5.1 shows the distribution of answers for the variable "benefit" for the three countries which have been studied through the additional EB 41 panel component.

Table 5.1 The responses in percentages in face to face and telephone interviews carried out by two different organisations for the benefit variable

	France		Spain		Belgium	
	EB41.0	FORSA	EB41.0	FORSA	EB41.0	FORSA
Benefited	39.4	43.7	38.6	48.6	48.3	63.4
Not benefited	39.3	31.5	43.4	29.4	27.8	16.2
DK/ No answer	21.4	24.8	18.0	22.0	23.9	20.4
N	1034	501	1003	500	1081	500

This table shows that there are considerable differences between the face to face and the telephone study. Table 5.2 presents these differences in detail. To calculate the percentage point differences, the response from the face to face study are subtracted from the percentages of the telephone survey.

Table 5.2 Percentage differences between the face to face interviews and the telephone interviews for the "benefit" variable

	France	Spain	Belgium
Benefited	4.3	10.0	15.1
Not benefited	-7.7	-14.0	-11.6
DK/No answer	3.4	4.0	-3.5
Adjusted total difference	7.7	14.0	15.1

The absolute differences of single categories range from 3.4 to 15.1. Clearly, there are considerable discrepancies between these two measurements. In order to obtain an impression of the difference between the two modes of data collection, all differences ignoring the signs are summed up and devided by two. The resulting total differences between the two modes for France Spain and Belgium are: 7.7%, 14% and 15.1%. The same calculation has been done for all variables in this study, and the results are reported in table 5.3.

The resulting figures are of a quite considerable magnitude: 14 out of 45 differences are higher than 10 percentage points. These results again demonstrate very clearly that findings from studies using different modes of data collection can not be compared directly. In this case the data are collected at the same time and supposedly for the same populations. Nevertheless, in all three countries for at least some questions large differences have been found, also the standard Eurobarometer questions are effected.

In the next sections we will analyse where these large differences come from.

Table 5.3 Adjusted total differences for 15 variables in three countries

	France	Spain	Belgium
Satisfaction with life	10.9	21.9	4.6
Satisfaction with democracy	4.9	8.8	3.4
Persuade others	17.1	11.6	7.8
Political discussion	1.7	6.3	10.4
News on TV	6.6	3.8	3.8
News daily papers	7.8	8.1	6.9
News on radio	13.8	11.2	13.3
Interest in European politics	10.5	13.8	9.3
Level of EU informedness	8.4	11.1	11.0
Membership in EU	6.7	7.0	16.6
Benefit from EU membership	7.7	14.0	15.1
Colour TV	4.1	0.7	2.5
PC	0.2	7.4	8.4
Two or more cars	4.7	10.6	9.3
Second home	0.8	4.9	1.0
Mean	7.1	9.4	8.2

5.3 Coverage errors

In chapter 4 large differences between owners and nonowners of telephones are revealed. It was also found that especially in countries with a low telephone density the differences between owners and nonowners can be very large. Whereas when the group of nonowners is small, the coverage differences will likely be small as well.

In order to assess the coverage error, the question is asked: What are the percentage point differences between a sample drawn from telephone owners and nonowners and a sample drawn only from telephone owners? For all variables tables presenting these differences can be constructed on the basis of the data from the Eurobarometer 41 face to face study. In these tables the owners of telephones are compared with the complete sample of owners and nonowners. Table 5.4 is an example of such a table for the “benefit” question.

Table 5.4 Coverage errors estimated for the benefit variable in three countries

	France		Spain		Belgium	
	EB41.0	Telephone owners EB41.0	EB41.0	Telephone owners EB41.0	EB41.0	Telephone owners EB41.0
Benefited	39.4	39.3	38.6	39.1	48.3	50.7
Not Benefited	39.3	39.9	43.4	43.6	27.8	27.0
DK/No answer	21.4	20.8	18.0	17.3	23.9	22.3
N	1034	972	1003	793	1081	888

If there were no coverage errors, the percentage differences would have been zero or close to zero. This is clearly not the case, but on the other hand the differences in table 5.4 are much smaller than in table 5.1 for the total differences.

Table 5.5 shows the differences derived from table 5.4 These differences indicate the size of the coverage error for the variable "benefit" in three countries caused by the systematic bias originating from ignoring the people without a telephone.

Table 5.5 Differences* between full sample and sample of telephone owners for the "benefit" variable in three countries in EB41.0

	France	Spain	Belgium
Benefited	-0.1	0.5	2.4
Not benefited	0.6	0.2	-0.8
DK/No answer	-0.6	-0.7	-1.6
Adjusted total difference	0.6	0.7	2.4

*Because of rounding errors the sum of positive and negative values is not exactly zero in each column.

The "adjusted total difference" is calculated like in table 5.2, that is by summing up the differences ignoring the signs and deviding by two. This value gives a clear indication of the size of the coverage error.

Table 5.6 shows the estimates of the coverage error for all variables for the three countries.

The mean values show that the coverage error is larger in the two countries with a lower telephone density than in France, a country with a higher ownership. At first sight, this seems a bit surprising because it was shown in chapter 4 that often the nonowners of telephones in countries with a high telephone density are quite deviant. On second thought, it is apparent that the size of this group necessarily plays a more important role than the extent of the deviation between the group of owners and nonowners.

Besides this general effect a clear pattern does not exist. Neither are the demographic variables particularly deviant nor the involvement variables conspicuous. Scores are also very small, ranging from 0 to 3.4.

From these results the conclusion must be drawn that the coverage problem exists as a systematic error of telephone interviews, but it appears to produce only a very small bias for almost every question for the three countries studied here.

Table 5.6 Coverage error across 15 variables: the adjusted total differences

	France	Spain	Belgium
Satisfaction with life	0.3	0.9	1.4
Satisfaction with democracy	0.3	1.7	0.8
Persuade others	0.6	0.8	1.6
Political discussion	0.7	2.1	1.7
News on TV	0.9	1.1	0.7
News daily papers	0.6	3.4	1.4
News on radio	0.9	1.0	1.0
Interest in European politics	1.0	1.6	2.1
Level of EU informedness	0.7	1.4	0.7
Membership in EU	0.5	0.8	2.0
Benefit from EU membership	0.6	0.7	2.4
Colour TV	0	0.5	0.2
PC	0.5	1.5	2.4
Two or more cars	1.3	1.9	2.2
Second home	0.2	2.0	0.4
Mean	0.6	1.4	1.4

5.4 Mode differences

Mode effects should be visible if answers reported in the face to face interviews are compared with answers reported in the telephone interviews for the same people. Such data have been produced by the panel component of the Eurobarometer study for France, Spain and Belgium and for all three countries. Table 5.7 presents the distribution of answers to the "benefit" question of people interviewed firstly face to face and secondly by telephone in the panel.

Table 5.8 gives the percentage differences derived from table 5.7. The percentages from respondents interviewed face to face are subtracted from the percentages of the same respondents interviewed by telephone. Again an overall measure of mode effect has been calculated.

Table 5.7 The distribution of the answers for the "benefit" variable in three countries from two modes of interviewing in the EB41.Panel

	France EB41.Panel		Spain EB41.Panel		Belgium EB41.Panel	
	face to face	telephone	face to face	telephone	face to face	telephone
Benefited	39.0	39.3	38.2	42.5	51.5	53.3
Not Benefited	41.6	40.2	45.0	44.4	27.0	23.1
DK/No answer	19.4	20.4	16.8	13.1	21.5	23.6
N	338	338	306	306	229	229

The results of the equivalent calculations for all variables and the three countries are presented in table 5.9. These results come as a little surprise. For many questions, the mode effects are substantial. Small mode effects are only found for the media involvement questions , the "benefit" question and the questions concerning the possession of goods. Large mode effects have been found for the satisfaction questions, political involvement, and involvement in the EU.

Table 5.8 Mode Differences* between face to face and telephone interviews for the "benefit" variable in three countries in the EB41.Panel

	France	Spain	Belgium
Benefited	0.3	4.3	1.8
Not benefited	-1.4	-0.6	-3.9
DK/No answer	1.0	-3.7	2.1
Adjusted total difference	1.3	4.3	3.9

*Because of rounding errors the sum of positive and negative values is not exactly zero in each column.

Actually, effects are found where they are less expected and not found where they were expected. Let us start with the last point. Because telephone interviews are done without visual aids, it could be expected that questions supported in this way in face to face studies produce the most different results. This, however, is not the case. The media involvement questions are asked with show cards in the face to face interview, but for these questions the deviation tends to be lower than in some attitudinal questions. This could be interpreted as an indication that the respondents do not need help by visual aids in that case. On the other hand questions about rather remote political topics like the European Union should produce higher deviations because here nonattitudes are highly probable and therefore the interviewing mode could affect the responses (Zaller, 1992). However, also in this case these effects are not very strong. For example, for the "benefit" question the mode differences are rather small.

Table 5.9 Mode across 15 variables in three countries: the adjusted total differences

	France	Spain	Belgium
Satisfaction with life	5.3	5.0	16.6
Satisfaction with democracy	9.3	5.4	10.8
Persuade others	3.6	9.9	10.9
Political discussion	3.1	2.2	9.1
News on TV	0.9	4.7	2.3
News daily papers	2.6	5.2	6.9
News on radio	6.9	6.7	2.7
Interest in European politics	7.4	8.6	2.4
Level of EU informedness	6.1	12.0	15.1
Membership in EU	6.0	6.3	7.0
Benefit from EU membership	1.3	4.3	3.9
Colour TV	2.6	1.0	1.7
PC	5.3	4.2	3.1
Two or more cars	2.8	1.5	1.2
Second home	2.0	2.6	0
Mean	4.3	5.3	6.2

On the other hand there are even differences regarding the factual information although real change was impossible between the two waves of the panel. For interpretation purposes an additional piece of information may be helpful: the differences are calculated on the basis of the category of possession of a particular good. The proportion of those who have been

interviewed by telephone and who said that they possessed something asked tends to be higher than in the face to face interviews. Thus there is no evidence that people are hesitating on telephone to tell what they possess.

The largest differences are found for the satisfaction and interest questions. For these questions the differences are considerable. Altogether the results differ from the standard literature on this issue suggesting that the differences due to mode are minor (Groves and Kahn, 1979; de Leeuw and van der Zouwen, 1989; de Leeuw, 1992). Therefore there are good reasons to study this issue further in the next part of the book.

5.5 Nonreponse differences

The nonresponse or organisational differences can not be estimated independently. This would require a design where the same people have been contacted by two different organisations at the same time, a design which is, of course, impossible. But an estimate of the differences due to fieldwork effects can be obtained from the previously presented results on the basis of equation 1.

From the estimates of the total differences, of the coverage and of the mode differences the differences due to nonresponse can be derived applying the formula

$$N = T - C - M.$$

On each category percentage for each question. Table 5.10 demonstrates this using the "benefit" variable in France.

Table 5.10 Nonresponse Differences in the case of France for the "benefit" variable*

	Organisational effect	Total % effect	Coverage effect	Mode % effect
Benefited	4.1	4.3	-0.1	0.3
Not Benefited	-7.1	-7.7	0.7	-1.3
DK / No answer	3.0	3.4	-0.6	1.0
Adjusted total difference	7.1	7.7	-0.7	1.3

*Because of rounding errors the sum of positive and negative values is not exactly zero in each column.

The same calculation can be done for all variables and all three countries. Table 5.11 presents the results of these calculations.

Neglecting the effects on the four demographic variables which tend to be lower than the others, the organisation effects vary from 3.6 to 21.7 percentage points. It seems that the organisational differences lead to differences which clearly are rather large and larger than the other two sources of differences.

Table 5.11 Estimates of the nonresponse or organisational effects across 15 variables in three countries: the adjusted total differences

	France	Spain	Belgium
Satisfaction with life	9.3	21.7	12.7
Satisfaction with democracy	10.4	9.4	13.9
Persuade others	15.8	3.6	8.8
Political discussion	4.2	3.9	16.8
News on TV	4.7	8.4	3.7
News daily papers	10.4	10.2	6.4
News on radio	7.6	7.8	10.5
Interest in European politics	2.9	6.2	7.3
Level of EU informedness	5.7	6.1	9.3
Membership in EU	12.3	9.9	20.2
Benefit from EU membership	7.1	13.6	10.9
Colour TV	6.7	2.2	4.4
PC	5.6	1.7	2.9
Two or more cars	0.6	7.2	5.9
Second home	1.4	0.3	0.6
Mean	6.9	7.4	8.9

It should be pointed out again here that these differences cannot be contributed directly to one of the organisations. Only differences, not the absolute biases caused by one of the organisations or both, can be ascertained. We have even to add that a part of these differences might be due to sampling fluctuation because we can not separate systematic effects of the organisations and effects of the random sampling in each study. On the other hand, it is clear from the data that large differences are found if two different organisations collect data from the same populations with the same mode of data collection and identical questions, since the figures in table 5.11 are corrected for mode effects and for coverage differences.

5.6 Conclusion

Table 5.12 summarises all calculated effects for the different variables in the different countries. It should be noted that the estimates of the different effects are based on calculations over all categories of variables. Due to that the equation 1 does not hold anymore. This equality holds for each category but not necessarily for the sum ignoring the signs. We prefer this presentation because it gives the maximum effect for each factor.

According to the size of the effects a clear rank order can be established. The coverage differences rank lowest with a mean of 1.1 over all countries and all questions. In contrast to the other sources of differences, it unavoidably occurs in telephone interviews, but its effect is rather small. Mode differences are remarkably larger with an average score of 5.2. It cannot be concluded that this is necessarily due to the telephone interviews. It may be that respondents produce more random answers in telephone interviews because the time pressure is stronger and they are not supported by visual aids. But on the other hand interviewers in telephone interviews are more controlled which means that they ask the question more precisely in the way expected than in the uncontrolled face to face interview situation.

Clearly the largest total differences come from the black box of the fieldwork of both survey organisations. The mean differences over topics and countries is 7.7. percentage points.

Concerning the individual variables some patterns can be observed. First of all we see the questions about different possessions have relatively small differences for all components except the question with respect to the ownership of two or more cars. Here especially the nonresponse effects of the different organisations make a difference. Given that the mode effects are relatively small, these variables will not be analysed with respect to mode effects in the next part.

The media involvement variables also display relatively small total differences except for the question about the radio. For the newspapers this is, however, partially true because the different effects compensate each other.

The unification questions on membership and benefit are acceptable with respect to coverage and mode effects but the nonresponse differences caused by the different organisations are very large so that incomparable results are obtained.

The other two EU involvement questions have mode effects which are even larger than the nonresponse effects which does not occur very often.

Finally for the satisfaction question and the political involvement questions the mode effects are especially large in one country (Belgium) while large nonresponse differences occur for many questions.

Overlooking all these results we have to say that the effects differ from question group to question group. This is not surprising because the strength of the effects is always dependent on the strength of the relationship between the error source and the substantive type of variable. The strength of these relationships differ of course from topic to topic and therefore there appear also differences between the differences in results for the different questions.

Nevertheless, it is clear that the coverage error is the smallest problem and that the other two factors can produce quite large differences between studies done with a different mode of data collection or by a different organisation. In general the effects are so large that the results can not be compared. Therefore we will discuss in chapters 10 and 11 procedures to correct for these differences in order at least to make the results comparable. But before this is done we will first give more attention to mode effects as the second largest source of differences. Unfortunately not much can be said about the organisational differences than the remarks which have been make in chapter 2. Therefore we will concentrate in the book further on mode effects and correction for differences between studies in general.

Table 5.12 A summary of all one directional differences in three countries

		Total	Coverage	Mode	Organisation
Satisfaction with life	France	10.9	0.3	5.3	9.3
	Spain	21.9	0.9	5.0	21.7
	Belgium	4.6	2.0	16.6	12.7
Satisfaction with democracy	France	4.9	0.3	9.3	10.4
	Spain	8.8	1.7	5.4	9.4
	Belgium	3.4	0.8	10.8	13.9
Persuade others	France	17.1	0.6	3.6	15.8
	Spain	11.6	0.8	9.9	3.6
	Belgium	7.8	1.6	10.9	8.8
Political discussion	France	1.7	0.7	3.1	4.2
	Spain	6.3	2.1	2.2	3.9
	Belgium	10.4	1.7	9.1	16.8
News on TV	France	6.6	0.9	0.9	4.7
	Spain	3.8	1.1	4.7	8.4
	Belgium	3.8	0.7	2.3	3.7
News daily papers	France	7.8	0.6	2.6	10.4
	Spain	8.1	3.4	5.2	10.2
	Belgium	6.9	1.4	6.9	6.4
News on radio	France	13.8	0.9	6.9	7.6
	Spain	11.2	1.0	6.7	7.8
	Belgium	13.3	1.0	2.7	10.5
Interest in European politics	France	10.5	1.0	7.4	2.9
	Spain	13.8	1.6	8.6	6.2
	Belgium	9.3	2.1	2.4	7.3
Level of EU informedness	France	8.4	0.7	6.1	5.7
	Spain	11.1	1.4	12.0	6.1
	Belgium	11.0	0.7	15.1	9.3
Membership in EU	France	6.7	0.5	6.0	12.3
	Spain	7.0	0.8	6.3	9.9
	Belgium	16.6	2.0	7.0	20.2
Benefit from EU membership	France	7.7	0.6	1.3	7.1
	Spain	14.0	0.7	4.3	13.6
	Belgium	15.1	2.4	3.9	10.9
Colour TV	France	4.1	0	2.6	6.7
	Spain	0.7	0.5	1.0	2.2
	Belgium	2.5	0.2	1.7	4.4
PC	France	0.2	0.5	5.3	5.6
	Spain	7.4	1.5	4.2	1.7
	Belgium	8.4	2.4	3.1	2.9
Two or more cars	France	4.7	1.3	2.8	0.6
	Spain	10.6	1.9	1.5	7.2
	Belgium	9.3	2.2	1.2	5.9
Second home	France	0.8	0.2	2.0	1.4
	Spain	4.9	2.0	2.6	0.3
	Belgium	1.0	0.4	0	0.6
Mean	France	7.1	0.6	4.3	6.9
	Spain	9.4	1.4	5.3	7.4
	Belgium	8.2	1.4	6.2	8.9

CHAPTER 6

MODE EFFECTS IN THE STANDARD EUROBAROMETER QUESTIONS

WILLEM E. SARIS AND JACQUES A. HAGENAARS

6.1 Introduction

In many studies it is suggested that a change from personal to telephone interviewing does not make much of a difference (Groves and Kahn, 1979; De Leeuw and Van der Zouwen, 1988; De Leeuw, 1992). There are, however, also studies which indicate quite large mode effects. Silberstein and Scott (1991) have discovered large mode effects in family expenditure research. Kalfs (1994) has shown for time budget research quite large differences for media use and transport between telephone interviewing and self administrated interviews. Scherpenzeel (1995) has also found large mode effects between telephone interviews and personal interviews for two topics.

These contradictory results seem to suggest that in controlled experiments small differences are found while in real life data collections where the procedures for the different modes are in some sense optimised differences can occur.

With respect to the reasons for differences in univariate distributions the research group suggested that three aspects had to be studied. The first obvious reason for differences was that the results of the telephone and personal interviews will be different due to the differences in penetration of telephones in the different countries (coverage error). A second issue concerns the effect of the organisational characteristics of the fieldwork. These activities lead to more or less participation and nonresponse.

A third component is the pure mode effect, an effect of the medium which is used in the data collection: a direct face to face interview or an interview using a mediating instrument like a telephone.

The aspect of sampling has been discussed in the first part of the book. In the previous chapter we have seen that the difference between telephone and face to face interviewing can be decomposed into three components of which the coverage error is the smallest and the organisational component the largest. It was also shown that the pure mode effects could be considerable. Therefore, in this second part of the book the main emphasis will be on mode effects.

In this chapter the mode effects for the standard Eurobarometer questions will be analysed. In chapter 9, several other questions, less frequently used in the Eurobarometers, will be studied which allows the evaluation of the mode effects and cross cultural differences for different types of questions. In between, in chapters 7 and 8, attention will be given to two special types of questions. First the effect of the mode of data collection on two open ended questions will be discussed. In chapter 8 the effect of the mode of data collection and the adjustment to the different modes of the formulation for the very commonly used Left Right scale will be scrutinised. In this chapter, a model will be formulated which can be used to describe and test mode effects for questions with precoded answers.

6.2 The latent class model applied for mode effects

In any data collection measurement errors are made. As long as these errors are similar except for random fluctuations, it is no problem to switch from one mode of data collection to another. If, however, different modes produce systematically different errors this switch is not so simply made. It has already been shown that for several questions systematic bias due to the mode of data collection has been observed. Therefore, an explanatory model for responses in surveys is presented.

For this purpose the formalisation of the latent class model developed by Lazarsfeld (1950a; 1950b) is used. Imagine the simplest case of a variable (x) with two categories, for example people who think that the own country has benefited of the EU (x=1) or not (x=2). The percentage of people of the population in each category π^x_1 and π^x_2 is by definition unknown. The only information which can be obtained is the percentage of people answering a question positively or negatively in a sample. But these questions can be formulated in many different ways, the data can be collected in many different ways, and each approach can lead to a different response distribution.

In this specific project the question remained the same, but it is asked either in a face to face interview or in a telephone interview. Using a slightly different notation of Goodman (1974a; 1974b) and Hagenaars (1990), the conditional probabilities to react positively or negatively, given the score on the variable x, can be presented in a matrix as follows

variable latent	x=1	x=2	marginal
observed			
1	π^f_{11}	π^f_{12}	π^f_1
2	π^f_{21}	π^f_{22}	π^f_2
	1.0	1.0	1.0

variable latent	x=1	x=2	marginal
observed			
1	π^t_{11}	π^t_{12}	π^t_1
2	π^t_{21}	π^t_{22}	π^t_2
	1.0	1.0	1.0

If we denote the proportion of people who have score 1 on variable x as π^x_1 and those with score 2 on variable x as π^x_2 than we can formulate that the proportion of people who say yes in the face to face study is equal to

$$\pi^f_1 = \pi^f_{11}\,\pi^x_1 + \pi^f_{12}\,\pi^x_2$$

and the proportion of people who say no is equal to

$$\pi^f_2 = \pi^f_{21}\,\pi^x_1 + \pi^f_{22}\,\pi^x_2$$

In the same way we can formulate for the telephone interviewing that the proportion of people who say yes is

$$\pi^t_1 = \pi^t_{11}\,\pi^x_1 + \pi^t_{12}\,\pi^x_2$$

and the proportion who say no is

$$\pi^t_2 = \pi^t_{21}\,\pi^x_1 + \pi^t_{22}\,\pi^x_2$$

In matrix algebra this can be simplified to

$$\pi^f = \Pi^f \pi^x \qquad (1)$$

and

$$\pi^t = \Pi^t \pi^x \qquad (2)$$

where π^f is the vector with the marginal distribution obtained by face to face interview

π^t is the vector with the marginal distribution obtained by telephone interview

π^x is the vector with the marginal distribution of x

Π^f is the response probability matrix in face to face interview given the score on x

Π^t is the response probability matrix in telephone interview given the score on x

If the probabilities that one will answer a question positively in a face to face interview (π^f_{11}, π^f_{12}) and in telephone interview (π^t_{11}, π^t_{12}), given the score on the latent variable x, are the same then the distribution of the observed distributions (π^f_1, π^f_2) and (π^t_1, π^t_2) will also be the same except for random fluctuations. If, however, the probabilities are unequal than the distribution for the different variables will also be different.

This point can be illustrated with a simple example. Imagine that π^x_1 =.9 and π^x_2 = .1 while the response probabilities are as given in table 6.1.

Table 6.1 Response probabilities

Face to face	x=1	x=2	**Telephone**	x=1	x=2
1	.8	.1	1	.9	.4
2	.2	.9	2	.1	.6

Due to this difference in tendency to say yes to the same question in personal and telephone interviews the distributions for the two variables will become different. The distribution on the face to face would be π_1 = .73 which is .8 x .9 + .1x.1 and π_2 = .27 while for telephone interviewing will be π^t_1 = .85 and π^t_2 =.15. This difference in distribution would, of course, not have occurred if the response probabilities would have been the same.

Such tendencies to prefer certain categories more in one mode than in another can, for example occur if in personal interviewing show cards are used and on the telephone the response categories are read by the interviewer in a fixed order. This is only one reason why such differences will be found. In the literature many other reasons can be found (Groves, 1989; De Leeuw, 1992).

In practice one does not know the distribution of x and the response probabilities but only the distributions for the two response variables. Schuman and Presser (1981), Billiet et al. (1986) and many others have shown how these differences in distributions can be tested using a research design with independent samples from the same population. Only under extreme experimental conditions one can use these so called “split ballot experiments” for a test on the effect of the mode of data collection. In general, also other differences such as coverage and nonresponse errors will play a role as was shown before. Another problem is that the response probabilities can not be estimated from such a design.

In case of panel studies this is possible by using the “turnover table”. This table presents the relationship between the responses collected with the different independent modes. In our example this table would look like table 6.2.

Table 6.2 The relationship between the observed variables if π_1^x = .9 and π_2^x = .1 and the response probabilities of table 6.1.

Face to face	**Telephone**		
	1	2	Total
1	.652	.078	.730
2	.198	.072	.270
Total	.85	.15	1.0

This table shows the distributions of the two variables in the marginals while the combinations of the response variables can be found in the cells of the matrix.

From the model specified before it follows that the table denoted by $\mathbf{T}^{ft}$ can be written as a function of the matrices with the response probabilities and the values of the latent variable if it can be assumed that the modes are independent of each other given the value of x and that x is stable over time. In order to do so, we first create a diagonal matrix **X** which contains on the diagonal the values of the latent variable in our example, thus the number or proportion of people in the classes x_1 and x_2:

$$\mathbf{X} = \begin{vmatrix} .9 & .0 \\ .0 & .1 \end{vmatrix} \tag{3}$$

Using this matrix the $\mathbf{T}^{ft}$ can be shown to be:

$$\mathbf{T}^{ft} = \Pi^{f}.X.\ \Pi^{t'} \tag{4}$$

So if the matrix with the proportions of people in the latent classes is pre- and post-multiplied by the two matrices representing the response probabilities, one gets table $\mathbf{T}^{ft}$. This formulation is attractive because it makes the connection between the table obtained from the panel study and the model characteristics one is really interested in. As one does not know the values of the probabilities in the two matrices $\Pi^{f}.\ \Pi^{t}$ and the matrix (**X**) the estimation of these values is the task to be done.

In this chapter, these response probabilities will be estimated in order to see whether they are different for the different modes of data collection. If they are different, one can expect differences in the distribution of the observed variables. If they are the same, no mode effect can be found given that X is the same for the two modes.

6.3 Research design

In order to test the equality of the response probabilities, data have to be collected from the same people in two different ways so that a turnover table can be constructed as indicated above. Such a panel study has been done as a continuation of the Eurobarometer study 41. In the personal interview the interviewer noted whether the people had a telephone. The households with a telephone have been contacted again within a period of one to two weeks for a second interview, this time by telephone, with a set of the most important questions of the Eurobarometer. This panel experiment has been done only in France, Belgium and Spain. These countries were selected because they had large differences in telephone penetration. In France approximately 350 people have completed a personal as well as a telephone interview, in Belgium approximately 250 and in Spain 320 people (see chapter 2 for details). Although these samples are much smaller than the original samples, it has been found that for most variables the distribution of the responses of the respondents did not deviate significantly from the responses of the original samples. This result suggests that the people who dropped out the study at the occasion of the second interview did not hold different opinions on the issues covered as the people who did not drop out. To continue this analysis, an important assumption was made:

The people who continued in this research did not differ in their response behaviour from the people who dropped out after the personal interview

While the research group believes that this assumption is not very strong and most likely true, the data nevertheless do not allow a test for this assumption.

In order to be able to test the equality of the response probabilities for the different response modes, it should be possible to estimate these parameters from the turnover tables. This can be done with the ML estimation procedure (Haberman, 1979) using the EM algorithm (Goodman, 1974a; 1974b; Hagenaars, 1993). The program LEM used in this study has been written by Vermunt (1995). The program uses turnover tables like the ones seen before as input. The user has to specify some mild restrictions on the probability matrices because otherwise the models of interest are not identified due to the fact that the number of unknown parameters is larger than the number of independent cells in the table.

The program (LEM) provides also a goodness of fit test for the whole model. The procedure will be illustrated below by an example.

The questions for which the analysis is done are a number of standard questions of the Eurobarometer which have been asked many times before and for which mode effects would be very troublesome. Therefore the following questions have been used:

1. Evaluation of membership of the EU

1.a Membership

Generally speaking, do you think that (our country's) membership of the EU is

a good thing / bad thing / good nor bad / DK/No answer

1.b Benefit

Taking everything into consideration, would you say that (our country) has on balance benefited or not from being a member of the (EU/EC) ?

benefited / not / DK/No answer

2. Satisfaction

2.a Life satisfaction

On the whole, are you very satisfied / fairly satisfied / not very satisfied / not at all satisfied with the life you lead ? Would you say you are ?

very satisfied / fairly satisfied / not very satisfied / not at all satisfied / DK/No answer

2.b Satisfaction with the way democracy works in (our country)

On the whole, are you very satisfied / fairly satisfied / not very satisfied / not at all satisfied with the way democracy works in (your country) ? Would you say you are ?

very satisfied/ fairly satisfied / not very satisfied / not at all satisfied / DK/No answer

3. Political interest

3.a Political discussion

When you get together with friends, would you say you discuss political matters frequently, occasionally, or never ?

frequently / occasionally / never / DK/No answer

3.b persuade others

When you hold a strong opinion, do you find yourself persuading your friends, relatives or fellow workers to share your views ? Does this happen ?

frequently / occasionally / never / DK/No answer

4. Media involvement

4.a Read newspapers

About how often do you read the news in daily newspapers ?

Every day / several times a week / once or twice a week / less often / never / DK/No answer

4.b Listen to radio

About how often do you listen to the news on the radio ?

Every day / several times a week / once or twice a week / less often / never / DK/No answer

4.c Watch TV

About how often do you watch the news on television ?

Every day / several times a week / once or twice a week / less often / never / DK/No answer

6.4 Results

The procedure which has been used for all questions will first be illustrated for one particular question. For this illustration the question concerning the frequency with which people try to persuade friends of political issues, was chosen. In table 6.3 the observed table from the French sample is given.[25]

Table 6.3 The table for the French sample with in the cells the frequencies of (absolute numbers) the answers for the different modes and within brackets the estimated frequencies on the basis of the latent class model with equal response probabilities.

Face to face	**Telephone**				
	Often	From time to time	Rarely	Never	Total
Often	13	14	1	1	29
	(13)	(12.5)	(2.0)	(1.5)	
From time to time	11	106	25	8	150
	(12.5)	(106)	(23.6)	(6.5)	
Rarely	3	22	24	23	72
	(2.0)	(23.5)	(24)	(18.0)	
Never	2	5	13	63	83
	(1.5)	(6.5)	(18.0)	(63)	
Total	29	147	63	95	334

On the basis of this table the response probabilities have to be estimated. This can be done with many different restrictions. The most interesting one in this case is the assumption that the response probabilities for the two modes are identical. In that case, using the model with equations (1) and (2) would mean that also the marginal distributions for the two modes have to be the same. The assumption of equal response probabilities can be specified in the model as

$$\Pi^f = \Pi^t \tag{5}$$

This assumption can in this case not be tested without further restrictions on the probabilities for identification reasons. The extra assumption made was that

$$\pi^f_{i,i+2}=\pi^f_{i+2,i} \text{ and } \pi^f_{i,i+3}=\pi^f_{i+3,i} \tag{6}$$

25 In this analysis we ignore the DK/No answer category because there are only a few cases and it complicates the analysis too much.

This assumption should also holds for the probabilities in telephone interviewing. These constraints concern probabilities which are very close to each other and also close to zero and therefore will have little effect on the fit of the model but help in the identification of the parameters.

With these restrictions and the assumptions in (5) the response probabilities have been estimated which were for both modes equal to the values in table 6.4.

Table 6.4 The estimated values for the response probabilities in France for the "Persuade" question using model (1) and (2) and assumptions (5) and (6).

Observed variable	**Latent variable**			
	Often	From time to time	Rarely	Never
Observed				
Often	.63	.04	.02	.01
From time to time	.34	.79	.10	.02
Rarely	.02	.15	.47	.01
Never	.01	.02	.41	.96
p^x	.10	.48	.27	.15

Applying equations (1) and (2) and using the results in table 6.4, one can compute the expected marginal distributions for the two modes. Using (4), the expected frequencies for table 6.3 can be obtained. The results are presented in brackets in the table. They show that in most cells the observed frequencies and expected frequencies do not deviate very much. Only in the cells (3,4) and (4,3) a larger difference emerges. This suggests a rather good fit of the model. As a formal test the likelihood ratio test is used for this purpose which gives in this case a value of $L^2 = 5.45$. With 4 degrees of freedom this test indicates that the model with the equality assumption (5) cannot be rejected. This result is rather remarkable because in the test with independent samples (chapter 5) it was found that there was a significant mode effect in France for this question. This test with panel data now suggests that the response probabilities might be equal and therefore the distributions of the observed variables will not differ more than by chance. This result is also surprising as the test, based on dependent samples, has more power than the test based on independent samples (Hagenaars, 1990). In chapter 5 this result is explained in the sense that a large part of the difference which was

detected before was due to the different characteristics of the fieldwork in the different studies and, only to a smaller extent, to the mode of data collection.

Table 6.5 The turnover table for the Belgian and Spanish sample with the cell frequencies (absolute numbers) of the answers for the different modes and within brackets the estimated frequencies on the basis of the latent class model with equal response probabilities.

			Belgium		
Face to face			**Telephone**		
	Often	From time to time	Rarely	Never	Total
Often	11	11	0	1	23
	(11.0)	(9.5)	(0.0)	(0.5)	
From time to time	8	66	17	11	102
	(9.5)	(66.0)	(19.5)	(7.0)	
Rarely	0	22	16	18	56
	(0.0)	(19.5)	(16.0)	(12.0)	
Never	0	3	6	25	34
	(0.5)	(7.0)	(12.0)	(25.0)	
Total	19	102	39	55	215

			Spain		
Face to face			**Telephone**		
	Often	From time to time	Rarely	Never	Total
Often	24	25	2	2	53
	(24.3)	(22.2)	(2.1)	(2.0)	
From time to time	20	59	26	8	113
	(22.2)	(60.1)	(22.1)	(5.6)	
Rarely	2	20	21	38	81
	(2.1)	(22.1)	(23.1)	(24.2)	
Never	2	3	13	38	56
	(2.0)	(5.6)	(24.2)	(39.2)	
Total	48	107	62	86	303

When the same model was also tested for Belgium and Spain, the likelihood ratio statistic was respectively L^2 = 13.6 and 16.9. With 4 degrees of freedom this means that the equality hypothesis has to be rejected for both countries. Table 6.5 presents the observed and expected frequencies for the two countries. Table 6.6 indicates again that the model fits the data rather well except, as in France, for the cells (3,4) and (4,3).

This finding seems to suggest that the probabilities for the categories 4 and 3 of the observed and latent variable have been constrained too much. Therefore the equality assumption (5) is now corrected by suggesting that the response probabilities π^{A}_{33} and π^{D}_{33} and π^{A}_{34} and π^{D}_{34} in each country do not have to be equal. This means that also π^{A}_{43} and π^{A}_{44} and also π^{D}_{43} and π^{B}_{44} can vary because the probabilities should add up to 1 for each column.

These two extra parameters were enough to obtain a very good fit for the model in each country. In France L^2 becomes 2.29, in Belgium 2.17 and in Spain 2.69. It follows that there are large differences between the modes in category 3 of the latent variable but not for category 4. This suggests that only the category 3 needs free parameters across modes. This turns out to be correct because the fit of the models does not change if the assumption is made that all probabilities in category 4 of the latent variable are the same for personal and telephone interviewing. The result of this analysis is therefore that people in category 3 of the latent variable behave differently when they get a personal interview or a telephone interview. The differences are indicated in table 6.6.

This table clearly indicates that there is quite a large change in response probability going from personal interviewing to telephone interviewing and that this change is in the same direction in all three countries: The probability to say “never” in a telephone interview increases considerably even though the probabilities are different in the different countries. Such a change in response probabilities can be an explanation for the significant differences which are found in the distributions of the responses in Belgium and Spain on this question for the different modes of data collection.

Table 6.6 The difference in reaction of respondents in category 3 of the latent variable in face to face (ftf) and telephone (tel.) interviews

Response categories	France		Belgium		Spain	
	ftf	tel.	ftf	tel.	ftf	tel.
Often	.02	.02	.00	.00	.00	.00
From time to time	.14	.14	.37	.37	.11	.11
Rarely	.64	.46	.54	.32	.56	.19
Never	.20	.38	.10	.32	.34	.70

This analysis was then repeated for all questions: First the model with equal response probabilities is tested. If this model fits the data, the analysis stops. If the model does not fit, a less restricted model allowing in one column differences in probabilities between the modes is used. This approach is applied for one country, and the obtained model is then also tested for the other countries. If the corrected model does not work, a better and parsimonious alternative model is tried for the other country. Generally, obtaining an identical model for each country for the same questions was regarded as the most desirable solution, taking into

account that, however, this goal was given up in order to achieve the most parsimonious model. This means that a model with more parameters was not accepted if a model with less parameters turned out to be equally good, even if it did not hold up for all countries.

The results obtained with this approach are presented in table 6.7. This table shows that for several questions the model with equal probabilities did not fit. In these cases the mode of data collection had an effect on the response probabilities and consequently on the distribution of the answers in the different modes.

Table 6.7 The fitted models for 9 different questions

Question	Country	Equality model	Necessary parameter	Size of probability ftf		tel.	
Persuade others							
	France	accepted*	π^t_{33}, π^t_{43}	.64	.20	.46	.38
	Belgium	rejected	π^t_{33}, π^t_{43}	.53	.10	.32	.32
	Spain	rejected	π^t_{33}, π^t_{43}	.56	.34	.19	.70
Political discussion							
	France, Spain	accepted					
	Belgium	rejected	π^t_{23}, π^t_{33}	.24	.76	.00	.99
Benefit of country from EU membership							
	all	accepted					
Evaluation of EU membership							
	France	rejected	π^t_{12}, π^t_{22}	.34	.42	.19	.81
	Belgium	rejected	π^t_{12}, π^t_{22}	.36	.45	.18	.82
	Spain	rejected	π^t_{12}, π^t_{22}	.34	.42	.19	.81
Newspaper							
Radio/TV	all countries	accepted					
Satisfaction with democracy							
	France	rejected	π^t_{23}, π^t_{43}	.00	.30	.30	.00
	Belgium	rejected	π^t_{33}, π^t_{43}	.41	.51	.77	.15
	Spain	accepted					
Satisfaction with life							
	France	accepted					
	Belgium	rejected	π^t_{12}, π^t_{22}	.47	.53	.18	.82
	Spain	accepted					

* Although the model fitted , the correction made a significant improvement.

For the “persuade” question a systematic pattern was found that the people in the third class out of four latent classes had a tendency to say “never” more frequently in telephone interviews than in personal interviews. This pattern was found in all three countries, and the difference in response probabilities was quite large.

A similar phenomenon was discovered for the variable “membership”. In all countries the model with equal response probabilities was rejected. The reason seems to be the same in all three countries, namely that there is a tendency for the people with a middle position to express this middle position more frequently in telephone interviews than in personal interviews. These effects are very similar in all three countries and very large.

For the variables “political discussion” and “satisfaction with life” only in Belgium the model with equal probabilities had to be rejected and the necessary changes are also considerable.

For the variable “satisfaction with democracy” in France and Belgium significant differences in response probabilities have been found, but not in Spain, and the reason for these differences are also different.

Finally for the “benefit” question and the question about the frequency of looking or listening to the news, the model assuming equal response probabilities could not be rejected. So only for these 4 questions there is no problem of unequal distributions of the variables in the different data collection modes.

6.5 Conclusion

In this analysis the mode effect was studied by specifying a latent class model and testing whether the response probabilities for the respondents in a given latent class are the same for personal and for telephone interviewing. If that were the case, no mode effects should emerge. If differences occur, mode effects will be detected in form of differences between the responses in a telephone and in a personal interview.

This analysis has clearly indicated that for several standard Eurobarometer questions differences in the response probabilities occur at least in some countries for some questions. This suggests that at least a part of the total mode effects can be explained by this factor. For some questions these effects are the same in all countries, but for other questions these effects are different for different countries. This might have to do with the specific formulation and interpretation of the labels of the categories in the different countries.

An attractive feature of this methodological approach is that the response probabilities for the different classes give an impression whether the questions are interpreted in the same way in the various countries. If the response probabilities for the same question are very different one can doubt whether the questions have the same meaning for the respondents. It is at least questionable whether the responses can be compared because the people in the different countries interpret the questions apparently in a different way. Such differences can be seen in table 6.8 for the variables “persuade” and “satisfaction with democracy” where the response probabilities are very different for the different countries. This is, however, a different

problem than the mode problem we have dealt with before. In chapter 9 this problem will be discussed in greater detail.

Returning to the issue of this chapter, one can say that this analysis has given strong evidence that the mode of data collection can cause considerable differences in response distributions. This finding suggests that correction methods should be developed in order to make the results for the different interviewing modes comparable. This topic will be discussed in the last part of the book. In the next chapters, mode effects will be studied for other questions.

CHAPTER 7

MODE EFFECTS ON OPEN-ENDED AGENDA QUESTIONS [26]

HERMANN SCHMITT, PETER SCHROTT AND MICHAELA THOMA

7.1 Introduction

It is good to know, in a number of contexts and for a variety of purposes, what the public considers the most important political issues to be. It is good for political leaders because only then are they in a position to consider peoples' concerns in what they are doing -- something absolutely essential for the functioning of representative democracy. It is good for social scientists as well because these concerns are known to be important determinants of socio-political behaviour, relevant for answers to the questions why people vote or don't, why they prefer a particular party over another, why they participate in demonstrations or block traffic, and so on.

There is a long debate in empirical social research about how to adequately assess issue preferences of the mass public. Two basic alternatives exist in the framework of survey research: "open-ended" and "closed-ended" questions. Both have their advantages and their shortcomings, and it depends on the purpose of the study which instrument to choose (Schuman and Presser, 1981). The main differences, in a nutshell, are as follows: open-ended questions are better suited to grasp the saliency dimension, that is, to establish which issues are felt to be important. Closed-ended questions are better in determining a structure in issue orientations, that is, in identifying issue dimensions and issue spaces in which parties and politicians can be placed. The problem with close-ended questions is to know which issues are of central importance to the citizenry, while the problem with open-ended questions is to code the resulting information in an intelligent and useful way (see Schwarz and Hippler, 1991, for a review on the impact of open- and closed-response formats).

The more salient an issue is, the more relevant it is for shaping political attitudes and behaviour. Any investigation into the determinants of political behaviour must therefore try to find out what people think the important issues are. Asking people "openly" is probably the best way to learn what they perceive as being most important.

To identify an appropriate way to assess what people think is one thing, to properly measure it is another. There are different survey methods on offer, and the question arises whether they produce similar or different outcomes. This analysis will concentrate on two modes of

26 The authors gratefully acknowledge criticism and advise given by the two editors of the volume and by Norbert Schwarz

surveying people, i.e. on telephone and face to face interviews. Are the answers people give to open-ended questions influenced by a particular mode of survey administration, that is, do responses to the question what the most important problems are, differ between telephone and personal interviews?

To answer this question responses to the open-ended question about the most important problem from the standard Eurobarometer are compared with responses to the same question asked in the two telephone surveys set up in the experiment. The question wording in the Eurobarometer surveys, by face to face and telephone interview, was as follows:

"Generally speaking, what is the most important problem facing (country) today?" „And what is the second most important problem facing (country) today?"

In the telephone survey the wording of the question was about the same, with only marginal changes:

„In your opinion what is the most important problem facing (country) at the moment?" „And in your opinion as well what is the second most important problem facing (country) at the moment?"[27]

7.2 Mode effects in responses to open-ended questions

Are there reasons to expect mode effects on the response to open-ended questions? There are some psychological differences between face to face and telephone interviews which might lead one to expect such differences (see Schwarz et al., 1991; for a similar discussion de Leeuw, 1992).

Interview modes differ in the presentation of stimuli. In the telephone interview situation stimuli are presented to the respondent auditory, via the "channel" telephone. This is a much poorer communication situation than an interview which is conducted face to face in the respondents private home. There, stimuli are presented both auditory and visually (by the interviewer's non-verbal behaviour and by presenting show cards etc. to the respondent).

Interview modes differ also in the time pressure they impose on respondents. Time pressure interferes with extensive recall processes and increases reliance on the first thing that comes to mind. In the telephone interview situation moments of silent reflection cannot be bridged by non-verbal communication. The degree of acceptable silence differs between the telephone interview situation and the face to face interview. Interviewers try to avoid silence in telephone interviews and time pressure here must be expected to be perceived much heavier than in face to face interviews.

The modes of data collection differ in the degree of possible interviewer impact. It is evident that interviewer characteristics are more likely to be noticed by the respondents in face to face contact while in a telephone interview interviewer characteristics can only be transmitted through paralinguistic cues and speech styles. In a face to face study, interviewers may

27 In the telephone interview this question was repeated for up to five problem mentionings. In the face to face survey only two answers were asked for and we therefore limit our analyses on these first two mentions.

convey their personal attitudes to the respondents. Through the perception of interviewer characteristics the amount of socially desirable responses may be increased, but may also serve to increase the rapport between the interviewer and the respondent.

The modes of data collection, finally, differ in their degree of perceived confidentiality. Responses to questions may be regarded as more or less confidential with respect to the interviewer, to the researcher and to other household members. It is obvious that the respondent is somewhat less "anonymous" in face to face interviews than he or she is in a telephone survey. Other things being equal, this should result in more socially desirable responses in face to face interviews.

These points have been summarised in figure 1 where a + indicates a possible effect of a mode characteristic on the results.

Figure 1 : Psychological aspects of mode of survey administration

	Face to face	**Telephone**
Perceived time pressure		+
Interviewer impact	+	
Perceived confidentiality		+

Adapted from Schwarz et al., 1991.

Some evidence exists from mode comparison studies for closed-ended questions. The general impression is that differences if they exist are small (e.g. De Leeuw and Van der Zouwen, 1988). Relatively little is known about the differences in answers to open-ended questions. Groves reported in an earlier study that higher time pressure in the telephone interview condition produced shorter answers to open-ended items compared to face to face interviews (Groves and Kahn, 1979; Groves, 1978). Especially higher income groups and younger respondents were found to give fewer mentions on the phone.

Our general expectations about mode differences -- formulated from a psychological point of view -- are unspecified with respect to the types of survey questions. How can one translate this logic into open-ended agenda questions, what can one specifically say about them? There are three more specific hypotheses that can be deduced from the above:

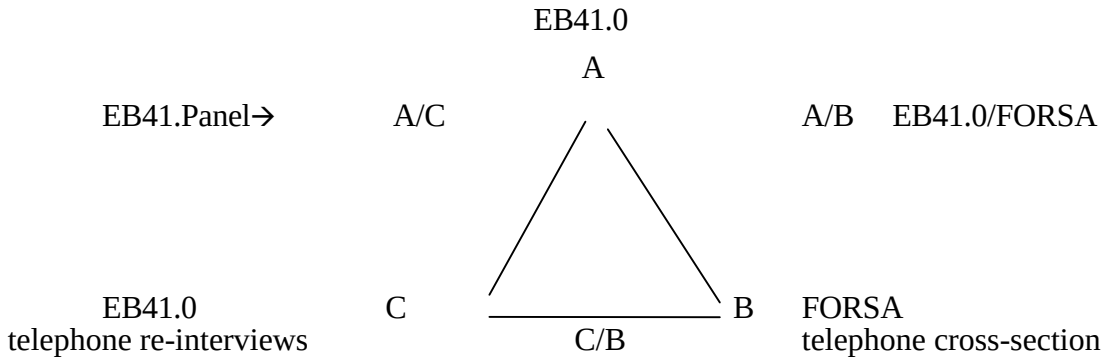
a. Elevated (perceived) time pressure in telephone interviews should result in a higher number of nonresponses as compared to face to face interviews. There is a counter hypothesis, however: greater (perceived) confidentiality might as well see to it that respondents speak out more freely and mention concerns which they would rather like to hide in face to face interviews; other things being equal, this would lead one to expect less rather than more nonresponses.

b. (a) Elevated (perceived) time pressure in telephone interviews should result in a rather limited number of problems mentioned in an open-ended agenda question, and the average thematic variety in face to face interviews is expected to be greater. It is much the same argument to say that (b) elevated (perceived) time pressure in telephone interviews should bring about more limited transcriptions, or verbatims, of answers to open-ended questions than to face to face interviews.

c. Elevated (perceived) confidentiality in telephone interviews should bring about a higher proportion of valid responses which run against social norms. The "foreigners" issue is perhaps a good example -- social norms prescribe that all humans are equal, that xenophobia is an extreme-right attitude, that extreme-right is close to fascism and holocaust and is therefore socially banned. If respondents indeed feel more anonymous in telephone interviews (and are concerned about the number of foreigners in their country), they should speak out more freely, and the foreigners issue should emerge more prominently.

7.3 Study design

In the following responses to identically worded open-ended agenda questions will be compared that were obtained in telephone and face to face interviews of the Eurobarometer experiment. In an ideal "mode comparison study", one would conduct two surveys of the same population which differ in nothing but the mode of interview (Biemer, 1988). Differences in the results could then be attributed to the mode of survey administration alone and would not be suspected to originate in other design factors. The design of the Eurobarometer experiment was somewhat more complicated. It can be visualised as follows:



One way to assess mode effects in the Eurobarometer experiment rests on a panel design where one panel wave is conducted face to face and the other by telephone (the "A/C" line in the above diagram). To do a mode comparison study within a panel design is a strategy also used by Woltman et al. (1980, on reporting of criminal victimisation) and by Hochstim (1967, on health behaviour). But even if one restricts comparison to panel respondents only, one cannot really rule out the possibility that the differences one may find and is tempted to interpret as mode effects, are caused by other factors, e.g. by learning effects known to occur in panel surveys. Such panel problems cannot disturb a comparison of results between two independent cross-sections, where one is interviewed face to face and the other by telephone

(the "A/B" line of the above triangle). But this comparison again is not without problems if the two surveys are done by different polling firms -- as in our case by INRA and FORSA. Differences one may find and is tempted to interpret as mode effects may then be as well caused by "house" or "agency" effects.[28] A conservative strategy is chosen to control for the possible complications "panel learning" and "agency effects". In this chapter, mode effects are only accepted as such if the result of the "A/C" comparison coincides with that of the "A/B" comparison; to put it in other words, a clear mode effect necessitates that both comparisons arrive at the same result (that is, that the "C/B" comparison shows no difference).

7.4 Do telephone surveys produce a higher proportion of nonresponse?

Elevated time pressure in telephone surveys should produce a higher proportion of item nonresponses, while the greater confidentiality of the telephone interview situation may make it easier to speak out -- this should cause a smaller rather than a greater proportion of nonresponses. The empirical evidence produced by the Eurobarometer experiment does not support any of these considerations (table 7.1). Nonresponses to the open-ended agenda question are somewhat more frequent, within the INRA panel, in the telephone wave than in the one conducted face to face (11 percent as compared to 8; "A/C" comparison). The situation is reversed if one looks at the cross-sections ("A/B" comparison) where telephone interviewing produces a few less missing cases than the face to face approach does (8 percent against 9).

No matter which way one looks at these figures, the differences are small. These results can best be summarised by saying that the mode of survey administration has no impact on the readiness to answer on open-ended agenda question, neither among panel respondents nor among the cross-section samples.

The panel design which was part of the Eurobarometer experiment suggests to go beyond the inspection of differences in marginal distributions. It allows to ascertain stability and change in individual response behaviour, i.e. to determine how many interviewees did or did not answer the open-ended question in both waves, how many answered in the first (face to face) and refused to answer in the second (telephone) wave, and vice versa (see table 7.2).

28 A more intensive discussion on house effects and their impact on responses to various questions asked by different survey organisations is provided by Wiseman et al. (1989).

Table 7.1 Proportion of missing cases, by mode of survey administration (column percent)

	EB41.Panel			
	First wave face to face	Second wave telephone	**EB41.0**	**FORSA**
Responses	92	89	91	92
No Responses	8	11	9	8
N	884	884	3124	1501

The usual panel picture -- aggregate stability and individual level flux -- does not emerge from this table. An overwhelming proportion of panel respondents, 85 percent altogether, does not vary between response and nonresponse from one wave to the next. Fifteen percent do change, however: nine percent responded in the face to face interview but refused to answer over the telephone (this is compatible with the time pressure proposition), while six percent answered the open-ended question in the telephone interview but not in the preceding face to face survey (this is compatible with the confidentiality proposition). Statistically, the behaviour of respondents in the first wave is significantly different from their behaviour in the second.

Table 7.2 Response and nonresponse within the EB41.Panel (total percent in parentheses)

First wave Face to face	**Second wave** Telephone		
	Responses	No responses	All
Responses	735 (83)	83 (9)	818 (92)
No Responses	52 (6)	14 (2)	66 (8)
All	787 (89)	97 (11)	884 (100)

McNemar Test: $c2 = 6,7$ $p < 0.01$

Further analyses have shown, however, that these differences are only marginally due to mode effects (results not shown). Higher educated people are somewhat less likely to change form response to nonresponse between waves. The major explanatory factor for the change seems to be national differences: Belgians are 2.4 times more likely to change than others[29]. Altogether, the statistical difference in table 7.2 should not be taken too seriously. More important is to highlight that 85 percent of the panel respondents did not change from the first to the second wave at all, and that among the few who did change, the time pressure proposition received somewhat more support than the anonymity proposition.

Overall, the difference between the two modes of data collection is remarkably small. With regard to responding or not responding to open-ended questions one finds that the mode does not make much of a difference.

7.5 Do respondents in telephone surveys mention fewer and different issues?

It was pointed out before that higher confidentiality in telephone interviews should produce results which give more emphasis to issues running against social norms. Furthermore, time pressure should make it that respondents mention fewer problems over the phone than they do in a face to face interview. In order to test these propositions, the institute-specific coding frames of INRA and of FORSA were recoded. It is always a tricky business to recode two institute-specific coding schemes into one scheme of higher generality. The problem is that coding rules (that is, decisions to assign a class of responses to one or another category) may have been different between the institutes and that therefore the "common" coding scheme does not necessarily contain identical, or at least equivalent, answers in each category. However, this reservation probably applies to some general category headings more than to more specific ones.

Ten "common" code categories were arrived at: (1) unemployment, (2) health, (3) deterioration of values, (4) foreigners, (5) political situation, (6) social problems, (7) security, (8) economy, (9) environment and (10) other. "Don't know-no answer" is an additional but technical "common" code. Topical issues like "unemployment" or "foreigners" have been assigned one or several separate codes in the coding schemes of both institutes, and we are confident that a comparison of such specific items, should lead to valid conclusions on the impact of modes of survey administration on substantive answers to open-ended questions.

In the following, one must keep in mind that the open ended question analysed throughout this paper asks for the most and the second most important problem. Both possible responses have been coded by the institutes (and recoded by the authors), and the analysis is based on the answers to both questions with respect to the most and second most important problem together. Table 7.3 gives an overview of the findings.

Of those giving at least one answer, one finds that the average face to face cross-section respondent records somewhat fewer replies (1,8 of 2 possible) than the average telephone

29 These results were obtained through a logistic regression with change between wave as the dependent and subjective social class, eduction, sex, age and nation as independent variables.

cross-section respondent does (1,9). The picture is reversed within the panel: there, the face to face wave on average produces more replies (1,8 of 2 possible) than the telephone wave (1,6).

Unemployment is by far the most frequently mentioned issue in every survey. The time pressure argument would lead one to expect this "modal issue" to be even more often mentioned in telephone surveys than in face to face interviews. However, the contrary is true. Unemployment appears considerably more often in the face to face survey (87 percent of respondents mentioning it) than in both telephone studies (63 percent each).

The "foreigners" issue should surface more often in the telephone surveys than in the face to face study as well, if for different reasons: the greater confidentiality and anonymity of a telephone interview might make it easier to speak out, to express concerns that are in conflict with social norms. Again, this expectation is not supported by the empirical evidence gathered in the framework of the Eurobarometer experiment. Telephone respondents mention the foreigners issue equally often (proportion of responses) or even somewhat less often (proportion of respondents) than respondents in the face to face interview do.

Are the differences in topical responses between face to face and telephone interviews statistically significant? One can test this again for both "angles" of the comparison -- "intra panel", and between the "cross-sections". The result is rather straightforward. Intra-panel comparisons produce only three statistically significant differences[30], indicated by a star behind the percentage in the telephone survey: unemployment and health is mentioned significantly less often, and social problems are mentioned significantly more often mentioned over the telephone than face to face. By contrast, there are eight statistically significant differences between the telephone and the face to face cross-sections; only two (foreigners and environment) are not significant.

The relative dissimilarity of responses between the two cross-sections can be interpreted as a result of a combination of an agency effect and a mode effect. The agency effect might be particularly strong in the meta-code-analysis performed here. The relative similarity of responses within the panel can be seen as a result of learning or recall processes.

On the other hand one can not completely rule out mode effects, given the large and statistically significant differences in the panel and the cross sectional study between the two modes for the issues "unemployment " and "health". Even according to the conservative test rules introduced earlier in this chapter, the results for these two issues suggest that there are mode effects present in the data. Although these differences do not lead to different conclusions with respect to the ordering of the issues with respect to importance, the effects are large enough to be taken seriously because the same can happen to subcategories of the issue unemployment which could indeed lead to differences in ordering of issues with respect to importance.

30 For the intra-panel test the Wilcoxon signed rank test was used; for the test between cross-sections the chi square test was used.

Table 7.3 Do respondents in telephone interviews mention fewer and different issues than in face to face interviews?

	EB41.0		FORSA		EB41.Panel First wave face to face		EB41.Panel Second wave telephone	
	Percentages	Cases	Responses	Cases	Percentages	Cases	Responses	Cases
Unemployment	**49**	**87**	**34** *	**63**	**50**	**89**	**39***	**63**
Health	9	16	5*	9	10	18	8*	12
Deterioration of values	9	16	2*	4	9	15	9	15
Foreigners	**7**	**12**	**6**	**10**	**6**	**11**	**6**	**10**
Political situation	6	11	11*	20	6	11	5	8
Social problems	6	10	6*	12	6	10	4*	6
Crime	6	11	11*	20	10	10	5	9
Economy	2	3	8*	16	2	3	1	2
Environment	1	2	1	2	1	1	1	1
Other	-	1	3	6	0	1	1	1
DK/No answer	5	9	13	25	4	8	21	34
Total %	100	178	100	187	100	176	100	161
N	5532	3124	2807	1501	1554	884	1422	884

7.6 Are telephone respondents less talkative?

After investigating nonresponses and the spread of topics mentioned, the final step in this analysis addresses the length of respondents' answers to the open-ended agenda question. Elevated time pressure and the poorer communication situation on the phone should lead to shorter answers; telephone respondents are expected to be less talkative.

The indicator of talkativeness chosen here is the length of the transcribed answers which respondents gave, measured in bytes (i.e. the number of letters used to write the answers down). Bytes were counted for the first and second most important problem mentioned. Respondents not naming any problem are not being included in this step of the analysis. Transcriptions of answers are only available for the cross-section surveys, and hence length comparisons are not possible within the INRA panel. This, of course, makes it impossible to clearly identify mode effects.

Empirical evidence again does not coincide with the expectations. One finds that telephone respondents are considerably more talkative than their face to face counterparts (their transcriptions are on average 10 bytes, (that is: letters), longer; see table 7.4).[31]

Table 7.4 Are telephone respondents less talkative?

	Mean	Std. dev.	N
EB41.0	31,5	26,8	2484
France	38,2	34,3	1027
Belgium *	30,9	21,0	463
Spain	24,9	16,7	994
FORSA	41,5	34,1	1172
France	33,6	28,8	377
Belgium *	55,1	37,3	359
Spain	37,2	32,4	436

* Flemish respondents only; the Walloon "verbatims" could not be matched to the standardised data set.

This overall discrepancy is not uniform across countries. Face to face respondents in France talk more than telephone interviewees while they seem to talk less in Belgium and Spain. If this is a mode effect the mode has an opposite effect in different countries. This is not impossible because there can be different habits in countries with respect to the use of the phone. It is however also possible that it is an effect of the interviewers who did the interviews in the different countries.

Given the scarcity of data in this section, it is not possible to conclusively argue that the significant difference between the telephone and face to face interviews is a mode effect. It may just as well be an agency effect due to polling practices, sampling routines or the like. However, if one controls for age, sex and social class (as a proxy for education) and simultaneously specifies interaction terms, one should be able to detect mode/agency effects holding all other factors constant. Doing such an analysis using regression[32] the FORSA + telephone (F+T) factor was still significant. This is a combination of a mode and agency

31 It is difficult to say what the interviewer's and what the respondent's contribution to this finding is. Face to face, the two may chat for a minute and the interviewer records one keyword because he has to maintain eye contact in the conversation; on the phone, the interviewer has nothing else to do but noting the given answers

32 Social class was used as a substitute for education because there was no education information available in the FORSA survey.

effect. This shows that there are considerable differences between the different studies but on the basis of this study it can not be said whether the differences are due to the mode effects or agency effects.

7.7 Conclusion

The aim of this analysis was to find out whether there are significant mode effects for open-ended questions in face to face and in telephone surveys. Given previous research findings, it was questionable whether clear and powerful mode effects would be discovered. De Leeuw (1992) in her meta analysis found only small differences between face to face and telephone surveys. Groves (1989:551) also reports that the " ...most consistent finding in studies comparing face to face and telephone interviews is the *lack* of difference in results obtained through the two modes".

Focusing on answers to open-ended questions in this chapter, the research strategy concentrated on various aspects of possible mode effects. First, the amount of nonresponses in the two survey modes was analysed. No substantial differences in the data were found.

The second step dealt with the content (or quality) of the answers. Here, much to the authors' surprise, the expectations were reversed: More diverse answers were obtained in telephone surveys than in face to face interviews. Large differences were found between the face to face study and the two telephone surveys for the issues unemployment and health. The similarity in results for the two telephone surveys seems to suggest that there are indications of pure mode effects in these cases. These differences did not change the ordering of the importance of the issues but could produce such a change quite well if more subcategories are used for the unemployment problem.

Finally, the talkativeness (and the possible obstructions for talkativeness) in respondents' behaviour was considered. Clear differences between the two polling firms were found which again ran against expectations about how response behaviour should differ between telephone and face to face interviews. Due to lack of data in this case it could not be determined whether the differences were due to agency effects or mode effects.

In general we can conclude that there is no doubt that quite large differences can be obtained when one compares the answers on open-ended questions of two companies one which is using a face to face approach and another using a telephone procedure for data collection. This study did not show clear evidence for pure mode effects except in case of the frequency with which unemployment and health are mentioned as one of the two most important problems. This happens less frequently on the telephone than in face to face surveys.

CHAPTER 8

THE LEFT-RIGHT SELF-PLACEMENT QUESTION IN FACE TO FACE AND TELEPHONE SURVEYS

HANS-DIETER KLINGEMANN

8.1 Introduction

In most countries in Western Europe a question concerning the left-right orientation has been regularly asked in political opinion surveys. In several countries the left right orientation has been used for explanation and prediction of party preference. For example, van de Eijk and Niemoller (1984) argue that more than 60% of the votes can be predicted correctly in the Netherlands using as predictor only the left-right scale. Similar studies have been done in other countries (Levitin and Miller, 1979).

The theoretical reason for this relationship is that the left-right schema has been used by citizens to orient themselves in a complex political world. This argument has been made by many people. This does not mean that the left-right question measures an ideological orientation of the respondents. Converse (1964; 1975) and Klingemann (1979) have shown that such an explanation cannot be given for the whole population; it probably holds mainly for the political elite.

Fuchs and Klingemann (1990) have made the argument that the left-right dimension plays such an important role in the Western European politics because it is a medium which can be used to connect all kinds of issues to the positions of parties. This is not only true for old issues like employment, salaries etc. but also postmaterialistic issues like environmental protection etc. In this way these concepts simplify for the citizen the complexity of the political spectrum and therefore these concepts also play an important role in the political systems in Western Europe. In their empirical study they found ample evidence for this phenomenon.

Given the relevance of the left-right schema in political science research, it is also important to know how this orientation can be measured and what happens to this measure if a different mode of data collection or a different formulation of the question is introduced. In this chapter, an experiment with the reformulation of the left-right question for telephone interviewing and the use of the standard question in face to face and telephone interviewing is reported.

The plan is as follows. First, the standard procedure for measuring the left-right orientation is discussed followed by the proposed alternative for telephone interviewing. Then the design of the experiment and the results which have been obtained are presented.

8.2 The standard measure of left right orientation

The most common way to measure left-right orientation is a question of the following format:

> In political matters people talk of "the left" and "the right".
> How would you place your views on a scale?
> (INTERVIEWER: SHOW CARD; DO NOT PROMPT. IF CONTACT HESITATES, ASK TO TRY AGAIN)
>
> Left Right
> 1 2 3 4 5 6 7 8 9 10
>
> 11 No answer/refusal
> 12 DK

The formulation given here has been used in the Eurobarometer and many other studies. Data is collected many times with this question using face to face interviews in all EU countries.

Table 8.1 gives the results of the Eurobarometer 41 study for a subset of the countries.

Table 8.1 The distribution of the responses in selected countries on the standard 10-point left-right scale in EB41.0

Countries	The categories of the standard left-right question									
	1	2	3	4	5	6	7	8	9	10
NL	3.8	6.1	13.9	14.1	22.0	15.2	14.1	8.5	0.9	1.3
W-G	1.9	3.0	9.8	15.0	26.5	20.1	9.6	8.7	3.0	2.3
E-G	5.5	5.7	15.3	16.1	33.7	13.1	5.9	3.4	0.4	0.8
DK	1.7	3.2	12.2	12.9	21.7	10.9	15.5	17.0	3.0	1.9
IRL	2.0	1.7	5.2	9.2	40.7	15.4	11.7	7.7	5.2	1.5
GR	3.4	4.9	3.0	9.6	38.7	10.6	9.4	9.1	2.7	8.6
P	4.8	4.5	12.4	18.3	27.2	15.4	7.0	5.6	2.2	2.5

This table indicates that in these countries the left-right orientation has an unimodal distribution where category 5 has the highest frequency and the frequencies go down if the

distance from 5 becomes larger. The interpretation of this result is not completely clear. The scale is constructed in such a way that the scale has no middle category so that the category 5, the category with the highest frequency, could be seen as a category indicating an opinion leaning to the left. However, a more likely interpretation is that many people chose the fifth category as a middle category and in doing so reduce the left side of the scale to 4 points while the right side has 5 points.

It cannot be excluded that a number of people chose 6 as the middle category. That would lead to the argument that categories 5 and 6 should be put together to make a middle category. Whatever one does, one thing is clear: the distribution suggests that most people are in the middle of the scale while a limited number of people has a more extreme orientation (left or right).

In table 8.2 the data are presented for the other EU-countries.

Table 8.2 The distribution of the responses in selected countries on the standard 10-point left-right scale in EB41.0

Countries	The categories of the standard left-right question									
	1	2	3	4	5	6	7	8	9	10
F	5.0	6.1	17.9	11.1	27.1	10.2	11.8	5.2	3.6	2.0
B	4.0	8.9	11.4	10.7	22.0	16.6	10.0	9.6	2.8	4.0
I	7.8	9.8	12.1	9.8	23.2	12.6	7.8	8.8	3.5	4.5
Lux	1.4	2.3	12.6	10.2	39.5	17.7	5.6	7.9	1.4	1.4
GB	3.7	2.8	11.9	10.2	32.9	14.7	10.4	9.1	1.9	2.4
ESP	8.1	7.5	16.7	15.3	23.9	9.2	6.1	4.9	2.6	5.8

For these countries 5 is also the modal category, but the distributions are not unimodal any more, i.e. the frequencies are not going down regularly when the category gets farther away from the middle. Categories 3 and sometimes 7 or 8 are higher than the surrounding categories. A possible explanation for this phenomenon is that in these countries left-wing parties exist with a large group of voters which identify with them. In such a situation one can expect a category at the left side with a relative high frequency. The same could be expected at the right side but such a phenomenon is only very weakly present in a few countries (for example Italy and Luxembourg).

The differences between the groups of countries in tables 8.1 and 8.2 cannot be an artefact of the data collection method because the same question format and data collection method has been used in all countries. On the other hand, one cannot be sure that the presented distributions are the correct distributions. It is possible that these results would look very different if another data collection mode or a different question would have been used. For

example, it is possible that too many people choose category 5 because the question was too difficult for them.

8.3 An alternative procedure

The above question using a show card is typically a procedure for a face to face interview. In telephone interviews this question with the show card cannot be used. But without the show card the question is rather complex. Therefore an alternative format has been proposed: a stepwise procedure. Such procedures have not only been suggested for this question but for several other, even simpler, questions. Groves and Kahn (1979) discuss the transformation of 7-point-category scales in what is called a stepwise procedure: first the direction is asked in three categories, and then the intensity for a specific direction. Similar experiments have been done by Sykes and Hoinville (1985) and Miller (1984). Locander and Burton (1976) and Monsees and Massey (1979) have done similar experiments for the income variable.

For the left-right scale the following stepwise procedure has been suggested:

When people talk about politics, the terms “left” and “right” are always used. We would very much like to ask you, where you put yourself, as rather “left” or rather “right?”

Rather “left”
Middle/neither nor (SPONTANEOUS)
Rather “right“
No answer/refusal
DK

Please imagine for a moment a scale, from 1 to 5, where 5 means very left and 1 not very left. Where would you put yourself?

Not very left				Very left
1	2	3	4	5

Please imagine for a moment a scale, from 1 to 5, where 5 means very right and 1 not very right. Where would you put yourself?

Not very right				Very right
1	2	3	4	5

The idea behind this formulation seems to be that people cannot respond directly to a bipolar left-right scale of 10 points but can to a 5-point scale after they have determined on what side they stand.

The problem with this question is that category 3 in both directions will probably be much larger than before, due to the stepwise procedure and the tendency of many people to choose a middle category. If this result would be obtained, it would be an artefact of the method.

On the other hand, it can also be argued that the alternative question becomes more understandable and that these results therefore are probably closer to the truth than the results obtained with the 10-point scale. It is difficult to say which argument is correct.

In this study this complex question cannot be answered What will be assessed first of all is whether it makes a difference if the standard or the stepwise version of the left right question is used, and secondly if it is really true that people cannot use a 10-point scale in telephone interviewing.

8.4 Research design

Although the alternative question was designed for telephone interviewing, it will not be used in a telephone interview because in that way two effects will be confounded: the different formulation of the question, and the different mode of data collection. In this study, the 10-point scale and the stepwise procedure will be used in two independent samples of the same populations. Such an experiment is called a split-ballot experiment (Schuman and Presser, 1981; Billiet et al., 1986). The data will be collected in a face to face-study. Besides that, a comparison will be made for the 10-point scale between a face to face study and a telephone survey. So the design of this study is as follows:

	10-point scale	Stepwise procedure
Face to face	+	+
Telephone	+	-

In the face to face interview a split-ballot experiment has been done with the question formulation. This study can show if a different distribution is obtained for the different questions. This is the only systematic difference between the two studies. So eventual differences must be due to formulation differences, except for sampling fluctuations.

The results of telephone and face to face interviews can be compared for the 10-point scale. Unfortunately, there are no repeated observations in the different modes for the 10-point scale so that coverage errors, nonresponse differences and mode differences cannot be distinguished. Only the total difference which occurs in these populations due to the difference in data collection can be compared.

The split-ballot experiment was done by INRA in the standard EB41.0 in the countries mentioned above where half of the sample got the standard question and the other half the alternative question.

The comparison between face to face and telephone interviews can be done by comparison of the INRA study with the special telephone study done by FORSA. The sample sizes in this case are also approximately the same (n=500).

8.5 Results

This section starts with the presentation of the results in tables 8.3 and 8.4 of the split-ballot experiment using the two versions of the left-right scale. Since stepwise procedure produces a scale with 11 points instead of 10, an adjustment had to be made in order to make the scales comparable. This has been done by putting categories 1 and 2 together so that both scales have, according to the interpretation given before, four left-side categories, one middle category and five right-side categories. In table 8.3 the results for the countries where the distribution was unimodal so far are compared.

Table 8.3 The differences between the stepwise procedure (2S) and the standard 10-point scale (S)

Countries		The categories of the standard left-right question									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Method											
NL	S	3.8	6.1	13.9	14.1	22.0	15.2	14.1	8.5	0.9	1.3
	2S	11.0	19.2	7.8	6.5	25.7	3.8	7.4	12.9	4.9	0.8
W-G	S	1.9	3.0	9.8	15.0	26.5	20.1	9.6	8.7	3.0	2.3
	2S	8.7	9.1	5.5	4.1	54.8	1.9	4.3	7.9	2.2	1.4
E-G	S	5.5	5.7	15.3	16.1	33.7	13.1	5.9	3.4	0.4	0.8
	2S	10.5	17.2	6.7	5.3	49.6	3.3	1.2	4.8	1.2	0.2
DK	S	1.7	3.2	12.2	12.9	21.7	10.9	15.5	17.0	3.0	1.9
	2S	6.1	12.0	6.6	6.8	27.7	2.5	10.4	17.2	5.9	2.7
IRL	S	2.0	1.7	5.2	9.2	40.7	15.4	11.7	7.7	5.2	1.5
	2S	5.3	7.8	4.8	4.8	49.3	2.2	2.8	9.5	9.5	3.9
GR	S	3.4	4.9	3.0	9.6	38.7	10.6	9.4	9.1	2.7	8.6
	2S	4.3	5.9	2.7	5.1	45.0	4.8	7.0	15.8	4.8	4.6
P	S	4.8	4.5	12.4	18.3	27.2	15.4	7.0	5.6	2.2	2.5
	2S	9.3	15.2	10.1	2.5	34.4	2.3	7.6	9.3	7.3	2.0

In each country presented in this table the differences between the two distributions for the two forms of the left -right question are highly significant. Even more so, these countries have been reported in table 8.1 as countries for which the distributions were unimodal in contrast to the countries of table 8.2 where the distributions were at least bimodal. According to table 8.3, in all countries the stepwise question procedure leads to a distribution with three peaks: the highest for the value 5 and two lower but clearly detectable ones for categories 2 and 8. Since the only difference between the two studies is the question formulation one has to conclude that these differences are due to the formulation of the question and therefore artefacts.

On the other hand, this does not mean that the correct distribution is known, as was mentioned before. But before entering this debate, also the effects which occur in the countries where already a bimodal distribution existed will be scrutinized. For these countries the results have been summarised in table 8.4.

Table 8.4 The differences between the stepwise procedure (2S) and the standard 10-point scale (S) in EU41.0 for selected countries

Countries		The categories of the standard left-right question									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Method											
F	S	5.0	6.1	17.9	11.1	27.1	10.2	11.8	5.2	3.6	2.0
	2S	15.8	12.8	9.0	2.7	29.7	2.0	5.2	12.8	6.8	3.2
B	S	4.0	8.9	11.4	10.7	22.0	16.6	10.0	9.6	2.8	4.0
	2S	8.0	9.7	6.9	1.7	50.5	3.6	3.6	7.4	4.8	3.8
I	S	7.8	9.8	12.1	9.8	23.2	12.6	7.8	8.8	3.5	4.5
	2S	17.9	10.3	4.6	2.2	31.2	1.9	4.9	10.3	10.2	6.5
Lux	S	1.4	2.3	12.6	10.2	39.5	17.7	5.6	7.9	1.4	1.4
	2S	4.7	12.2	4.3	3.9	56.7	3.1	3.5	7.5	1.2	2.8
GB	S	3.7	2.8	11.9	10.2	32.9	14.7	10.4	9.1	1.9	2.4
	2S	8.6	10.8	5.2	1.9	53.4	1.5	4.5	9.1	3.2	1.7
ESP	S	8.1	7.5	16.7	15.3	23.9	9.2	6.1	4.9	2.6	5.8
	2S	18.4	16.8	7.2	2.8	28.7	2.2	5.0	10.6	4.7	3.7

Also in this case the distribution in all countries significantly differs from each other in that the distribution as a whole shifted to the left side. Categories 1 and 2 are often more than twice as large as before, and people say less frequently that they are right-wing oriented. On

the other hand, there clearly is a peak in category 8 which was not there when using the 10-point scale. So, also in this table considerable differences between the distributions emerge which cannot be due to any other cause than the formulation of the question.

The general conclusion based on these results is that one cannot change the question on left-right orientation in the way it has been done above because the results will be absolutely incomparable. This also means that one cannot use the stepwise procedure in telephone surveys as the equivalent form for the standard 10-point scale in face to face interviewing.

These findings lead to the second question of this study: Does one have to change the formulation of the question if telephone interviewing is used?

The answer is of course 'yes' because one cannot use a show card. Without a show card, the question must be raised whether explaining the 1 to 10 scale where 1 is left and 10 is right, on the phone is enough to enable people to answer this question.

The only way to obtain an answer in this study is to look at the nonresponse and DK answers to the different questions in the different data collection modes and at the distribution of the responses again to see if large differences are found between the different modes.

Starting with the nonresponse or no answers, table 8.5 shows the results.

Table 8.5 The nonresponse and DK/No answer in the different questions and in the different modes in EB41.0 and FORSA in 12 EU countries

	EB41.0 stepwise face to face	EB 41.0 standard face to face	FORSA
DK/No answer/ Refusal	21	16	15
Answer	79	84	85
N	6704	6706	6650

In contrast to the predictions, the number of DK/No answer responses are the lowest for the standard 10-point scale in telephone interviewing. The second best is the 10-point scale in face to face interviewing, and the worst is the question which has been suggested as the presumed solution for the problems of the standard 10-point scale. More precise analysis of the responses indicates that almost all DK/No answer or refusal responses relate to the first question in the two-step procedure which compels people to make a choice between left and right.

Furthermore, the data give no indication that the standard question is more difficult for the people on the telephone than in a face to face study with a show card. It seems that this card is not needed for getting a response.

This does not mean that a change in the mode of data collection does not have an effect on the responses. These effects can be seen in tables 8.6 and 8.7.

Table 8.6 shows for the countries with a unimodal distribution in the face to face study that the change of mode also leads to differences between the distributions, but these differences are much smaller. If a test is done at a 5% level, the results in the Netherlands, East Germany and Denmark are not significant, while at the 1% level the results in Ireland and Greece are not significant, too. Larger differences occur in West Germany and Portugal. In both countries the most likely middle category 5 is increased and the end points of the scale contain more cases than before. This can indicate a response behaviour of people who have problems with the scale. Possible solutions are to mention the middle or the end points of the scale. Nevertheless, the differences are relatively small, and the pattern that the most people are in the middle and that farther away from the middle fewer people can be found still holds, except in many cases for the lowest and the highest category which suggest in general the pattern mentioned before.

Table 8.6 The differences between the response on the 10-point scale in face to face (F) and telephone interviewing (T) for the countries of table 8.1

Countries		The categories of the standard left-right question									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Mode											
NL	F	3.8	6.1	13.9	14.1	22.0	15.2	14.1	8.5	0.9	1.3
	T	2.8	3.5	15.1	14.9	26.8	12.5	14.5	5.7	1.2	3.0
W-G	F	1.9	3.0	9.8	15.0	26.5	20.1	9.6	8.7	3.0	2.3
	T	3.9	4.5	11.6	11.9	40.4	12.2	7.2	3.9	0.0	4.4
E-G	F	5.5	5.7	15.3	16.1	33.7	13.1	5.9	3.4	0.4	0.8
	T	5.5	4.8	19.0	17.1	35.6	10.1	4.0	2.4	0.0	1.5
DK	F	1.7	3.2	12.2	12.9	21.7	10.9	15.5	17.0	3.0	1.9
	T	2.5	3.6	10.3	13.0	25.5	12.0	14.3	11.7	3.2	3.9
IRL	F	2.0	1.7	5.2	9.2	40.7	15.4	11.7	7.7	5.2	1.5
	T	4.7	2.5	4.8	5.9	35.6	18.0	13.5	9.2	2.6	3.4
GR	F	3.4	4.9	3.0	9.6	38.7	10.6	9.4	9.1	2.7	8.6
	T	8.2	2.9	5.6	10.1	36.1	9.3	8.0	10.3	2.9	6.7
P	F	4.8	4.5	12.4	18.3	27.2	15.4	7.0	5.6	2.2	2.5
	T	12.1	3.2	8.7	7.1	38.8	8.1	5.8	4.5	2.2	9.4

Table 8.7 The differences between the response on the 10-point scale in face to face and telephone interviewing for the countries of table 8.2

Countries		The categories of the standard left-right question									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Mode											
F	F	5.0	6.1	17.9	11.1	27.1	10.2	11.8	5.2	3.6	2.0
	T	4.3	2.5	8.5	13.7	35.1	12.7	10.0	8.2	2.6	2.4
B	F	4.0	8.9	11.4	10.7	22.0	16.6	10.0	9.6	2.8	4.0
	T	2.2	3.6	9.2	8.0	40.5	13.7	11.8	5.1	1.4	4.6
I	F	7.8	9.8	12.1	9.8	23.2	12.6	7.8	8.8	3.5	4.5
	T	7.1	2.0	9.6	10.5	29.1	9.5	10.6	7.5	2.6	11.4
Lux	F	1.4	2.3	12.6	10.2	39.5	17.7	5.6	7.9	1.4	1.4
	T	4.4	2.0	9.9	11.3	44.7	9.1	10.1	3.3	1.2	4.1
GB	F	3.7	2.8	11.9	10.2	32.9	14.7	10.4	9.1	1.9	2.4
	T	3.0	2.9	10.1	10.6	39.0	13.5	9.4	7.1	1.6	2.9
ESP	F	8.1	7.5	16.7	15.3	23.9	9.2	6.1	4.9	2.6	5.8
	T	15.5	3.7	8.9	8.7	26.6	5.9	9.1	5.9	1.5	14.2

With respect to the group of countries which had a bimodal distribution for the 10-point scale in face to face interviews the results have been summarised in table 8.7.

Larger differences are found in this table. In this case all differences are significant except for Great Britain. In all countries the same tendency emerges which was mentioned before that the middle category 5 and/or the lowest and the highest categories have been chosen more frequently than before. As a consequence, the other categories got fewer cases, and this led to the disappearance of the bimodal feature of the distribution in 4 out of the 6 countries discussed in table 8.2. Thus the change of the mode of data collection has the opposite effect of the change in question formulation. The stepwise procedure caused an increase of bimodal or trimodal distributions while the change of mode made them disappear.

8.6 Conclusion

One has to conclude that the differences in responses to a left-right scale between telephone and face to face interviews are less than the differences for the change of formulation of the question but, nevertheless, they are large enough to cause problems with respect to the comparison of the responses across modes.

One also has to admit that it looks as if a number of people cannot cope with the 10-point scale on the telephone and opted for simple solutions like the middle category or 1 or 10. But this result can have also been caused by other reasons. For instance, it might be an effect of a

coverage error, but it can also be due to the selection which occurred by the procedures used by the different organisations. Possibly less sophisticated people are reached by telephone and they have more difficulty with this scale on the telephone. Only further research can clarify this issue.

The alternative question using a stepwise procedure does not seem to be a wise choice. It has been shown that the number of people who do not give an answer to this question is higher than for the standard question, both face to face and telephone. Besides that, this alternative formulation leads to a very different distribution of the responses than the standard question. It is not clear which distribution is the correct one, but for purposes of comparison these two questions are not equivalent. So the stepwise procedure is not recommended for telephone interviewing.

CHAPTER 9

COMPARABILITY ACROSS MODE AND COUNTRY

WILLEM E. SARIS

9.1 Introduction

Sofar only the mode effects for a limited set of standard Eurobarometer questions were discussed. In this chapter the discussion shall be extended in two ways. First of all, all questions of the panel study which can be analysed with the latent class model introduced in chapter 6 will be covered. Secondly variation in response probabilities across countries will be studied. This data set offers a unique opportunity to look at differences in response probabilities between modes of data collection and between different countries. This issue is relevant because it is doubtful whether responses can be compared if the response probabilities - or response functions, as they are also called (Saris, 1986) - are different. In that publication which concentrated on continuous data for individuals, the argument was made that it was highly questionable that responses can be compared if the response functions are different for different people. Here, two sources of variation will be in the centre of attention: the mode of data collection, which was discussed before, and the different meanings of questions in different languages.

The chapter starts with a discussion of the data and the model used for this purpose. Next, the research design will be treated and the method of data analysis will be illustrated. Finally, the results of the analysis and an explanation for the results which have been found will be presented.

9.2 The model for comparison of response probabilities

The basic model is the latent class model as used in chapter 6. In that model the distribution of the response variables for the face to face study and the telephone study can be described using the following equations:

$$\pi^f = \Pi^f \pi^x \quad (1)$$

and

$$\pi^t = \Pi^t \pi^x \quad (2)$$

where π^f is the vector with the marginal distribution of the responses in the face to face study

π^t is the vector with the marginal distribution of the responses in the telephone study

π^x is the vector with the marginal distribution of the latent variable x

Π^f is the response probability matrix for the face to face mode given the score on x

Π^t is the response probability matrix for the telephone mode given the score on x.

For more details on this model the reader should turn again to chapter 6. There, it has been indicated that from the model specified above it follows that the table denoted by T^{ft}, which gives the relationship between the responses of the the same people in a face to face and in a telephone interview, can be written as a function of the matrices with the response probabilities and the values of the latent variable. In order to do so, first a diagonal matrix **X** was created which contains on the diagonal the proportion of people (π(k) in each of the classes of the latent variable:

$$\mathbf{X} = \begin{vmatrix} \pi_1{}^x \ldots .0 \\ 0 \ldots . \pi_k{}^x \end{vmatrix} \qquad (3)$$

Using this matrix the $\mathbf{T}^{ft}$ can be shown to be:

$$\mathbf{T}^{ft} = \Pi^f.\mathbf{X}.\ \Pi^{t\prime} \qquad (4)$$

If the matrix with the number of people in the various latent classes is pre- and post-multiplied with the two matrices representing the response probabilities, then one obtains table $\mathbf{T}^{ft}$.

This formulation is attractive because it connects the table obtained from this research with the model characteristics one is interested in but does not know i.e. the response probabilities Π^f and Π^t and the matrix with the values for the latent variable **X**.

In chapter 6 tables presenting the relation between the responses in the face to face and in the telephone mode by the same people have been used to test the equality of the response probabilities in Π^f and Π^t . In this chapter this approach will be generalised in order to also test the equality of the response probabilities across countries. This analysis is just as important for crosscultural comparative research as the mode effect study for the comparison of results from different types of interviews. The reasons for the attention for cross cultural similarity of response probabilities is in fact also the same as it is for mode effects: it is questionable whether one can compare results of studies when the response probabilities for the questions of interest are not the same. In crosscultural research, usually an effort is made to keep the questions the same for the different groups. Nevertheless, it can happen that in one country the interpretation of a word is different than in another country, this way, the

same formulation can lead to different results. An indication of differences in linguistic interpretation can be obtained from the fact that the response probabilities are different for groups having the same score on the latent variable. In order to test for such equalities across countries the model needs to be extended to take into account the possibility of differences between countries.

This generalisation of equations (1) and (2) is rather simple and works in the following way:

$$\pi^{fi} = \Pi^{fi}\, \pi^{xi} \tag{5}$$

and

$$\pi^{ti} = \Pi^{ti}\, \pi^{xi} \tag{6}$$

where i indicates the ith country. From this follows as before:

$$\mathbf{T}^{fti} = \Pi^{fi}.\mathbf{X}^{i}.\, \Pi^{ti}, \tag{7}$$

This formulation suggests that for country i a table presenting the relationship between the responses in face to face and telepehone interviews are determined by the distribution on the latent variable in the population ($\mathbf{X}^{i}$) multiplied by the matrices presenting the response probabilities (Π^{fi}, Π^{ti}).

A similar equation could be constructed for each country:

$$\begin{aligned} \mathbf{T}^{fti} &= \Pi^{fi}.\mathbf{X}^{i}.\, \Pi^{ti}, \\ \mathbf{T}^{ftj} &= \Pi^{fj}.\mathbf{X}^{j}.\, \Pi^{tj}, \\ \mathbf{T}^{ftk} &= \Pi^{fk}.\mathbf{X}^{k}.\, \Pi^{tk}, \end{aligned} \tag{8}$$

Restrictions can be introduced in this model by specifying equalities between the different matrices with response probabilities.

A first possibility is that the reponse probabilities are the same for telephone and face to face studies in all countries:

$$\Pi^{fi} = \Pi^{ti} = \Pi^{fj} = \Pi^{tj} = \Pi^{fk} = \Pi^{tk} \tag{9}$$

A second possibily is to assume that the response probabilities are the same across countries but different for telephone and face to face interviewing:

$$\Pi^{fi} = \Pi^{fj} = \Pi^{fk}$$
$$\Pi^{ti} = \Pi^{tj} = \Pi^{tk} \quad (10)$$

A third possibility is that the response probabilities for face to face and telephone interviewing are the same but the probabilities across countries are not the same:

$$\Pi^{fi} = \Pi^{ti}$$
$$\Pi^{fj} = \Pi^{tj} \quad (11)$$
$$\Pi^{fk} = \Pi^{tk}$$

The first assumption is the most attractive one because comparisons are possible across studies and countries. If (9) does not hold but (10), then comparisons across countries are possible but not across the mode of data collection without correction. If (11) holds, comparisons across modes within each country are possible but comparisons across countries are not possible without further consideration. In the following, it will be shown how the data were analysed to test these restrictions.

9.3 Research design

In order to test the equality of the response probabilities data have to be collected from the same people in two different ways so that turnover tables can be constructed as indicated above. The panel experiment on which the following analysis was based was described in detail in chapters 1 and 2. France, Belgium and Spain were selected for this panel study because of the difference in telephone penetration. In France approximately 350 people have completed a personal as well as a telephone interview, in Belgium approximately 250 and in Spain 320 people. Although these samples are much smaller than the original face to face samples, it has been found that for most variables the distribution of the responses of the respondents did not deviate significantly from the responses of the original samples. This result suggests that the people who dropped out the study did not hold different opinions on the issues as the people who did not drop out. To be able to do the further analysis, an important assumption needs to be made:

> The people who participated both in the face to face and in the telephone interview did not differ in their response behaviour from the people who dropped out after the face to face interview

Although without proof, this assumption is not regarded as a very strong one and is most likely true. However, the data do not allow for a test of this assumption.

9.4 The research questions

The questions for which the analysis is done, are first of all the standard questions of the Eurobarometer which have been studied before. For details see appendix 1:

1. Evaluation of membership of the EU

1.a) Membership

Generally speaking, do you think that (our country's) membership of the EU is

a good thing /a bad thing / neither good nor bad / DK/No answer

1.b) Benefit

Taking everything into consideration, would you say that (our country) has on balance benefited or not from being a member of the (EU/EC)?

Benefited / not benefited / DK/No answer

2. Satisfaction

2.a) Satisfaction with life

On the whole, are you very satisfied/ fairly satisfied/ not very satisfied/ not at all satisfied with the life you lead? Would you say you are

very satisfied / fairly satisfied / not very satisfied / not at all satisfied? / DK/No answer

2.b) Satisfaction with democracy

On the whole, are you very satisfied/ fairly satisfied/ not very satisfied/ not at all satisfied with the way democracy works in (our country)? Would you say you are

very satisfied / fairly satisfied / not very satisfied / not at all satisfied? / DK/No answer

3. Political interest

3.a) Political discussion

When you get together with friends, would you say you discuss political matters

frequently / occasionally / never? / DK/No answer

3.b) Persuade others

When you hold a strong opinion, do you find yourself persuading your friends, relatives or fellow workers to share your views? Does this happen

frequently / occasionally / never? / DK/No answer

4. Media involvement

4.a) Newspaper

About how often do you read the news in daily newspapers?

Every day / several times a week / once or twice a week / less often / never / DK/No answer

4.b) Radio

About how often do you listen to the news on the radio?

Every day / several times a week / once or twice a week / less often / never / DK/No answer

4.c) TV

About how often do you watch the news on television?

Every day / several times a week / once or twice a week / less often / never / DK No answer

There is also a number of questions which have not been asked literally in the same way in each barometer and will not be asked in each tracking study. These questions concern specific opinions and knowledge. The content of these questions can be changed, but this format will not be changed. Therefore, it makes sense to study the effect of these formats in the different modes. Questions of this type which are scrutinized here, are:

Opinions

What is your opinion on each of the following proposals? Please tell me for each proposal, whether you are for it or against it .

READ OUT IN ROTATING ORDER

a) There should be a European Monetary Union with one single currency replacing by 1999 the (national currency) and all other national currencies of the member states of the (EC/EU).

b) The (EC/EU) member states should work towards a common defence policy.

c) Any citizen of another (EC/EU) country who resides in (your country) should have a right to vote in local elections.

d) The (EC/EU) member states should work towards a common defence policy.

e) Any citizen of another (EC/EU) country who resides in (our country) should have the right to vote in local elections.

f) Any citizen of another (EC/EU) country who resides in (our country) should have the right to vote in European elections.

g) Any citizen of another (EC/EU) country who resides in (our country) should have the right to be a candidate in local elections.

h) Any citizen of another (EC/EU) country who resides in (our country) should have the right to be a candidate in European elections.

i) The (EC/EU) should be responsible only for matters that cannot be effectively handled by national, regional and local governments.

Knowledge of White paper

Have you ever heard about the "White paper" by the European Commission in Brussels about growth, competitiveness and employment in Europe?

Yes
No
DK/No answer

Information on European Parliament

Have you recently seen or heard, in the papers, on the radio or on tv, anything about the European Parliament?

Yes
No
DK/No answer

Knowledge about next European Election

Do you know the date on which the next European election will take place in (your country), or not? (If yes) On which date?

Yes and correct date
Yes, but the date mentioned is not correct
No, does not know the date

9.5 Data analysis

The equality of the response probabilities can be tested by the turnover tables from the Eurobarometer panel. This is done like in chapter 6 (see there for details). The user

formulates a model by specifying specific restrictions on the response probability matrices for each country and for each mode. The program then estimates the response probabilities and provides a goodness of fit test for the whole model. The procedure will be illustrated below by an example. The following question is used for this illustration:

What is your opinion on each of the following proposals? Please tell me for each proposal, whether you are for it or against it?

Table 9.1 The relationships between the face to face and telephone responses in EB41.Panel for the variable "vote in local elections" in three countries

	Pro	Against	DK/ No Answer	Total
France		**Telephone**		
Face to face				
Pro	159	5	9	173
Against	7	128	4	139
DK/No answer	16	5	1	22
Total	182	138	14	334
Belgium		**Telephone**		
Face to face				
Pro	103	5	1	109
Against	3	80	1	84
DK/No answer	21	10	1	32
Total	127	95	3	225
Spain		**Telephone**		
Face to face				
Pro	175	4	14	193
Against	6	63	5	74
DK/No answer	25	8	5	38
Total	206	75	24	305

The third item presented is:

> Any citizen of another (EC/EU) country who resides in (our country) should have the right to vote in local elections.
>
> Pro / Against / DK/No answer

The tables representing the relationship between the face to face and telephone responses in the panel study for this variable in the three countries are presented in table 9.1.

This table has been used as data input to test different models in the three countries. Many different models can be formulated as was indicated before.

The simplest model is the model specified by equations (8) and (9) which means:

All response probabilities are the same across countries and modes *(Model 1)*

If this model does not fit one can move into two directions: relax the assumption of the equality of the probabilities across countries, or relax the assumption of the equality of the probabilities across modes. More problems between the modes are expected as was shown in chapter 6 where for many variables mode effects were found. Therefore model 2 is specified by (8) and (10):

All response probabilities are the same across countries but one or more probabilities can be different for the different modes *(Model 2)*

If this model does not fit, model 3 is tested as specified by (8) and (11):

All response probabilities are the same for the different modes but the response probabilities do not have to be the same from country to country. *(Model 3)*

If this model also does not fit, the only possibility is model 4:

At least some response probabilities for the different modes are different while also at least some probabilities are different across countries *(Model 4)*

The advantage of testing the models in this sequence is that one can stop when a model fits the data which saves time given the number of variables. Furthermore, these models are hierarchical so that the $\chi 2$ statistics for the different models can be subtracted from each other, and a test can be done on the improvement of the model by the additionally introduced parameters.

Model 1 is specified by the following patterns for the matrices Π^{fi} and Π^{ti}:

face to face telephone

$$\text{country 1 } \Pi^{fl} = \begin{vmatrix} fr & \pi^{f1}_{12} & \pi^{f1}_{13} \\ \pi^{f1}_{12} & fr & \pi^{f1}_{23} \\ \pi^{f1}_{31} & \pi^{f1}_{32} & fr \end{vmatrix} \qquad \Pi^{tl} = \begin{vmatrix} fr & \pi^{f1}_{12} & \pi^{f1}_{13} \\ \pi^{f1}_{12} & fr & \pi^{f1}_{23} \\ \pi^{f1}_{31} & \pi^{f1}_{32} & fr \end{vmatrix}$$

$$\text{country 2 } \Pi^{f2} = \begin{vmatrix} fr & \pi^{f1}_{12} & \pi^{f1}_{13} \\ \pi^{f1}_{12} & fr & \pi^{f1}_{23} \\ \pi^{f1}_{31} & \pi^{f1}_{32} & fr \end{vmatrix} \qquad \Pi^{t2} = \begin{vmatrix} fr & \pi^{f1}_{12} & \pi^{f1}_{13} \\ \pi^{f1}_{12} & fr & \pi^{f1}_{23} \\ \pi^{f1}_{31} & \pi^{f1}_{32} & fr \end{vmatrix}$$

$$\text{country 3 } \Pi^{f3} = \begin{vmatrix} fr & \pi^{f1}_{12} & \pi^{f1}_{13} \\ \pi^{f1}_{12} & fr & \pi^{f1}_{23} \\ \pi^{f1}_{31} & \pi^{f1}_{32} & fr \end{vmatrix} \qquad \Pi^{t3} = \begin{vmatrix} fr & \pi^{f1}_{12} & \pi^{f1}_{13} \\ \pi^{f1}_{12} & fr & \pi^{f1}_{23} \\ \pi^{f1}_{31} & \pi^{f1}_{32} & fr \end{vmatrix}$$

Where “fr” means that this parameter is free and should be equal to 1 minus the other probabilities in the column.

As can be seen; the response probabilities are the same for the different countries. Furthermore we see that also the face to face and the telephone response probabilities are assumed to be the same. Finally the elements π_{12} and π_{21} are assumed to be identical for identification reasons.

For the specific example, model 1, assuming all probabilities across modes and countries being identical, gave a $\chi 2$ statistic of 41.12 with 13 degrees of freedom which means that a test on 5% level would lead to rejection of this model. This rejection did not come as a surprise because in all three tables cell 31 is much larger than cell 13. This suggests that the people react differently in face to face than in telephone interviewing. It is hypothesised that the people who are at least weakly in favour of this issue are saying more quickly “Don’t know” or giving no answer in a face to face interview than in a telephone interview. So π_{31} should be different for telephone than for face to face interviews. Since in each column of the response matrix one parameter is free to make the probabilities add up to 1, also π_{11} can now vary between the face to face and the telephone mode, but not across countries. In this case the model is adjusted as follows:

$$\text{country 1 } \Pi^{fi} = \begin{vmatrix} fr & \pi^{f1}_{12} & \pi^{f1}_{13} \\ \pi^{f1}_{12} & fr & \pi^{f1}_{23} \\ \pi^{f1}_{31} & \pi^{f1}_{32} & fr \end{vmatrix} \qquad \Pi^{t1} = \begin{vmatrix} fr & \pi^{f1}_{12} & \pi^{f1}_{13} \\ \pi^{f1}_{12} & fr & \pi^{f1}_{23} \\ \pi^{t1}_{31} & \pi^{f1}_{32} & fr \end{vmatrix}$$

$$\text{country 2 } \Pi^{f2} = \begin{vmatrix} fr & \pi^{f1}_{12} & \pi^{f1}_{13} \\ \pi^{f1}_{12} & fr & \pi^{f1}_{23} \\ \pi^{f1}_{31} & \pi^{f1}_{32} & fr \end{vmatrix} \qquad \Pi^{t2} = \begin{vmatrix} fr & \pi^{f1}_{12} & \pi^{f1}_{13} \\ \pi^{f1}_{12} & fr & \pi^{f1}_{23} \\ \pi^{t1}_{31} & \pi^{f1}_{32} & fr \end{vmatrix}$$

$$\text{country 3 } \Pi^{f3} = \begin{vmatrix} fr & \pi^{f1}_{12} & \pi^{f1}_{13} \\ \pi^{f1}_{12} & fr & \pi^{f1}_{23} \\ \pi^{f1}_{31} & \pi^{f1}_{32} & fr \end{vmatrix} \qquad \Pi^{t3} = \begin{vmatrix} fr & \pi^{f1}_{12} & \pi^{f1}_{13} \\ \pi^{f1}_{12} & fr & \pi^{f1}_{23} \\ \pi^{t1}_{31} & \pi^{f1}_{32} & fr \end{vmatrix}$$

It can be seen that only one extra parameter is added. In all other countries the coefficients are assumed to be the same.

The fit of this model is acceptable because the $\chi 2 = 19.3$ with 12 degrees of freedom. Only one extra parameter, the difference between the telephone and face to face interviews, is sufficient to obtain a good fitting model for these data. To get a fitting model, no differences between the countries needed to be allowed for.

For illustrative purposes also the analysis for model 3 is presented where differences between countries but not between face to face and telephone interviewing are allowed for. In that case, the model is formulated as follows:

$$\text{country 1 } \Pi^{fl} = \begin{vmatrix} fr & \pi^{f1}_{12} & \pi^{f1}_{13} \\ \pi^{f1}_{12} & fr & \pi^{f1}_{23} \\ \pi^{f1}_{31} & \pi^{f1}_{32} & fr \end{vmatrix} \qquad \Pi^{tl} = \begin{vmatrix} fr & \pi^{f1}_{12} & \pi^{f1}_{13} \\ \pi^{f1}_{12} & fr & \pi^{f1}_{23} \\ \pi^{f1}_{31} & \pi^{f1}_{32} & fr \end{vmatrix}$$

$$\text{country 2 } \Pi^{f2} = \begin{vmatrix} fr & \pi^{f2}_{12} & \pi^{f2}_{13} \\ \pi^{f2}_{12} & fr & \pi^{f2}_{23} \\ \pi^{f2}_{31} & \pi^{f2}_{32} & fr \end{vmatrix} \qquad \Pi^{t2} = \begin{vmatrix} fr & \pi^{f2}_{12} & \pi^{f2}_{13} \\ \pi^{f2}_{12} & fr & \pi^{f2}_{23} \\ \pi^{f2}_{31} & \pi^{f2}_{32} & fr \end{vmatrix}$$

$$\text{country 3 } \Pi^{f3} = \begin{vmatrix} fr & \pi^{f3}_{12} & \pi^{f3}_{13} \\ \pi^{f3}_{12} & fr & \pi^{f3}_{23} \\ \pi^{f3}_{31} & \pi^{f3}_{32} & fr \end{vmatrix} \qquad \Pi^{t3} = \begin{vmatrix} fr & \pi^{f3}_{12} & \pi^{f3}_{13} \\ \pi^{f3}_{12} & fr & \pi^{f3}_{23} \\ \pi^{f3}_{31} & \pi^{f3}_{32} & fr \end{vmatrix}$$

The model specification in each country is the same with the requirement that the face to face and telephone response probabilities are the same. On the other hand, these coefficients do not have to be the same from country to country. Therefore, 15 parameters have to be estimated now and not only five like in model 1 or six like in model 2. In this specific case the fit of this model with 10 parameters more is hardly better that the fit of model 1. The $\chi 2 = 38.2$ with 3 degrees of freedom which leads to the rejection of this model. For 10 extra parameters only an improvement in $\chi 2$ of three points was achieved. Going from model 1 to model 2 only one parameter more was introduced and the reduction in $\chi 2$ was 22 points which is a large improvement. Therefore, in this case a difference in response probabilities across modes is necessary for a fitting model but not a difference in response probabilities across countries.

If model 2 and model 3 had not fitted the data, then the only possible solution would have been to allow for differences between the countries and the modes as suggested in model 4. This model is, however, only used if no other model fits.

9.6 Results

Using this procedure for the different categories of questions the results were obtained which have been summarised in table 9.2. First of all, all the ***media involvement*** questions have not been affected by the mode of data collection. The categories are relatively detailed and ask for

estimates of frequencies. In such a case it seems that the mode of data collection has no effect. It is also important that there are no differences in response probabilities across countries. So the response categories for these questions can be used for comparison across modes and across countries.

The second category of questions concerning ***political involvement*** contains one question which has the same response probabilities for the different modes and different countries, while the other question produces differences between modes and across countries. The reason for this difference is that the question on political discussion has as categories: frequently, occasionally, never and DK/No answer while the question "persuade" has as categories: often, from time to time, rarely, never, DK/No answer, that is one additional category. Comparing the meaning of the categories in the two questions one could conclude that "rarely" is the extra category and it is precisely this category which causes differences between the modes and the countries. This suggests that if the categories "rarely" and "never" are collapsed, the problems might disappear. Testing this hypothesis, it turned out that the model with equal probabilities across modes and across countries now indeed fitted to the data. This suggests that for purposes of comparison these two categories should be combined. To leave out the category "rarely" in the data collection at all is another viable option.

The third set of questions concerns ***satisfaction***. Both questions have unequal response probabilities across modes and countries. Looking at the response categories, however, the problems do not come as a surprise. The response categories for both questions are: very satisfied, fairly satisfied, not very satisfied, not at all satisfied and DK/No answer. The problems are that it can be confusing for translators and respondents how to interpret especially the labels "fairly satisfied" and "not very satisfied". In fact, logically one could argue that after the "very satisfied" category "not very satisfied" contains all other possible answers and that therefore it is not clear when to use the category of "fairly satisfied".

In the analyses it was indeed the case that people in category 3 on the latent variable had a different probability for answering fairly satisfied and not very satisfied in the different modes. But this problem may also carry over into the translation of these categories into different languages.

Checking this hypothesis, differences in the translations in the different languages were indeed discovered. In French and Dutch the translation of the labels were as follows: "very satisfied, rather satisfied, rather dissatisfied, not at all satisfied". If such differences exist in the translation between the countries, it cannot come as a surprise that also differences in the reactions of the respondents across countries are found. The differences across modes must also have to do with the problematic categories which are solved differently in face to face than in telephone research.

Interesting in this context are the findings for the questions on ***involvement in the EU***. The question on knowledge uses the same category system as the satisfaction questions: very well informed, quite well informed, not very well informed, not at all well informed and DK/No answer. So, if the above speculation is correct, this question should have the same problems as the satisfaction question. In table 9.2 we see that this is indeed the case. This strengthes the argument given before.

Table 9.2 The evaluation of difference between modes and countries for the EB41.Panel

	Equality of the parameters across countries		not across countries	
	across modes (model 1)	not across modes (model 2)	across modes (model 3)	not across modes (model 4)
Media involvement				
Radio	+			
Newspaper	+			
TV	+			
Political involvement				
Political discussion	+			
Persuade others	-	-	-	+
Satisfaction				
Life in general	-	-	-	+
Democracy in country	-	-	-	+
Involvement in EU				
Interest	-	-	-	+
Knowledge	-	-	-	+
Opinion on EU membership				
Benefit for country from EU membership	+			
Evaluation of membership for country	-	+		
Opinion on EU policies				
European Monetary Union	+			
EU defence	+			
Participation local elections	-	+		
Participation EU elections	-	+		
Candidacy local elections	-	+		
Candidacy EU elections	-	+		
Division of tasks between various levels of government	-	+		

The question on “interest in EU matters” had different labels: a great deal, to some extent, not much, not at all and DK/No answer, but the problem is comparable. The term “not much” is a negation of “much”. So if one is less than “much interested” in the EU, one could choose “not much”. But then the position of the category “to some extent” is again not clear. It could be seen as a part of the category “not much”, but that would lead to confusion. Given this situation the same problems as for the other questions were expected and were indeed found, as can be seen in table 9.2. Thus, again the translations in the different countries were checked, and also in this case the translation of the labels is different.

The next set of questions concerns the so called ***unification*** questions. It was found that for “benefit” no mode effect occurs and comparison across countries is also possible. For “membership”, on the other hand, it turned out that people had a significantly higher probability to say “good” if they were in the category “good” on the latent variable in face to face than in the telephone interviews. The difference was .92 against .80. The hypothesis that this difference existed in all three countries was not rejected. So this phenomenon seems to be a crosscultural difference between telephone and face to face interviews.

Finally seven ***opinion*** questions, all with the same format; Are you pro, against or do you have no opinion, were analysed. As can be seen in table 9.2, the first two questions concerning the introduction of the European Monetary Union and a common defence policy for the EU did not indicate any mode effect and differences across countries. On the other hand, all questions concerning the elections indicated a mode effect where in the telephone interview people with a score of “DK/No answer” on the latent variable have a higher probability to say “pro” than in the face to face interview. Besides that, the last opinion question concerning the division of tasks between local, national and EU government produced the same effect. It is difficult to explain these effects. It cannot be a general acquiescence bias (Schuman and Presser, 1981) because then it should occur for all questions. It is also not an effect of the topic because then one would have to find a different explanation for the last question. An explanation as a learning effect is also difficult because then one would also expect this outcome for all questions and not only for a limited number. Besides that, why would all people learn that they have to respond “yes” to this question instead of “no”, there is also no obvious reason for that. So for the time being one has to accept the so very systematic findings, and one will have to wait for further research to clarify this matter.

9.7 Conclusion

In this chapter the comparability of the responses across modes of data collection and across countries was studied by testing the equality of the response probabilities for the different questions across mode and countries.

Table 9.2 has shown that for 11 out of the 18 questions the responses are affected by the mode of data collection. This is a rather large number and indicates that one cannot switch between modes without having to expect differences in the results. This finding which was also corroborated in chapter 5 with a different approach, is contradicting the standard literature on mode effects (Groves and Kahn, 1979; de Leeuw and van der Zouwen, 1988; de Leeuw, 1990) but agrees with other studies which found considerable effects (Silberstein et

al., 1989; Kalfs, 1994; Scherpenzeel and Saris, 1997). The last mentioned study is the most comparable one to ours since there also panel data has been used. In panel studies the confounding factors are better controlled although one faces the additional problem of memory effect. In that study, also mode effects were found, for example, for the satisfaction variables. In a meta analysis of similar studies (Scherpenzeel, 1995) in different countries also country-specific effects have been found as reported in table 9.2.

Such country-specific differences are discovered much less frequently than mode effects. For only five questions the response probabilities across countries were different. In four questions the category labels were overlapping which led to confusion in the translations and in the responses. Therefore, it would be worthwhile to use a different categorisation in order to avoid these problems in the future:

- very satisfied,
- satisfied,
- dissatisfied,
- very dissatisfied.

It is most likely that this categorisation would lead to less problems. An unattractive feature of this scale is that since most people in Europe are satisfied, this scale is made into a two-point scale for most people.

An alternative would be :

- very satisfied,
- rather satisfied,
- little satisfied,
- not satisfied.

Both formulations will probably be comparable across modes and countries but this has to be tested. Also this classification should lead to less confusion, but more differentiation could be introduced.

So far it was shown that many differences exist between modes therefore corrections for these mode effects are necessary. In order to make this possible, response probabilities for all questions are reported in appendix 2. If the matrices are different for different countries, also a matrix is presented for each country. These matrices will be used in the last part of the book for the corrections of the distributions of the variables.

CHAPTER 10

ADJUSTMENT FOR DIFFERENCES BETWEEN FACE TO FACE AND TELEPHONE INTERVIEWS

WILLEM E. SARIS

10.1 Introduction

So far the effect of different aspects of the mode of data collection on the results has been studied. Especially, it was shown for the standard questions of the Eurobarometer that the coverage errors moving from face to face to telephone interviewing are relatively small but the mode effects and the effect of the differences in the fieldwork between research organisations can be considerable. In the last part of this book the aim is higher: the aim is to evaluate if it is possible to develop a procedure to make the results of studies done with different modes comparable? In order to do so, this chapter will discuss the statistical basis for this approach and illustrate this for the Eurobarometer experiment done. This approach cannot be directly applied on the Eurobarometer data as collected in this experiment because in the tracking studies of the European Commission a different market research company will do the field work. This requires a similar study as we report here. This chapter starts with a theoretical discussion of the approach. After that the procedures will be illustrated.

10.2 The notation and basic assumptions

In the previous chapters three kinds of problems in comparative survey research have been discussed: coverage errors, nonresponse errors and pure mode errors. The first two kinds of errors are due to a kind of process which will be called ‘selection’. The mode errors are caused by a process which will be called ‘transformation’. These processes can be formulated in very similar ways but nevertheless produce very different results. First, the selection process will be addressed.

10.2.1 Selection processes

One of the simplest selection processes is sampling. In the Eurobarometer and other survey research, people are interested in the distribution of the opinions of people in a population. For example they would like to know: How many people think that their country has

benefited or not benefited from the membership in the EU, and to what extent people have no opinion about this topic.

Normally it is assumed that in the population a frequency distribution exists for the opinion one is interested in. That this is an assumption has been elaborated by Zaller (1992). Following Converse (1964), Zaller suggests that people have no fixed opinion about many issues before the interview but create an opinion when they are asked about it. Whether one assumes the existence of an opinion or the creation of an opinion will not change the argument in this chapter.

Whatever assumption is made: The existing or created opinion for a specific question[33] will have a frequency distribution which will be denoted by **f**. Thus **f** contains three numbers for the benefit question, the sum of which gives the total number (N) of people in the population (see table 10.1). This distribution is, of course, not known. One of the purposes of the Eurobarometer studies is to estimate this distribution.

Table 10.1 A potential frequency distribution of the variable "benefit from EU membership"

Opinions	Absolute frequency **f**	Relative frequency **f**/N
Benefited	10.0 million	. 6250
Not benefited	5.0 million	. 3125
DK/No answer	1.0 million	. 0625
Total population	16.0 million	1. 0000

Research can be done in different ways, for practical reasons the population as a whole will hardly ever be used. This means that almost always a sample is drawn from the population at large. In principle the sample of size n should be chosen in such a way that the expected relative frequency distribution of the sample s ($\mathbf{f}_s$/n) is identical to the relative frequency distribution in the population (**f/**N**)**. If the sample size is n, from each class of the population frequency distribution the same proportion of cases should be drawn, namely p_s = n/N**.** In table 10.2 the example is continued with a sample of size 16.000.

33 The description also covers the ideas of Zaller (1992) who suggests that different considerations exist which lead to a response on the basis of the saliency. If we assume a specific combination of considerations as salient for a specific question, one can represent some aggregated result of these considerations as the opinion of the person at that moment of that question. This is all we need for the formulation in the chapter.

Table 10.2 A potential frequency distribution of the variable "benefit from EU membership" in the sample

Opinions	Absolute frequency in the population **f**	Expected frequency in the sample $\mathbf{f_s=p_sf}$	Relative frequency in both $\mathbf{f_s/n}$
Benefited	10.0 million	10 thousand	.6250
Not benefited	5.0 million	5 thousand	.3125
DK/No answer	1.0 million	1 thousand	.0625
Total population	16.0 million	16 thousand	1.0000

The selection process does not determine who is chosen but only how many are chosen. If the number in category k is represented by f(k) for the population and by $f_s(k)$ for the sample, we could represent the consequences of this selection process in a relationship between $\mathbf{f}_s$ and **f** as follows:

$$f_s(1) = p_s f(1)$$
$$f_s(2) = p_s f(2)$$
$$f_s(3) = p_s f(3)$$

If the probabilities are placed in a diagonal matrix, the outcome of this sampling procedure with equal probabilities can also be presented in matrix notation:

$$\begin{vmatrix} f_s(1) \\ f_s(2) \\ f_s(3) \end{vmatrix} = \begin{vmatrix} p_s & 0 & 0 \\ 0 & p_s & 0 \\ 0 & 0 & p_s \end{vmatrix} \begin{vmatrix} f(1) \\ f(2) \\ f(3) \end{vmatrix}$$

or

$$\mathbf{f}_s = \mathbf{S}_s.\mathbf{f} \qquad (1)$$

where $\mathbf{S}_s$ gives the effects of the selection mechanism. In this case this is a diagonal matrix with equal probabilities on the diagonal. It is essential that the probability of drawing a person from a class is the same for all classes of the variable.

However, such a procedure is very unlikely. There might be coverage errors or nonresponse errors as discussed before. Often such errors are related to the variables of interest. This means that for a specific variable for the members in the different groups, no equal probability exists.

The consequences of such a selection process can be presented in the same way as above, but now with unequal probabilities. For example, the occurrence of coverage errors suggests a selection process with unequal probabilities:

$$f_c(1) = p_{c1}f(1)$$

$$f_c(2) = p_{c2}f(2)$$

$$f_c(3) = p_{c3}f(3)$$

where c stands for coverage and p_{ck} is the probability to end up in $f_c(k)$ coming from f(k). The probabilities are different as a consequence of coverage error. Using matrix notation as before this reads:

$$\begin{vmatrix} f_c(1) \\ f_c(2) \\ f_c(3) \end{vmatrix} = \begin{vmatrix} p_{c1} & 0 & 0 \\ 0 & p_{c2} & 0 \\ 0 & 0 & p_{c3} \end{vmatrix} \begin{vmatrix} f(1) \\ f(2) \\ f(3) \end{vmatrix}$$

or $\mathbf{f}_c = \mathbf{S}_c.\mathbf{f}$ (2)

$\mathbf{S_c}$ is again a diagonal matrix but now with unequal values. As a consequence this matrix $\mathbf{S_c}$ produces a selection ($\mathbf{S}_c$) of the cases in the sample which is biased in some direction. In chapter 5 it was shown that this might occur, for example, if a sample is drawn from telephone owners, and the ownership of the telephone is related to the opinion on the variable of interest.

A similar problem will emerge due to nonresponse. A fieldwork organisation might use a procedure which is such that certain respondents have a higher probability to participate than others. If this selection process is related to the variable of interest bias will occur in the sample. For nonresponse the consequences of this selection process will be denoted by a matrix $\mathbf{S}_n$, and the formulation of the problem is, of course, the same as for coverage errors, i.e.:

$$\begin{vmatrix} f_n(1) \\ f_n(2) \\ f_n(3) \end{vmatrix} = \begin{vmatrix} p_{n1} & 0 & 0 \\ 0 & p_{n2} & 0 \\ 0 & 0 & p_{n3} \end{vmatrix} \begin{vmatrix} f(1) \\ f(2) \\ f(3) \end{vmatrix}$$

or $\mathbf{f}_n = \mathbf{S}_n.\mathbf{f}$ (3)

This selection process is formally analogous to the previous one. Typical for these selection processes is that people keep their score on a variable but they are selected or not in a certain process. So changes in the responses do not occur. In the case of mode effects this is not the case; therefore one can speak of transformation processes.

10.2.2 Transformation processes

The last process to be formulated is the response process. This process has been discussed in chapter 6 and chapter 9. It was suggested that people in, for example, class 1 of the opinion variable not necessarily also say 1 if they are asked for their opinion. This means that they can change their score on the variable. There is possibly a high probability that they say 1 but there is possibly also a nonzero probability that they say 2 or 3. This is not a selection process as discussed above where a person remains in the same class but is selected or not. Rather, here people can move from one class to another. This process can be formulated with a latent class model as before:

$$\begin{vmatrix} f_m(1) \\ f_m(2) \\ f_m(3) \end{vmatrix} = \begin{vmatrix} \pi_{m11} & \pi_{m12} & \pi_{m13} \\ \pi_{m21} & \pi_{m22} & \pi_{m11} \\ \pi_{m31} & \pi_{m32} & \pi_{m33} \end{vmatrix} \begin{vmatrix} f(1) \\ f(2) \\ f(3) \end{vmatrix}$$

or

$$\mathbf{f}_m = \Pi_m \mathbf{f} \qquad (4)$$

In this case **f** represents, as before, the number of people in the classes before the response is given and $\mathbf{f}_m$ the distribution of the answers if mode m is used.

The difference with the selection process is that people in, for example, class 1 have a probability π_{m11} to go to class 1 of the response variable, a probability π_{m21} to go to class 2 and a probability of π_{m31} to go to class 3. In the selection process all probabilities were zero except the probabilities in the diagonal. Therefore, the people will always keep the same score in the selection process. In the response model these probabilities are not zero, and therefore people can move to a different class then they were in before. This is typical for the transformation process.

Above the basic processes were presented which play a role in any survey research.

It should be clear that $\mathbf{f}_c$ and $\mathbf{f}_n$ cannot be observed without asking a question. Thus, also these distributions represent latent classes. Since a response process for the whole population can not be seen as realistic, combinations of the above mentioned processes need to be specified for real-life research.

10.3 Data collection and the assumption of independence

In a face to face interview the following steps are carried out:

1. a specific sample is drawn ($\mathbf{S}_s$)
2. fieldwork is done by organisation *i* leading to a specific nonresponse selection ($\mathbf{S}_{ni}$)
3. data are collected with the face to face mode of data collection (Π_f).

In the above specification a sequence of steps is identified, while in the previous section only single steps were considered. In order to make the formulation simple, one has to assume independence of the different steps.

This means that the following assumptions need to be made:

Assumption 1: The selection within the fieldwork is not different whether the whole population would have been contacted or only a sample.

Assumption 2: The response process can be described by the same response probabilities whether one is concerned with the population at large or a sample or a subsample which is willing to co-operate.

If these assumptions can be made, the resulting frequency distribution of the sequence of steps of the face to face interview ($\mathbf{f}_{ftf}$) can be described as:

$$\mathbf{f}_{ftf} = \Pi_f.\mathbf{S}_{ni}.\mathbf{S}_s.\mathbf{f} \qquad (5)$$

One can read this as follows: The resulting frequency distribution in face to face interviewing ($\mathbf{f}_{ftf}$) will be realised by the sample selection ($\mathbf{S}_s$) from the population distribution (**f**) which is again changed by the selection in the fieldwork($\mathbf{S}_{ni}$) where finally the people give their responses with a certain response probability (Π_f)

In telephone surveys the following steps are taken:

1. a sample is drawn ($\mathbf{S}_s$)
2. from this sample some people drop out because of lack of a telephone ($\mathbf{S}_c$)
3. the fieldwork causes a certain nonresponse selection ($\mathbf{S}_{nj}$) due to organisation j
4. the people answer the questions through the telephone (Π_t).

Using the assumption of independence, the resulting frequency distribution of this telephone ($\mathbf{f}_t$) interview will be:

$$\mathbf{f}_t = \Pi_t.\mathbf{S}_{nj}.\ \mathbf{S}_c\ .\mathbf{S}_s.\mathbf{f} \qquad (6)$$

In this process one additional selection step is necessary due to the fact that not all people have a telephone which might bias the results.

Finally also the panel study of the Eurobarometer experiment should be defined in the same way. This approach started as a face to face study:

1. a specific sample is drawn (S_s)
2. fieldwork is done leading to a specific nonresponse selection ($\mathbf{S}_{ni}$)
3. data are collected with a certain mode of data collection (Π_f)

The data are not used immediately but first some further steps are done in line with the telephone interviewing:

4. from this sample some people drop out because of lack of a telephone ($\mathbf{S}_c$)

5. the people are asked to participate in the panel which causes a certain nonresponse selection ($\mathbf{S}_{pi}$) due to the way organisation i works

6. the people answer the questions through the telephone in the panel (Π_{tp})

If step 3 is ignored for the moment, one can specify this process as follows:

$$\mathbf{f}_{pt} = \Pi_{tp}.\mathbf{S}_{pi}.\ \mathbf{S}_c\ .\mathbf{S}_{ni}.\mathbf{S}_s.\mathbf{f} \qquad (7)$$

In this formula the steps mentioned above can be observed: first the selection for the sample, then the selection for the face to face study, then the reduction to telephone owners, and finally the drop out in the panel. The people who are left after all these steps are asked the questions by telephone which leads to the final result denoted as $\mathbf{f}_{pt}$ for the telephone answers of the panel.

For the estimation of all effects one additional assumption is essential:

Assumption 3: The response probabilities in the panel do not differ from the probabilities in a normal telephone or face to face interview.

This assumption is less certain than the previous two assumptions because here one deals with repeated observations and the previous answer can have an effect. However, Van Meurs and Saris (1989) have shown that such effects disappear in most cases after 20 minutes in the same interview so these effects will have most certainly evaporated after one week or more. If this assumption can be made it means that:

$$\Pi_{tp} = \Pi_t \qquad (8)$$

and it follows that

$$\mathbf{f}_{pt} = \Pi_t.\mathbf{S}_{pj}.\ \mathbf{S}_c\ .\mathbf{S}_{ni}.\mathbf{S}_s.\mathbf{f} \qquad (9)$$

This case equals selection processes as already seen before except for the new selection effect ($\mathbf{S}_{pj}$) due to the use of a panel.

Finally, in chapter 6 the fact was used that for the panel the responses from the face to face interview are available. This results in:

$$\mathbf{f}_{pf} = \Pi_f.\mathbf{S}_{pj}.\ \mathbf{S}_c\ .\mathbf{S}_{ni}.\mathbf{S}_s.\mathbf{f} \qquad (10)$$

Now in a more formal way all procedures used in this study are defined. In order to give an idea of the possible consequences of the different processes in survey research, table 10.3 presents the relative frequency distributions for the "benefit" question in France as an example.

Table 10.3 The relative frequency distributions for the "benefit from the EU membership" question in France

Category	Face to face	Telephone
Benefited	39.1	45.0
Not benefited	39.4	30.4
DK/No answer	21.5	24.6
N	1000	500

The table shows clearly that the differences are considerable. It is, however, not clear where these differences come from. Therefore it will be explored in the next section whether the selection and response procedures can be estimated on the basis of the available data and the previous assumptions.

10.4 The estimation of selection and response processes

In the chapter 6 and 10 the latent class model was used to estimate the response process. This was based on the following simplification. In (9) and (10) it can be seen that the selection for the panel leads to the following frequency distribution ($\mathbf{f}_p$):

$$\mathbf{f}_p = \mathbf{S}_{pj}.\ \mathbf{S}_c\ .\mathbf{S}_{ni}.\mathbf{S}_s.\mathbf{f} \qquad (11)$$

The resulting distribution is unobserved (latent) because so far no response process is specified. Starting from here there are two modes in which people have responded: face to face and telephone. So one can write instead of (9) and (10):

$$\mathbf{f}_{pt} = \Pi_t\ \mathbf{f}_p \qquad (12)$$

and

$$\mathbf{f}_{pf} = \Pi_f\ \mathbf{f}_p \qquad (13)$$

Note that it is essential that $\mathbf{f}_p$ is the same for both modes of data collection.

It has been shown in chapters 6 and 10 that the latent class model makes it possible under certain mild assumptions to estimate the response probabilities and the distribution in the latent classes (see chapter 6 for the details).

For example for the "benefit" variable the response probabilities were the same for telephone and face to face research, and the values were as specified in table 10.4.

Table 10.4 The response probabilities for the "benefit from EU membership" question, given the score on the latent variable, estimated from EB41.Panel

Category	Benefited	Not benefited	DK/No answer
Benefited	.8508	.0159	.0147
Not benefited	.0159	.8719	.1197
DK/No answer	.1333	.1121	.8656

These response probabilities were not only the same for telephone and face to face research but also within France, Belgium and Spain. It seems that the errors made in each of the classes were approximately the same in all three countries. However, the distribution over the latent class was different as can be seen in table 10.5.

Table 10.5 The relative frequency distribution in three countries for the "benefit from EU membership" question estimated from EB41.Panel

Country	Benefited	Not benefited	DK/No answer
France	.4589	.4715	.0696
Belgium	.6163	.2511	.1326
Spain	.4676	.4923	.0402

Having shown that the panel data can be used for the estimation of the response process these results will now be used to explore the estimation of the different selection processes. In this context the third assumption plays an important role. Without this assumption no further estimation could be performed.

The coverage errors can easily be estimated so it is reasonable to start with them. $\mathbf{f}_{ftf}$ includes owners of telephones and nonowners. If the selection effect of telephone ownership ($\mathbf{S}_c$) is applied to this result, one gets:

$$\mathbf{f}_{ftf.c} = \mathbf{S}_c\mathbf{f}_{ftf} \qquad (14)$$

Since both frequency distributions can be obtained from the data, $\mathbf{S}_c$ can be obtained as well. It presents the proportions which have to be applied to move from the total sample to the sample of telephone owners. These proportions do not have to be identical for all classes. In case of the "benefit" variable the result is presented in table 10.6.

Table 10.6 The coverage errors in three countries for the variable "benefit from EU membership"in EB41.Panel

Category	France	Belgium	Spain
Benefited	.94	.86	.80
Not benefited	.96	.80	.80
DK/No answer	.91	.77	.76

These data indicate an unequal effects of penetration in the different countries. This is, however, less relevant than the possible biasing effect of the selection by telephone ownership within each country. In this specific case only in Belgium a significant difference between the different categories has been found. In the same way this selection process can be estimated for all variables and countries.

Since the response probabilities are known one can obtain the distribution of the latent opinions of the respondents f_{ni} in face to face research by applying equation (15) on the distribution of the responses in the face to face study.

$$\mathbf{f}_{ni} = \Pi_f^{-1}.\mathbf{f}_{ftf} \tag{15}$$

In the same way the distribution of the latent opinion of the telephone respondents f_{nj} can be obtained by equation (16) from the distribution of the responses in the telephone study.

$$\mathbf{f}_{nj} = \Pi_t^{-1}.\mathbf{f}_t \tag{16}$$

and for the panel study the distribution of the latent opinion (f_{np}) from the distribution of the observed responses in the panel study by applying equation (17)

$$\mathbf{f}_{np} = \Pi_t^{-1}.\mathbf{f}_{pt} \tag{17}$$

The results of these calculations for the first two equations are presented for the "benefit" variable in France in columns 3 and 4 of table 10.7. The resulting frequency distributions represent the estimated frequency distributions of the two studies corrected for mode effects.

Before, it was indicated that these frequency distributions will not be the same because they are effected by different selections processes (coverage errors and nonresponse errors). These selection processes can be written as:

$$\mathbf{f}_{ni} = \mathbf{S}_{ni}.\mathbf{S}_s.\mathbf{f} \tag{18}$$

$$\mathbf{f}_{nj} = \mathbf{S}_{nj}.\ \mathbf{S}_c\ .\mathbf{S}_s.\mathbf{f} \tag{19}$$

$$\mathbf{f}_{np} = \mathbf{S}_{pj}.\ \mathbf{S}_c\ .\mathbf{S}_{ni}.\mathbf{S}_s.\mathbf{f} \tag{20}$$

From these three equations the effects of nonresponse for the different organisations have to be estimated. Since $\mathbf{S}_c$ is also known, (20) can be used to estimate $\mathbf{S}_{pj}$. This can be done by substitution of the estimated values for $\mathbf{f}_{ni}$ from (15) in (20).

Now, the only remaining task is to estimate from (18) and (19) the selection processes specified by $\mathbf{S}_{ni}$ and $\mathbf{S}_{nj}$ for the normal face to face and telephone studies. It is, however, simple to show that these selection processes cannot be estimated separately.

These equations have the form:

$$f_{nt}(k) = p_{nt}f_s(k) \qquad \text{for } k = 1\text{- } K,\ t = 1,2 \tag{21}$$

where k is a category number, t the research organisation and s denotes the sample.

From research 2K numbers $f_{ni}(k)$ are known, but with this information 3K unknowns (2K p_{ni} and K $f_s(k)$ elements) have to be estimated. This is impossible. This also means that one can not get an estimate of f which is the distribution in the population. This means that one has to adjust the aim of the study.

A less attractive result but still very valuable is that one can get the relative size of the different errors. Since this result is also useful it will be presented here although it is not exactly what was wanted.

From (18) and (19) it follows that:

$$\mathbf{f}_{ni} = \mathbf{S}_{ni}.\ (\mathbf{S}_{nj}.\ \mathbf{S}_c\ )^{-1}\ \mathbf{f}_{nj} \tag{22}$$

which is in normal algebra:

$$f_{ni}(k) = (p_{ni}/p_{nj}.p_{c.})\ f_{nj}(k) \qquad \text{for each } k \tag{23}$$

which gives :

$$w_{in}(k) = p_{ni}/p_{nj}.p_c = f_{ni}(k)/ f_{nj}(k) \qquad \text{for each k} \qquad (24)$$

Both frequencies can be estimated if the response probabilities are known and p_c is also known. So the ratio p_{ni}/p_{nj} can also be estimated. For the "benefit" example, the results of these calculations are presented in table 10.7.

Table 10.7 The estimates of the nonresponse effects on the "benefit from EU membership" question for the two studies compared (EB41.0 and FORSA)

k	$f_{ltft}(k)$	$f_t(k)$	$f_{ni}(k)$	$f_{nj}(k)$	$w_{ij}(k)$	p_c	p_{ni}/p_{nj}
Benefited	391	450	520	449	.8635	.94	.8117
Not benefited	394	304	317	427	1.3470	.96	1.2931
DK/No answer	216	246	163	125	.7669	.91	.6969

In this case the organisation which did the face to face interviews has reached relatively many people with a negative opinion, and the other company which organised the telephone survey obtained co-operation of relatively many respondents with a positive opinion or no opinion at all (DK/No answer). More cannot be said about these differences on the basis of the data. With respect to the last point it should be made clear that in the comparison between the (non) response of the different companies the mode effects do not play a role any more because a correction was already made for this factor.

Although these results are interesting in itself they are not what was desired. So far there is no possibility to estimate the size of the errors and, therefore, no correction can be made for them. Therefore, in the next section an alternative will be formulated.

10.5 Prediction of the face to face results from telephone data

Since the estimation of all errors is not possible, the aim should be to obtain at least a procedure to predict the face to face results from the telephone data or vice versa. If this is possible one can use one mode of data collection to report about the other mode. In this way one can avoid differences in the reporting.

In case a panel study is done using the two modes it seems obvious that one can use the "turnover table" giving the relationships between the responses in the different modes for this correction. As an example table 10.8 presents such a table which has been produced with the latent class proportion of .9 and .1 and unequal response probabilities for the face to face (A) and telephone mode (B).

This table shows the distributions of the two variables in the marginals while the combinations of the values on A and B are in the cells. Within brackets column percentages are presented which could be used to compute the distribution of the variable A if the distribution of variable B is obtained. With the row percentages of the table one could create the distribution of the variables B from the distribution of the variable A. This result seems to suggest that this table can be used to estimate the distribution of A from B or the distribution of B from A.

Table 10.8 The relationship between responses in A and B if π_1^x = .9 and π_2^x = .1

		Variable B			
		1		2	Total
Variable A					
	1	.652	(.767)	.078	(.52) .730
	2	.198	(.232)	.072	(.48) .270
	Total	.85	(1.000)	.150	1.000 1.000

There are two objections against this idea. The first concerns possible changes in the latent classes. It is indeed true for the given data that the turnover table can be used, but if one would like to use the same turnover table which has been obtained at some point in time at a different point in time this procedure is quite doubtful unless the distribution of the latent variable x has not changed. If the distribution of this variable has changed, one should use a different turnover table even if the response probabilities remained exactly the same.

This point will be illustrated by an example. Imagine that the only difference with the previous example is that the people have changed their opinions. Now π_1^x = .7 and π_2^x = .3 and not 9 and .1 as before, while the response probabilities remain the same as before. Then quite a different turnover table is obtained (Hagenaars, 1994) which is shown below.

Table 10.9 The relationship between responses in mode A and B if A and B if π_1^x = .7 and π_2^x = .3

		Variable B				
		1		2		Total
Variable A	1	.515	(.69)	.074	(.31)	.590
	2	.235	(.31)	.166	(.69)	.410
	Total	.750	(1.00)	.240	1.00	1.00

This table shows that the probabilities which should be used in this table to calculate the distribution of A from the distribution of B are very different from the previous table even though the only difference is the distribution of the latent variable. This means that this table cannot be used for these calculations because this transformation is needed at different points in time, and at each occasion one can expect changes in the opinion. So equality of opinion cannot be assumed.

The second objection is that using this table corrects only pure mode effects, and it was shown in chapter 5 that the nonresponse effects are often at least as large. But non response effects are ignored in this approach. Therefore, one has to use a more complex approach.

Although the simple estimation procedure using the turnover table is not possible, the turnover table obtainable by panel data is nevertheless useful because it can be used to estimate the response probabilities. If these probabilities remain stable, which is much more likely than the stability of the distribution of the opinion, an estimate of the distribution of the latent variable from the distribution of the observed variables is possible. Combining these results with the results of the previous section, a correction procedure can be formulated.

This can be done starting with equation (6). From (6) follows:

$$\mathbf{S}_s.\mathbf{f} = (\Pi_t.\mathbf{S}_{nj}.\ \mathbf{S}_c)^{-1}.\ \mathbf{f}_t \qquad (26)$$

and substitution of this result in (5) gives

$$\mathbf{f}_{ftf} = \Pi_{ftf}.\mathbf{S}_{ni}.\ (\Pi_t.\mathbf{S}_{nj}.\ \mathbf{S}_c)^{-1}.\mathbf{f}_t \qquad (27)$$

which is the same as

$$\mathbf{f}_{ftf} = \Pi_{ftf}.\mathbf{S}_{ni}.\ (\mathbf{S}_{nj}.\ \mathbf{S}_c)^{-1}\ \Pi_t^{-1}.\mathbf{f}_t \qquad (28)$$

and simplifies to

$$\mathbf{f}_{ftf} = \Pi_{ftf} \,.\, \mathbf{W}_{ij} \,.\, \Pi_t^{-1} . \mathbf{f}_t \qquad (29)$$

where **W**ij is a diagonal matrix with as elements on the diagonal the values w_{ij} which represent the relative effects of the different organisations on nonresponse and response in the different categories of the variable (including the coverage error). In the last section it was shown that these coefficients can be estimated (23).

Since the response probabilities and the weights are known, this equation can be used for estimating the face to face results from the telephone results even if the studies are done by different companies. For the "benefit" variable the calculations are illustrated in table 10.10.

Table 10.10 The estimation for France of the face to face results from the telephone data for the variable "benefit from EU membership"

Categories	$\mathbf{f}_t$	$\mathbf{f}_{nj}$	$\mathbf{w}_{ij}$	$\mathbf{f}_{nj}$	$\mathbf{f}_{ftf}$
Benefited	450	520	0.8633	449	391
Not benefited	304	317	1.3470	427	394
DK/No answer	246	163	0.7669	125	216

According to equation (29), first $\mathbf{f}_t$ is corrected for mode effects using Π_t^{-1} to obtain the distributions of the latent variable for the telephone survey. Next the nonresponse and coverage errors are corrected using ($\mathbf{W}_{ij}$) so that the latent variable for the face to face survey is obtained. Finally, the results have been made comparable by applying the mode error Π_t of the face to face study on this latent variable in order to get an estimate of the frequency distribution of the face to face study ($\mathbf{f}_{ftf}$).

In this case, it should not come as a surprise that the results are exactly correct because all estimates are based on the same data and determined by these data. The real test can only be done with new data where the response probabilities of this study are used and the nonresponse weights are obtained from a comparison of two companies who are doing the standard Eurobarometer study and the tracking study. However, this theoretical analysis shows that a prediction from the telephone data to the face to face data is possible.

10.6 Conclusion

In this chapter, first the consequences of research designs for response distributions were formally defined. In doing so it was shown that differences in results can come from selection processes like sampling, coverage errors, nonresponse errors and from transformation processes like response processes.

Next, an effort was made to estimate the potentially biasing factors which turned out not to be completely possible in this experimental design. The response probabilities could be estimated, as could the coverage errors, but the nonresponse selection process for the two

different modes could not be estimated separately. Only a ratio of the effects of the two procedures could be assessed.

Furthermore it was shown that the possibility to estimate the response probabilities, the coverage errors and the ratios of the nonresponse errors is enough to estimate the face to face frequency distribution from the distribution in the telephone survey.

It should be remarked, however, that for this the weights (ratios) have to be estimated for all variables separately because they can be different for all variables, as was the case for the response probabilities.

Furthermore, it is also required that the procedures of the research companies doing the research are not changing. If a change happens, the correction factors will probably also change, especially those factors which correct for nonresponse error.

CHAPTER 11

SUMMARY AND DISCUSSION

WILLEM E. SARIS AND MAX KAASE

11.1 Introduction

In this book a series of studies has been reported related to the effects of different modes of survey data collection in the social sciences. The methodological study covered in this book was triggered by a decision of the European Commission in favour of a partial shift of the Eurobarometer from face to face data collection to telephone data collection. Since this switch can cause a variety of problems, the consequences of this change need to be explored. After the explanation of the design of the study, a detailed description of the data collection methods used and an overview of the different problems, in this chapter the results of the analysis will be summarised. After that, some practical conclusions will be drawn, and various scientific issues will be addressed in the concluding paragraphs.

11.2 The design of the methodological study

Given the relevance of the Eurobarometer data for all those interested in the development of political orientations in Europe and for social science research in general, the changes which were expected to occur in the Eurobarometer data because of the change in data collection modes was enough reason to suggest that methodological research should be done to evaluate the consequences of that change. The argument in favour of such research was based on prior knowledge on mode effects. For example, Groves (1989) gave nine reasons why one should expect differences between face to face and telephone data collection. They can be condensed as follows:

1. The coverage of the population will be different for face to face interviews and telephone interviews, since those people who do not have a telephone will not be representative of the general population from which the sample should be drawn. In Europe, this difference can be substantial because in some areas and countries household telephone density is close to 100% while in other areas and countries the coverage is closer to 50%. In the USA it has been found that this coverage error leads to considerable differences in responses on several dimensions (Groves and Kahn, 1979; Cannel et al., 1987).
2. The field work of the organisations doing the surveys can and usually will be quite different with respect to the interviewers used, their training and supervision, the number of times that a respondent is contacted, and the rules by which a refusal is accepted.

Differences in these management aspects will lead to differences in nonresponse and consequently to differences in findings.

3. The mode of data collection itself can also lead to different results. It is possible that people react differently to the same question in a telephone interview and in a face to face interview. For instance, it has been found that open-ended questions result in more elaborated answers in face to face interviews than in telephone interviews. Also, more acquiescence and an extremeness bias might be expected (Groves, 1989). However, the general picture is that these mode effects, after correcting for all other factors, are rather small (De Leeuw and van der Zouwen, 1988).

Also, mode-connected effects are possible, that is effects which might occur due to the fact that changes in the approach to the respondent are necessary depending on the mode of interview, and that these changes will matter. For example, the use of show cards is not possible in telephone interviews, and as a consequence the procedure for complex questions has to be adjusted. In order to cope with this problem, in telephone surveys, commonly a two-step approach is used where first a small number of crude categories is presented which are later split up into more differentiated ones. The idea here is to obtain the same kind of precision by telephone as in personal interviews where show cards are used to present ten or so categories at the same time. These mode-related changes in the questionnaire can lead to substantial discrepancies in the results, as has been shown by Groves and Kahn (1979), Miller (1984), and Monsees and Massey (1979).

This brief overview points to the main reasons why a change from face to face to telephone interviews will most likely lead to different results. As indicated in chapter 2, one can expect that the total difference (T) between face to face and telephone interview responses in percentages or in mean score will be equal to the difference due to coverage (C) plus the difference due to difference in nonresponse (N) plus the difference due to the mode of data collection (M):

$$T = C + N + M \quad (1)$$

Given the possible confusion due to discrepancies in results, notably the issue of errors in both procedures needs to be addressed more precisely, but also ways should be found to adjust the findings in such a way that the results become comparable.

In order to help with the methodological study, the Berlin-based FORSA research institute offered to collect data for a limited set of questions in all countries through telephone interviewing, while at about the same time the INRA institute conducted the standard Eurobarometer 41 face to face. This approach is a good simulation of the future situation in Europe when two studies will be done on identical topics at the same time by different survey organisations, each using a different data collection mode. The design used here permits to estimate the total difference in responses for two specific organisations (T).

There is, however, one major weak point here. The problem is that there are too many factors creating differences between the various approaches, and that one therefore cannot determine precisely which one causes these differences. Anticipating this lack of strength in the design, a panel element was included in this study.

Here, the respondents were first confronted in a face to face interview with the normal Eurobarometer questionnaire. In addition, they were asked whether they had a telephone and were willing to answer some questions some time later through the telephone. If they agreed, they were called back after about a week to respond to a small number of questions already put to them before in the Eurobarometer. This panel design offers better insights into the effects of the two different sources of error, as follows.

First, when the telephone owners and non-owners are compared, an estimate can be obtained of the effect of telephone ownership on the distribution of responses in the panel to the relevant variables. In this comparison no other variables intervene because the same people are studied and all questions are presented in a face to face interview. So the only possible explanation for differences is telephone ownership, and thus a good estimate of the coverage error (C) which will occur, is provided.

A second effect that can be studied with this design relates to the mode of data collection (M) since one can compare the answers of the respondents to the same questions in the personal interview and in the telephone interview. This evidence is not so strong as in the case of the comparison of telephone owners and non-owners because there are other factors besides the mode effects which can come into play.

This design does not allow for an independent estimate of the effects of the fieldwork organisation on the nonresponses (N), but one can at least deduce this effect. The direct comparison of personal interviews with telephone interviews gives an estimate of T. Using the panel design, C and M can be assessed. Using the combination of the two designs, the effect of the difference in nonresponse due to different organisational procedures will be:

$$N = T - C - M \qquad (2)$$

It should be recalled here that the coverage error (C) is an estimate which for the largest part is independent of the organisation that did the research because the effect is determined by the difference between owners and non-owners in the population. This difference will only minimally be influenced by the specific procedure used for data collection, as long as this procedure is not completely flawed.

The same point can be made for the estimate of the mode effect M as was argued above. On the other hand, the estimates for nonresponse (N) and for the total difference (T) are clearly influenced by the organisations which perform the studies. The total difference varies directly with the difference in nonresponse which is produced by the two organisations in question. So general statements are difficult to make about these two components although they can be properly assessed for a specific case.

Furthermore, the estimates of the coverage error and of the mode effect can also vary with the topics being addressed. Telephone non-owners can differ in their opinions on certain questions, and this will lead to differential effects although for other questions the differences can be very small. In the literature, some questions have been mentioned to be more effected than others, like open-ended questions, questions placing a heavy cognitive burden on respondents such as long questions or questions with a large number of categories. The same holds true for questions which are normally asked with a show card, a procedure presently not

available in telephone interviewing. Given the effects of the type of question asked, the different types were studied separately. For an overview of the questions we refer to the questionnaires in Chapter 1.

11.3 Results

In this section the results of this study will be summarised, following most of the time the sequence in which the results have been presented in the book.

11.3.1 Sample differences

In chapter 2 an overview has been given of the data collection procedures of the two survey organisations which collected the standard Eurobarometer data by face to face interviews and by telephone interviews. It has been shown that the procedures used were different on several points as can be seen in table 11.1.

Chapter 3 reported an effort to make the samples as comparable as possible by weighting on the basis of variables for which information about the population distributions is available. It has been shown that the weighting procedures could not reduce the serious differences which existed between the two studies on several variables. It seems that the correlation between the weighting variables, the variables which cause the problems and the variables of interest are not strong enough to produce an acceptable level of adjustment.

In chapter 4 the coverage error was studied in detail. It turned out that the group of respondents without a telephone can be very different from the group with a telephone. It must be understood, though, that this does not necessarily bias the results of telephone studies very much, especially when the size of the group is rather small.

In order to study this phenomenon, chapter 5 presented an effort to estimate the size of the different component of the total differences for the three participating countries. The result of this analysis has been reprinted in table 11.2. It should be noted that since the estimates of the

different effects are based on calculations over all categories of variables, equation 1 does not hold anymore (this equality holds for each category but not necessarily for the sum ignoring the signs).

When one looks at the size of the effects, a clear rank order can be established. Averaged across the three countries, the coverage differences rank lowest with a mean of 1.1 percentage points over all questions and countries. Mode differences are remarkably larger with an average score of 5.2 percentage points. However, it cannot be concluded that this is necessarily due to the telephone interviews. It may be that respondents produce more random answers in telephone interviews because the time pressure is stronger and they are not supported by visual aids. On the other hand, interviewers in telephone interviews are more controlled which means that they ask the questions more precisely in the way expected than in the uncontrolled face to face interview situation. Clearly, the largest differences come from

the black box of the fieldwork of the two survey organisations. The mean differences over topics and countries is 7.7. percentage points.

Table 11.1 Summary of the sampling methods

	EB41.0	**EB41.Panel**	**FORSA**
Type	face to face	telephone	telephone
Fieldwork	April 4th - May 6th	April 5th - April 30th	April 28th - June 3rd
Countries	12 EU member states	France, Belgium, Spain	12 EU member states
Completion rate	EU: 44,7%	37,6% (% of eligible households)	EU: 43,4%
Sample frame	- Census enumeration units (or otherwise)	- respondents of EB41.0 with a telephone who have given their number	Telephone directories
Selection Method	- more than 100 sampling units per country are randomly chosen as start address - a random increment provides up to 10 addresses - one person/per household selected by next birthday or Kish method or an other	- all possible respondents are contacted - controlled by Age, Sex, Occupation and Subjective Social Class	- From 10 to 22 'provinces' per country samples are drawn according to the size of the province's population - one person/per household selected by next birthday method
Interviewers' testing and supervision	INRA's national associates are responsible	INRA central - computerised dialling	FORSA central - tests in advance - computerised dialling
Call backs	2 revisits	8 call backs	12 call backs
Refusals	no refusal reversion	no refusal reversion	no refusal reversion
Substitution	random walk	no substitution	by random number

Table 11.2 A summary of all one-directional differences in three countries

		Total (T)	Coverage (C)	Mode (M)	Organisation (N)
Satisfaction with life	France	10.9	0.3	5.3	9.3
	Spain	21.9	0.9	5.0	21.7
	Belgium	4.6	2.0	16.6	12.7
Satisfaction with democracy	France	4.9	0.3	9.3	10.4
	Spain	8.8	1.7	5.4	9.4
	Belgium	3.4	0.8	10.8	13.9
Persuade others	France	17.1	0.6	3.6	15.8
	Spain	11.6	0.8	9.9	3.6
	Belgium	7.8	1.6	10.9	8.8
Political discussion	France	1.7	0.7	3.1	4.2
	Spain	6.3	2.1	2.2	3.9
	Belgium	10.4	1.7	9.1	16.8
News on TV	France	6.6	0.9	0.9	4.7
	Spain	3.8	1.1	4.7	8.4
	Belgium	3.8	0.7	2.3	3.7
News daily papers	France	7.8	0.6	2.6	10.4
	Spain	8.1	3.4	5.2	10.2
	Belgium	6.9	1.4	6.9	6.4
News on radio	France	13.8	0.9	6.9	7.6
	Spain	11.2	1.0	6.7	7.8
	Belgium	13.3	1.0	2.7	10.5
Interest in European politics	France	10.5	1.0	7.4	2.9
	Spain	13.8	1.6	8.6	6.2
	Belgium	9.3	2.1	2.4	7.3
Level of EU informedness	France	8.4	0.7	6.1	5.7
	Spain	11.1	1.4	12.0	6.1
	Belgium	11.0	0.7	15.1	9.3
Membership in EU	France	6.7	0.5	6.0	12.3
	Spain	7.0	0.8	6.3	9.9
	Belgium	16.6	2.0	7.0	20.2
Benefit from EU membership	France	7.7	0.6	1.3	7.1
	Spain	14.0	0.7	4.3	13.6
	Belgium	15.1	2.4	3.9	10.9
Colour TV	France	4.1	0	2.6	6.7
	Spain	0.7	0.5	1.0	2.2
	Belgium	2.5	0.2	1.7	4.4
PC	France	0.2	0.5	5.3	5.6
	Spain	7.4	1.5	4.2	1.7
	Belgium	8.4	2.4	3.1	2.9
Two or more cars	France	4.7	1.3	2.8	0.6
	Spain	10.6	1.9	1.5	7.2
	Belgium	9.3	2.2	1.2	5.9
Second home	France	0.8	0.2	2.0	1.4
	Spain	4.9	2.0	2.6	0.3
	Belgium	1.0	0.4	0	0.6
Mean	France	7.1	0.6	4.3	6.9
	Spain	9.4	1.4	5.3	7.4
	Belgium	8.2	1.4	6.2	8.9
Average across three countries		8.2	1.1	5.2	7.7

These effects differ from question group to question group. This is not surprising because the strength of the effects is always dependent on the strength of the relationship between the error source and the substantive type of variable, and this differs from topic to topic. Nevertheless, it is clear that the coverage error is the smallest problem and that the two other factors can produce quite large differences between studies done with different modes of data collection or by different organisations. In general the effects are so large that without correction the results cannot be compared. Therefore, more attention will be devoted to mode effects as the second largest source of differences. Unfortunately nothing more can be said about the organisational differences than what has already been remarked in chapter 2. As a consequence, this chapter will now concentrate on mode effects and on possible correction for differences between studies in general.

11.3.2 The mode effects

Given the considerable contribution of the pure mode effects to the total differences between the results obtained with the two data collection methods, a more detailed analysis of the pure mode effects was conducted on the basis of the panel study.

The mode effects are different for different types of questions, suggesting a separate look at open-ended questions, simple closed questions and complex closed questions.

Starting with the open-ended question, the study reported in chapter 7 concentrated on an agenda question asking for the two most important problems for the own country and for Europe. The study evaluated three aspects: the mode effect on nonresponse, frequency with which the different problems were mentioned, and the amount of information provided by the respondents.

First, the amount of nonresponses in the two survey modes was analysed. No substantial differences in the data were found. The second step dealt with the content (or quality) of the answers. Here, much to the authors surprise the expectations were reversed: more diverse answers were obtained in telephone surveys than in face to face interviews. Large differences were found for the issues of employment and health. The similarity in results for the two telephone surveys seems to suggest that there are indications of pure mode effects in these cases. These differences did not change the ordering of the importance of the issues but could produce such a change quite well if more subcategories are used for the unemployment problem.

Finally, the talkativeness (and the possible obstructions for talkativeness) in respondent behaviour was considered. Clear differences between the two polling firms were found. Due to the lack of data it could not be determined whether the differences were due to agency effects or mode effects.

Among the complex close-ended questions, the left-right orientation has been highlighted in chapter 8. Normally in face to face research a show card is used which presents the 10 categories ranging from the extreme left to the extreme right. Some researchers has suggested that in telephone interviewing this approach is not feasible, given the complexity of the question and the large number of response categories. This has lead to the idea to apply a two-

step procedure. Fortunately, in this study a comparison could be made between the standard question and the two step procedure in face to face interviews and between the standard 10 point scale in face to face and the 10 point scale in telephone interviewing. It was found that the change of format led to much larger differences in the response distributions in the different countries than the use of the 10 point scale in the different modes of data collection. Thus, strong evidence suggests that the two-step procedure for the left-right scale is not a good telephone alternative for the standard 10 point scale in face to face interviews. The results obtained with the two-step-version will not be comparable. On the other hand, the two 10 point scales remained more comparable for the two different modes of data collection, although also in that case significant differences between the two could be detected. As a consequence, corrections for these differences are required in order to compare the results obtained with the same scale in different data collection modes.

With respect to the standard Eurobarometer questions, mode effects occurred for some questions and not for others. In chapter 6, for the evaluation of the EU membership significant differences between face to face and telephone interviewing have been found, as was true for the satisfaction questions and for the persuade questions.

Using the same approach, in chapter 9, a larger set of close-ended questions was scrutinised whether there was a mode effect and also whether there was a difference in response probabilities in the different countries. This last point is interesting because comparison of responses across countries is only possible if in the different countries response probabilities, i.e. the probability of a specific answer given the (latent) opinion of a person, are the same. For example, regarding the "benefit question" the results are only comparable if people in different countries believing that their country has benefited from EU membership have an equal probability to also say that their country has benefited. If these response probabilities vary, the difference in responses does not result from a difference in opinion but from a difference in response probabilities. In table 11.2, the results of the respective test in this study is presented once more.

Starting with the ***media involvement*** questions, they have not been affected by the mode of data collection. The categories are relatively detailed and require separate estimates of frequencies, but apparently the mode of data collection has no effect. It is also important to note that there are no differences in response probabilities across countries. So the response categories for these questions can be used for comparison across modes and across countries.

The second set of questions concerns ***political involvement***. One of the two questions has the same response probabilities for the different modes and different countries, while the other produces differences between modes and across countries. The reason for this difference is that the question on "political discussion" entails the categories frequently, occasionally, never, and a DK/No answer, and the question "persuade" the often, from time to time, rarely, never, and a DK/No answer, that is one additional category. Comparing the meaning of the categories in the two questions just on face value, one could conclude that "rarely" is an unnecessary extra category. In fact, it turns out that it is precisely this category which causes the differences between the modes and the countries. Consequently, if the categories "rarely" and "never" are collapsed, the problems might disappear. After a test, it turned out indeed that

the model with equal probabilities across modes and across countries now fitted the data. This suggests that for purposes of comparison the categories “rarely” and “never” should be combined. To leave the category “rarely” completely out in the data collection is, of course, another viable option.

The third set of questions concerns ***satisfaction***. These two questions have unequal response probabilities across modes and countries. Looking at the response categories, however, the problems do not come as a surprise. The response categories for both questions are: very satisfied, fairly satisfied, not very satisfied, not at all satisfied, and a DK/No answer. The problems are that it can be confusing for translators and respondents how to interpret especially the labels “fairly satisfied” and “not very satisfied”. In fact, logically one could argue that after the “very satisfied” category “not very satisfied” contains all other possible answers and that therefore it is not clear when to use the category of “fairly satisfied”. In the latent class analyses it was indeed found that people in category 3 on the latent variable had a different probability for answering fairly satisfied and not very satisfied in the different modes. But this problem may also carry over into the translation of these categories into different languages.

Checking this hypothesis, differences in the translations in the different languages were indeed discovered. In French and Dutch, the translation of the labels were as follows: very satisfied, rather satisfied, rather dissatisfied, not at all satisfied. This is quite different from very satisfied, fairly satisfied, not very satisfied, not at all satisfied. If such differences exist in the translation between countries, it cannot come as a surprise that also differences in the reactions of the respondents across countries are found. The differences across modes must also have to do with the above mentioned problematic categories which may work differently in face to face than in telephone research.

Interesting in this respect are the findings for the ***involvement in the EU***. The question on knowledge uses the same category system as the satisfaction questions: very well informed, quite well informed, not very well informed, not at all well informed, and DK/No answer. So, if the above interpretation is correct, this question should have the same problems as the satisfaction question. In table 11.2 one can see that this is indeed the case, strengthening the argument given before.

The question on “interest in EU matters” uses different labels: a great deal, to some extent, not much, not at all, and a DK/No answer, but the problem is comparable. The term “not much” is a negation of “much”. So if one is less than “much interested” in the EU, one could choose “not much”. But then the position of the category “to some extent” is again not clear. It can be regarded as a part of the category “not much”, but that would lead to confusion. Thus, the same problems as for the other questions were expected and were indeed found, as can be seen in table 11.2. Checking the translations in the different countries, it was found that also in this case the translation of the labels is not equivalent.

Finally, seven ***opinion*** questions, all with the same format “pro, against and no opinion”, were analyzed. According to table 11.2, the questions concerning the introduction of the European Monetary Union and a common defense policy for the EU did not indicate any mode effect and differences across countries. On the other hand, all questions concerning the

elections showed a cross-cultural mode effect. In the telephone interview people with a score of “DK/No answer” on the latent variable have a higher probability to say “pro” than in the face to face interview. Besides that, the question concerning the division of tasks between local, national and EU government produced the same result.

It is difficult to explain these effects. It cannot be a general acquiescence bias (Schuman and Presser, 1981) because then it should occur for all questions. It is also not an effect of the topic because then one would have to find a different explanation for the last question. An interpretation as a learning effect is also questionable because then one would also expect this outcome for all topics and not only for a limited number. Besides that, there is no obvious reason why all people should learn that they have to respond “yes” to this question instead of “no”. So for the time being one has to accept these findings and will have to wait for further research to clarify the matter.

11.3.3 Adjustment for mode differences

Finally, an effort has been made in chapter 10 to see if it would be possible to adjust the results obtained with face to face interviewing to the results obtained with telephone interviewing. This activity makes sense because it is very inconvenient for users that the two data collection modes produce different results for the same questions. This could lead to a lot of confusion if these results would be reported without further comment. Imagine that the standard Eurobarometer (face to face) would present a positive opinion of the public in one month, and in the next month the telephone tracking study would present a much more negative opinion. Is there reason for alarm or not politicians and journalists might ask. In such a situation one has to be able to correct the tracking study data for the mode effects in order to check whether there are indeed significant differences to the standard Eurobarometer once this correction has been performed.

Chapter 10 suggests that a procedure can indeed be found which transforms the results from the telephone interviews to the results from the face to face interview and vice versa. However, this procedure needs to be specific for the response-nonresponse structures of those companies conducting the studies. This means that the numeric solution obtained in chapter 10 cannot be generalised to data collections which will occur in the future. Rather, they must be estimated again for new data because different companies collect these data. Furthermore, this chapter has suggested that the adjustments are specific for each question because the effects of coverage errors, nonresponse and mode effects are different for each question. The most important finding, however, is that such adjustments are possible. How stable the results are across time is an interesting issue for further research.

Table 11.2 The evaluation of differences between modes and countries for EB41.Panel

Variables	Equality of the parameters across countries		not across countries	
	across modes (model 1)	not across modes (model 2)	across modes (model 3)	not across modes (model 4)
Media involvement				
Radio	+			
Newspaper	+			
TV	+			
Political involvement				
Political discussion	+			
Persuade others	-	-	-	+
Satisfaction				
Life in general	-	-	-	+
Democracy in country	-	-	-	+
Involvement in EU				
Interest -	-	-	+	
Knowledge	-	-	-	+
Opinion on EU membership				
Benefit for country from EU membership	+			
Evaluation of membership for country	-	+		
Opinion on EU policies				
European Monetary Union	+			
EU defense	+			
Participation local elections	-	+		
Participation EU elections	-	+		
Candidacy local elections	-	+		
Candidacy EU elections	-	+		
Division of tasks between various levels of government	-	+		

11.4 Conclusion

In this study the comparability of the responses in surveys across modes of data collection and across countries was studied, and considerable differences were found. In this last section some consequences of this finding for survey research will be drawn, and some scientific issues will be discussed.

11.4.1 Practical consequences

The results clearly indicate that quite large differences can be expected for many questions of the Eurobarometer surveys if one compares the results of the standard face to face Eurobarometer and the telephone-based tracking study.

It has been shown that there are at least three factors which can explain why these differences occur. For the Eurobarometers where two different companies do the face to face study and the telephone study, it is hard to suggest ways to reduce these errors by adjustment of the existing procedures. Partially the differences are due to conditions which exist in telephone interviewing (many call-backs) but which do not exist in face to face interviewing, and vice versa (show cards).

Some of the mode effects can be avoided. It was found that for four questions the mode effects had to do with the formulation of the response categories. For these four questions the category labels were overlapping which led to confusion in the translations and to differences between modes. With adjustment of the categorisations these problems can be avoided in the future. For example, for the satisfaction question in some countries the labels very satisfied, fairly satisfied, not very satisfied, not at all satisfied should be substituted by the labels very satisfied, rather satisfied, rather dissatisfied, very dissatisfied.

Quite likely this categorisation will lead to less problems if used in all countries. The substantive disadvantage of this categorisation is that since usually more people in Europe are satisfied than dissatisfied, this scale is de facto used only as a two-point scale.

An alternative categorisation might be very satisfied, rather satisfied, little satisfied, not satisfied. Both formulations will probably be comparable across modes and countries, but this has to be tested. This categorisation will also lead to less confusion accross modes. For the two questions about involvement in the EU similar labels should be used to avoid problems.

It was also found that simple weighting procedures cannot be used to correct for the differences between the two types of studies. However, for all questions the correction procedure can be used which has been discussed in chapter 10 to make the results of the standard Eurobarometer as comparable as possible to a telephone study. For this purpose the pure mode effects and the coverage error as estimated in this study can be used while the fieldwork effects have to be calculated again because fieldwork organisations will probably change across time. If the connected “nonresponse’ difference has been assessed, the correction can be done for all questions in the same way as was demonstrated in chapter 10. If the procedures used by field organisations are not changed, these new estimates should be expected to remain the same, and the corrections can be used until a change in the procedure is introduced. At that moment again new estimates have to be obtained.

11.4.2 Scientific issues

Table 11.1 shows that large differences are found across modes for different variables. While the methodological literature gives most attention to coverage errors, it turned out that the coverage errors are relatively small compared with the other two types of errors. Especially

the effects of the procedures used by the different research organisations are relatively large. Unfortunately, this effect because of the many contributing factors cannot be decomposed into the different possible causes in the present study.

Table 11.2 has shown that for 11 out of the 18 questions the responses are effected by the pure mode of data collection when estimated on the basis of the panel data. This finding which is in agreement with the results in chapter 5 using a different approach, is also contradicting some of the standard literature on mode effects (Groves and Kahn, 1979; de Leeuw and van der Zouwen, 1988; de Leeuw, 1990) but agrees with other studies which found considerable effects (Silberstein et al., 1989; Kalfs, 1994; Scherpenzeel and Saris, 1997). This last study is the most comparable one to the one discussed in this book since there also panel data have been used. In panel studies the confounding factors can be better controlled although there one faces the additional problem of memory effect. In that study also mode effects were discovered, for example for the satisfaction variables. In a meta analysis of similar studies (Scherpenzeel 1995) in different countries also country-specific effects have been found like the ones reported in table 11.2. Such country-specific differences occur much less frequently than mode effects. For only five questions the response probabilities across countries were different. In four questions the category labels were overlapping which led to confusion in the translations and in the responses. These confusing category labels apparently are the major reason for the cross-cultural differences obtained. It would be desirable to pursue this idea in further research.

In this context the importance of the equality of the response probabilities has been emphasised because cross-cultural comparison requires these equalities. If they have not been found, then observed differences can be explained by the differences between the response probabilities instead of the differences in opinions. The analysis in chapters 6 and 9 has shown that the test of such equalities is relatively easily done using the latent class model and the program LEM (Vermunt, 1996).

In sum, this book has served three major purposes. For one, the methodological study reported here has indicated that it is worthwhile, consequential and possible for survey research to systematically study effects of coverage, interview modes and nonresponse on the answers of those questioned. There is no excuse for methodological naiveté in the matter. Secondly, it has pointed to the need to consistently continue to pursue these effects in ongoing everyday research in order to present methodologically enlightened findings to the public and to the clients. Thirdly, the study has also indicated where additional methodological research is necessary. It is hoped that this book contributes to all three of those prerogatives.

References

Alexander, C. (1987). A class of methods for using person controls in household weighting. Survey Methodology 2, pp. 183-198.

Almond, A. and S. Verba (1963). The civic culture: Political attitudes and democracy in five nations. Princeton. Princeton University Press.

Backstrom, C.H. and G. Hursh-Cesar (1981). Survey research. New York:

Barnes, S., M. Kaase, et al. (1979). Political action. Mass participation in five Western democracies. Beverly Hills and London: Sage Publications.

Beck, U. (1986). Die Risikogesellschaft, Frankfurt a.M.

Biemer, P.P. (1988). Measuring data quality. In Groves, R.M., P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls II, and J. Waksberg (Eds.), Telephone survey methodology, New York: John Wiley, pp. 273-282.

Billiet, J., G. Loosveldt and L. Waterplas (1986). Het survey interview onderzocht. Leuven: Departement Sociologie.

Bradburn, N.M. (1992). A response to the nonresponse problem. Public Opinion Quarterly, 56, pp. 391-397.

Cannel, C.F., R.M. Groves, L. Magilavy, N. Mathiowetz and P.V. Miller (1987). An experimental comparison of telephone and personal health surveys. National Center for Health Statistics, Technical Series 2, no. 106.

Champness, B.G. (1972). The perceived adequacy of four communication systems for a variety of tasks. London: Communication Studies Group, Paper no. E/72245/CH.

Clark, T.N. and S.M. Lipset (19). Are social classes dying? International Sociology 6, No. 4, pp. 397-410.

Converse, P.E. (1964). The nature of mass belief systems in mass publics. In: Apter, D.E. (Ed.), Ideology and discontent, New York: The Free Press, pp. 206-261.

Converse, P.E. (1975). Some mass-elite contrasts in the perception of political spaces. Social Science Information, 14, pp. 49-83.

Deming, W.E. and F.F. Stephan (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. The Annals of Mathematical Statistics, 11, pp. 427-444.

Deville, J.C. and C.E. Särndal (1992). Calibration estimators in survey sampling. Journal of the American Statistical Association, 87, pp. 376-382.

Deville, J.C., C.E. Särndal and O. Sautory (1993). Generalized ranking procedures in survey sampling. Journal of the American Statistical Association, 88, pp. 1013-1020.

Dillman, D.A. (1978). Mail and telephone surveys. The total design method. New York:

Drews, H.P. (1994). Nicht zu verantworten. Telefonische Bevölkerungsumfragen in Ostdeutschland. Planung und Analyse, 1, pp. 64-65.

DuToit, S.H.C., A.G.W. Steyn and R.H. Stumpf (1986). Graphical exploratory data analysis. New York:

Ekman, P. (1965). Differential communication of affect by head and body cues. Journal of personality and social psychology, 2, pp. 726-735.

Elliot, D. (1991). Weighting for non-response. London: Office of Population Censuses and Surveys.

ESOMAR (1994). Annual market study on market statistics, Amsterdam.

ESOMAR (1996). Market research industry turnover. European Society for Opinion and Marketing Research. Trend report 1990-1994. Amsterdam.

Ester, P., L. Halman and R. de Moor (1993). The individualizing society. Value change in Europe and North America. Tilburg: Tilburg University Press.

Euler, M. (1989) Ausstattung privater Haushalte mit ausgewählten langlebigen Gebrauchsgütern im Januar 1988. Wissenschaft und Statistik, 5, pp. 307-315.

Eyk, C. van de and B. Niemoller (1984). Theoretical and methodological considerations in the use of left-right scales. Paper presented at the ECPR Joint Sessions of Workshops, Salzburg, Austria.

FORSA (1994). Research method of the FORSA European Telephone Survey 1994: A feasibility study. In: ZEUS Eurobarometer experiment: Integrated data set of EB 41.0 variables, of EB 41.0 panel survey variables and of the FORSA European survey. Unpublished manuscript. Mannheim: Mannheim University.

Frey, J., G. Kunz and G. Lüschen (1990). Telefonumfragen in der Sozialforschung, Opladen: Westdeutscher Verlag.

Fuchs, D. and H.-D. Klingemann (1990). The left/right schema. In: Jennings, M.K., J.W. van Deth, et al., Continuities in political action: A longitudinal study of political orientations in three Western democracies, Berlin: de Gruyter, pp. 204-234.

Gabler, S. (1994). Eine allgemeine Formel zur Anpassung an Randtabellen. In: Gabler, S., J.H.P. Hoffmeyer-Zlotnik and D. Krebs (Eds.), Gewichtung in der Umfragepraxis, Opladen: Westdeutscher Verlag,

Gabler, S., H.P.J. Hoffmeyer-Zlotnik and D. Krebs (Eds.) (1994). Gewichtung in der Umfragepraxis. Opladen: Westdeutscher Verlag.

Garnham, N. (1988). Universal Service - Objectives and practice in international comparison. In: Neu, W. and K.H. Neumann (Eds.), Die Zukunft der Telekommunikation in Europa. Berlin,

Gölz, R. (1983). Mehr über das Telefon, Stuttgart:

Goodman, L.A. (1974a). The analysis of systems of qualitative variables where some of the variables are unobserved. Part I - A modified latent structure approach. American Journal of Sociology,79, pp. 1179-1259.

Goodman, L.A. (1974b). Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika, 61, pp. 215-231.

Groves, R.M. (1978). On the mode of administering a questionnaire and responses to open-ended items. Social Science Research, 7, pp. 257-271.

Groves, R.M. (1989). Survey errors and survey costs. New York: Wiley.

Groves, R.M. and R.L. Kahn (1979). Surveys by telephone: A national comparison with personal interviews. New York: Academic Press.

Haberman, S.J. (1979). Analysis of qualitative data, Vol. 2: New developments, New York: Academic Press.

Häder, S. (1994). Auswahlverfahren bei Telefonumfragen. ZUMA-Arbeitsbericht 94/03.

Hagenaars, J.A. (1990). Categorical longitudinal data: Loglinear panel, trend and cohort analysis. London, Sage.

Hájek, J. (1981) Sampling from a finite population. New York: Marcel Dekker Inc.

Harding, S.; Phillips, D. (with Michael Forgarty). 1986: Contrasting values in western Europe. Unity, diversity and change. Houndmills, Basingstoke, Hampshire and London: The Macmillan Press Ltd.

Heyde von der, Ch. (1994) Gewichtung am Beispiel: Einwohnermeldeamt versus Random Route. In: Gabler, S., J.H.P. Hoffmeyer-Zlotnik and D. Krebs (Eds.): Gewichtung in der Umfragepraxis. Opladen: Westdeutscher Verlag.

Hochstim, J.R. (1967) A critical comparison of three strategies of collecting data from households. Journal of the American Statistical Association, Vol. 62, pp. 976-989.

Hradil, S. (1992) Sozialstruktur und gesellschaftlicher Wandel. In: Gabriel, O.W. (Ed.): Die EG-Staaten im Vergleich, Opladen, pp. 50-94.

Inglehart, R. (1977) The silent revolution: Changing values and political styles among western publics. Princeton, Guildford: Princeton University Press.

Inglehart, R. (1990) Culture shift in advanced industrial society. Princeton NJ: Princeton University Press.

Inglehart, R. (1997) Modernization and Postmodernization. Cultural, economic and political change in 43 societies. Princeton: Princeton University Press.

INRA (1994) "Codebook Eurobarometer 41.0: Explanatory note. In: ZEUS Eurobarometer experiment: Integrated dataset of EB 41.0 variables, of EB 41.0 panel survey variables and of variables of the FORSA European survey. Mannheim: ZEUS.

INRA (1994) Codebook Eurobarometer 41.0: Explanatory note. In ZEUS Eurobarometer experiment: Integrated data-set of EB41.0 variables, of EB41.0 Panel survey variables and of variables of the FORSA European Survey, Mannheim, ZEUS.

INRA (1994) Eurobarometer experiment: survey documentation. In ZEUS Eurobarometer experiment: Integrated data-set of EB41.0 variables, of EB41.0 Panel survey variables and of variables of the FORSA European Survey, Mannheim, ZEUS.

INRA (1994) Software Manual of the Eurobarometer Studies.

INRA (1995) Standard Eurobarometers: Methodological report. Intern report INRA European Co-ordination Office.

Institut für Sozialstudien (ISSAS) (1990) Armut in Zahlen. Luxemburg.

Ireland, C.T. and S. Kullback (1968) Contingency Tables with given marginals. Biometrika 55, pp. 179-188.

Jennings, M.K. and J.W. van Deth (1990) Continuities in political action. A longitudinal study of political orientations in three western democracies. Berlin: de Gruyter.

Jung, M. (1990) Ausschöpfungsprobleme bei repräsentativen Telefonumfragen, in: Forschungsgruppe Telefonkommunikation (Ed.): Telefon und Gesellschaft. Berlin. Vol. 2, pp. 386-398.

Kaase, M. and K. Newton (1995) Beliefs in government. Vol. 5 of the Beliefs in Government Series. Oxford: Oxford University Press.

Kalfs, N. (1994) Hour by hour, Amsterdam, Nimmo.

Kass, G.V. An exploratory technique for investigating large quantities of categorial data, Applied Statistics, Vol. 29, pp. 119-127.

Kish, L. (1965) Survey sampling. New York: John Wiley.

Kish, L. (1995) Methods for design effects. Journal of Official Statistics 11, pp. 55-77.

Klingemann, H.D. (1979) The background of ideological conceptualization. In: Barnes, S.H., Kaase, M., et al. Political action. Mass participation in five western democracies, Beverly Hills: Sage pp. 255-279.

Koch, A., S. Gabler and M. Braun (1994) Konzeption und Durchführung der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften. ZUMA-Arbeitsbericht 94/11.

Kohl, J. (1992) Armut im internationalen Vergleich. Methodische Probleme und empirische Ergebnisse. In: Leibfried, S.; Voges, W. (Eds.): Armut im Wohlfahrtsstaat, Sonderheft der Kölner Zeitschrift für Soziologie und Sozialpsychologie, Opladen Vol. 32, pp. 272-299.

Kormendi, E. (1988) The quality of income information in telephone and face to face surveys. In: Groves, R.M. et al. (Eds.). New York: Wiley.

Kreiselmaier, J. and R. Porst (1989) Methodische Probleme bei der Durchführung telefonischer Befragungen: Stichprobenziehung und Ermittlung von Zielpersonen, Ausschöpfung und Non-Response, Qualität der Daten, ZUMA-Arbeitsbericht Mannheim: ZUMA.

Lange, U. and A. Zernick (1990) Gedanken zur Vielfalt des Telefons in Europa. In: Forschungsgruppe Telekommunikation (Ed.): Telefon und Gesellschaft, Berlin Vol. 2, pp. 88-123.

Lazersfeld, R.F. (1950a) The logical and mathematical foundation of latent structure analysis. In: S. Stouffer (Ed.) Measurement and prediction Princeton: Princeton University Press, pp. 362-412.

Lazersfeld, R.F. (1950b) Interpretation of statistical relations as research operation. In: Stouffer, S. (Ed.) Measurement and prediction Princeton: Princeton University Press, pp. 413-472.

Leeuw, E. de (1992) Data quality in mail, telephone and face to face surveys. Ph. D. Thesis. Amsterdam.

Leeuw, E. de and H. van der Zouwen (1988) Data quality in telephone and face to face surveys: A comparative metaanalysis. In: Groves, R.M. et al. (Eds.): Telephone Survey Methodology. New York: Wiley.

Levitin, T.E. and W.E. Miller (1979) Ideological interpretations of presidential elections. American Political Science Review, Vol. 73, pp. 751-771.

Little, R.J.A. and M.M. Wu (1991) Models for contingency tables with known margins when target and sampled populations differ. Journal of the American Statistical Association 86, pp. 87-95.

Lyberg, L. and Kasprzyk, D. (1991). Data collection methods and measurement error: An overview. In P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman (Eds.), Measurement errors in surveys, pp. 237-257. New York: John Wiley and Sons.

Meier, E. (1995) The response rates of the National Readership Survey. Paper for the Worldwide Readership Symposium 1995.

Mekrabian, A. 1968: Silent messages. Belmont: Wadsworth.

Meurs, A. van and W.E. Saris (1990) Memory effects in MTMM studies. Chapter 6 in: W.E. Saris and A. van Meurs (Eds.). Evaluation of measurement instruments by meta-analysis of multitrait multimethod studies. Amsterdam: North-Holland.

Meurs, A. van and W.E. Saris (1989) Memory effects in MTMM studies. In: Saris, W.E.; Meurs, A. van 1996. (Eds.): Evaluation of measurement instruments by meta analysis of multitrait multimethod studies. Amsterdam: North-Holland.

Miller, P. (1984) Alternative question forms for attitude scale questions in telephone interviews. Public Opinion Quarterly, Vol. 49, no. 4, pp. 766-768

Miller, P.V. (1984) Alternative question forms for attitude scale questions in telephone interviews. Public Opinion Quarterly, Vol.49, no. 4, pp. 766-768

Monsees, M. L. and J.L. Massay (1979) Adapting procedures for collecting demographic data in personal interviews to a telephone interview. Proceedings of the section on survey research methods, American Statistical Association, pp. 130-135.

Moyal, A. (1990) Women and the telephone in Australia: The study of a national culture. In: Forschungsgruppe Telekommunikation (Ed.) Telefon und Gesellschaft. Berlin, pp. 190-198.

Niedermayer, O. and R. Sinnott, R. (1995) Public opinion and internationalized governance. Vol. 2 of the Beliefs in Government Series. Oxford: Oxford University Press.

Noble, G. (1990) Social aspects of domestic telephone use. In: Forschungsgruppe Telekommunikation (Ed.): Telefon und Gesellschaft, Berlin. Vol. 2, pp. 176-189.

Porst, R. 1993. Ausschöpfungen bei sozialwissenschaftlichen Umfragen. Annäherung aus der ZUMA-Perspektive. Mannheim: ZUMA.

Reif, K. and Inglehart, R. (1991) Eurobarometer: The dynamics of european public opinion. Essays in honour of Jacques-René Rabier. Houndsmill, Basingstoke and London: The Macmillan Press.

Rösch, G. (1994) Kriterien der Gewichtung einer nationalen Bevölkerungsstichprobe. In: Gabler, S., J.H.P. Hoffmeyer-Zlotnik and D. Krebs (Eds.): Gewichtung in der Umfragepraxis. Opladen: Westdeutscher Verlag.

Rothe, G. (1994) Wie (un)wichtig sind Gewichtungen? Eine Untersuchung am ALLBUS 1986. In: Gabler, S., J.H.P. Hoffmeyer-Zlotnik and D. Krebs (Eds.): Gewichtung in der Umfragepraxis. Opladen: Westdeutscher Verlag.

Rothe, G. and M. Wiedenbeck (1994) Stichprobengewichtung: Ist Repräsentativität machbar? In: Gabler, S., J.H.P. Hoffmeyer-Zlotnik and D. Krebs (Eds.): Gewichtung in der Umfragepraxis. Opladen: Westdeutscher Verlag.

Saris, W.E. (1986) Variation in response functions: A source of measurement error in attitude research. Amsterdam: SRF.

Scherpenzeel, A.C. (1995) Meta analysis of a European comparative study. In: Saris, W.E.; Münnich, A. (Eds.) The Multitrait-Multimethod Approach to Evaluate Measurement Instruments. Budapest: Eötvös University Press.

Scherpenzeel, A.C. and W.E. Saris. (1997) The validity and reliability of survey question: A meta analysis of MTMM studies. Sociological methods and research, Vol. 25, pp. 341-383.

Schnell, R. (1993) Die Homogenität sozialer Kategorien als Voraussetzung für 'Repräsentativität' und Gewichtungsfaktoren. Zeitschrift für Soziologie 22, pp. 16-32.

Schnell, R. and U. Kohler (1995) Empirische Untersuchung einer Individualisierungs-hypothese am Beispiel der Parteipräferenz von 1953-1992. Kölner Zeitschrift für Soziologie und Sozialpsychologie, Vol. 47, pp. 634-657.

Schuman, H. and S. Presser (1981) Questions and answers in attitude surveys: experiments on question form, wording and context. New York: Academic Press.

Schwarz, N. and Hippler, H.-J. (1991). Response alternatives: The impact of their choice and presentation order. In P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman (Eds.), Measurement errors in surveys, pp. 41-56. New York: John Wiley and Sons, Inc.

Schwarz, N., Strack, F., Hippler, H.-J. and Bishop, G.F. (1991). The impact of administration mode on response effects in survey measurement. Applied Cognitive Psychology, 5, pp. 193-212.

Silberstein, A.S. and S. Scott (1991) Expenditure diary surveys and their associated errors. In: P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz and S. Sudman (Eds.). measurement errors in Surveys. New York: Wiley. pp. 303-327.

Sosdian, C.P. and L.M. Sharp (1980) Nonresponse in mail surveys: Access failure or response resistence. Public Opinion Quarterly, Vol. 44, pp. 396-402.

Statistisches Bundesamt (1994) Statistisches Jahrbuch.

Sykes, W. and G. Hoinville (1985) Telephone interviewing on a survey of social attitudes: A comparison with face to face procedures. London: Social and Community Planning Research.

Szalai, A. and R. Petrella (1977) (in collaboration with Stein Rokkan and Erwin K. Scheuch). Cross-national comparative survey research. Theory and practice. Oxford, New York, Toronto, Sydney, Paris and Frankfurt: Pergamon Press.

Thomassen, J. (1994) The intellectual history of election studies. Special issue of the European Journal of Political Research, 25, 237-385.

Verba, S. (1971) Cross-national survey research: The problem of credibility. In: Vallier, Ivan (Ed.): Comparative methods in sociology. Essays on trends and applications. Berkeley, Los Angeles, London: University of California Press, pp. 309-356.

Vermunt, J.K. (1995) Loglinear event history analysis. A general approach with missing data, latent variables and unobserved hetrogeneity. Tilburg: WORC.

Vermunt, J.K. (1996) Loglinear event history analysis. A general approach with missing data, latent variables and unobserved heterogeneity. Tilburg, WORC.

Williams, E. (1974) A summary of the present state of knowledge regarding the effectiveness of the substitution of face to face meetings by telecommunicated meetings. London: Communication studies group, Paper no. P/74294/WL.

Williams, E. (1977) Experimental comparison of face to face and mediated communication: A review. Psychological Bulletin, Vol. 84, pp. 963-976.

Wiseman, F., Rogers, T.F., Groves, R.M., and Smith, T.W. (1989). Mode of administration. In E. Singer and S. Presser (Eds.), Survey research methods: A reader, pp. 187-244. Chicago, IL, US: University of Chicago Press.

Woltman, H.F., Turner, A.G., and Bushery, J.M. (1980). A comparison of three mixed-mode interviewing procedures in the national crime survey. Journal of the American Statistical Association, 75, pp. 534-543.

Zaller, (1992) The nature and origins of mass opinion. Cambridge: Cambridge, University Press.

Zaslavsky, A.M. (1988) Representing local area adjustments by reweighting of households. Statistics Canada 14, pp. 265-288.

ZEUS (1994) Eurobarometer experiment: Integrated data-set of EB41.0 variables, EB41.0-panel survey variables and of variables of the Forsa European survey.

ZEUS (1994) ZEUS Eurobarometer experiment : Integrated data set of EB 41.0 variables, of EB 41.0 panel survey variables and of variables of the FORSA European Survey. Mannheim: ZEUS.

Biographical notes on contributers

SIEGFRIED GABLER is the Director of the Department of Statitics at the Center for Survey Research and Methodology (ZUMA) and lecturer at the University of Mannheim where he received his doctorate in mathematics (1978). His main research interest is sampling theory. He published numerous papers in different journals, wrote a book on "Minimax Solutions in Sampling from Finite Populations" and was co-editor of two books on Sampling and Weighting.

ANGELIKA GREIL received her M.A. degree in English, German and Sociology from the University of Mannheim (1997). She was a research assistant at ZEUS since 1989. After her studies at the University of Alabama (1991-1992), she assisted Peter Schubert in taking over responsibility for the Question Bank, one part of the dual ZEUS Database system. She is enrolled in the Ph.D. program of the English Department at the University of Mannheim, writing her dissertation about Djuna Barnes, an American author of the early 20th century.

JACQUES A. HAGENAARS if Full Professor of Methodology of the Social Sciences at Tilburg University, Netherlands. His current research interest centers on categorical and longitudinal causal analysis. He is author of LCAG, a program for loglinear modeling with latent variables, author of Categorical Longitudinal Data: Loglinear Panel, Trend, and Cohort Analysis (Sage, 1990) and has written many articles that have appeared in national (Dutch) and international scholarly journals. At present, he is involved in research projects concerning general loglinear modeling and causality and quasi-experimentation.

SABINE HÄDER is a member of the Department of Statistics at the Center for Survey Research and Methodology (ZUMA). She studied statistics in Berlin and received her doctorate from the Hochschule für Ökonomie, Berlin, in 1987. She is mainly interested in sampling methods, especially for telephone surveys, and weighting for nonresponse. Another field of interest is the application of the Delphi technique to the Social Science. She published her research results in several contributions.

MAX KAASE, born in Krefeld, Germany, on May 14, 1935 is Professor of Political Science and Comparative Survey Research at the University of Mannheim. He has been on leave from that position since 1993 and now is a research professor at the Wissenschaftszentrum Berlin für Sozialforschung. Since 1992, he has been a member of the Senate and of the Executive Committee (Hauptausschuß) of the Deutsche Forschungsgemeinschaft (German Research Council) as well as a member of the Standing Committee for the Social Sciences of the European Science Foundation. He is also head of board of the Mannheim-based academic survey research center "Zentrum für Umfragen, Methoden und Analysen" of which from 1974 to 1980 he was its first director. He has published widely in the fields of comparative politics, political, sociology and mass communication. His book publications range from "Political Action" (1979) jointly with Samuel H. Barnes and others) to "Beliefs in Government" (1995, with Kenneth Newton), "Estranged Friends? The Transatlantic Consequences of Societal Change" (1996, with Andrew Kohut) and "Politisches System" (1996, with Andreas Eisen and others).

HANS-DIETER KLINGEMANN (born 1937 in Eibeck, Germany) earned his academic degrees from the University of Cologne (1961: Dipl.-Kfm.; 1966: Dr. rer. pol.) and from the University of Mannheim (1978: Dr. habil). He has held positions at the University of Cologne (1966-74: Assistant Professor; Zentralarchiv für Empirische Sozialforschung), the Center for Survey Research (ZUMA) Mannheim (1974-80: Deputy Director), the Free University of Berlin (1980-: Professor of Political Science) and the Social Science Research Center Berlin (1989-: Director, Research Unit on Institutions and Social Change). Professionally, Professor

Klingemann has served, among other appointments, as President of the International Society of Political Psychology (1986/87), as member of the Executive Committee of the European Consortium for Political Research (1988/94) and as Vice-President of the International Political Science Association (1994/97). Recent publications are: Hans-Dieter Klingemann, Richard I. Hofferbert and Ian Budge. 1994 „Parties, Policies, and Democracy. Boulder, Colorado: Westview; Hans-Dieter Klingemann and Dieter Fuchs (eds.). 1995 „Citizens and the State". Oxford: Oxford University Press; Robert E. Goodin and Hans-Dieter Klingemann (eds.). 1996. „A Handbook of Political Science". Oxford: Oxford University Press. His current research is in the field of democratic transition and consolidation. In this area he conducts a major study on values and value change in Central and Eastern Europe jointly with Professor Ronald F. Inglehart, The university of Michigan and Max Kaase, Social Science Research Center Berlin.

JÜRGEN LASS is senior research manager in GFM-GETAS/WBA, Hamburg. Before he was research assistant at the Social Science Research Center Berlin (1991-1996). He earned his academic degrees from the Free University Berlin (1990: Dipl. Pol., 1993: Dr. phil.). He has published several articles and book chapters on political culture and electoral sociology. His current interests concentrate on methodological problems of survey research.

WILLEM E. SARIS is Full Professor of Methodology of the Social Science at the University of Amsterdam. He has published several papers on application of Structural equations models in th social sciences, on quality of measurement and data collection procedures. With an International Research group on Methodology and Comparative Survey research (IRMCS) he is doing research on comparative research which had led to a book on the comparative methodology (Saris and Muennich, 1995) and one on a substantive application of this approach in the field of life satisfaction (Saris, Veenhoven, Scherpenzeel and Bunting, 1996).

HERMANN SCHMITT is the Director of the Center for European Survey Analysis and Studies (ZEUS) and a senior research fellow at the Mannheim Center for European Social Research (MZES) and the University of Mannheim. He is the coordinator of the European Election Study research group and the convenor of the German national election study group. His main research topics are comparative electoral behavior, political representation and the transformation of parties and party systems. Among his publications are: „Neue Politik in alten Parteien" (1987); „Das Parteiensystem der Bundesrepublik Deutschland" (1987); „The European Elections of 1989" (1991, with Renato Mannheimer); „Wahlen und Europäische Einigung" (1994, with Oskar Niedermayer); „Parteien, Parlamente und Wahlen in Skandinavien" (1994 with Franz U. Pappi); „Das Europäische Mehrebenensystem" (1996, with Thomas Koenig and Elmar Rieger); „The European Elections Study 1994" (Datafile and Codebook, 1997, with Cees van der Eijk, Franz U. Pappi and Evi Scholz).

PETER R. SCHROTT is Director of the Department of Text Analysis, Media Analysis and Coding at the Center for Survey Research and Methodology (ZUMA). He studied Political Science and Sociology at the University of Tübingen/Germany, Law at the University of Caen/France and received his Ph.D. in 1986 from the State University of New York at Stony Brook/USA. He is interested in electoral behavior, political communication, public opinion and research methodology. He has widely published in journals like Political Behavior, Public Opinion Quarterly, British Journal of Political Science and numerous German journals. He also co-authored „The Joint Press Conference", a book on American Presidential televised debates.

PETER SCHUBERT received his diploma in Sociology from the University of Mannheim. Since 1990, he was the administrator of the dual ZEUS data base system coordinating the numeric Data Base and the Question Bank. He was responsible for the establishment of trends, the harmonization of trend and weight variables, the design of a unique coding system for

political parties and a complex region coding scheme. In this function, he provided the EB41.Experiment data set. Over the last year, he concentrated on establishing internet accessibility of the ZEUS Data Base system.

MICHAELA THOMA received her degree in Sociology from the University of Mannheim. She was Project Director at ZUMA from 1992 to 1997. Her current research interests focus on the application of social cognition research to survey methodology, nonresponse and the impact of traditional and computer assisted data collection on data quality. With support by a stipendium from the University of Mannheim she is about to finish her Ph.D. thesis on an experimental mode comparison study comparing different computer assisted to paper-and-pen methods of survey data collection.