Comparing societies and cultures: challenges of cross-cultural survey research as an approach to spatial analysis

Baur, Nina

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:
Baur, N. (2014). Comparing societies and cultures: challenges of cross-cultural survey research as an approach to spatial analysis. *Historical Social Research*, *39*(2), 257-291. https://doi.org/10.12759/hsr.39.2014.2.257-291

Comparing Societies and Cultures. Challenges of Cross-Cultural Survey Research as an Approach to Spatial Analysis

Nina Baur *

Abstract: *»Gesellschafts- und Kulturvergleich. Die Leistungsfähigkeit der interkulturell-vergleichenden Umfrageforschung als Methode der Raumanalyse«.* The paper shows how cross-cultural, cross-societal, cross-national, multinational and international comparative survey researchers have been handling space since the 1950s and how it can be used for spatial analysis. Using the concepts of the Survey Life Cycle and the Total Survey Error (TSE), the paper discusses two major methodological problems cross-cultural survey methodology faces: The problems of (1) equivalence and (2) demarcation.

Keywords: Spatial sociology, cross-cultural research, survey methodology, equivalence, demarcation, survey life cycle, total survey error.

1. Cross-Cultural Survey Research and Spatial Analysis

Since the 1920s, survey data have been an important data source for analyzing (inter)action and distribution within space, as well as relations and movements between spaces (Scheuch 1993 [1989], 174). In *survey research* (Baur 2009a; Behnke et al. 2010, 97-103, 107-12, 223-32), especially in *cross-cultural survey research*[1] (Rippl and Seipel 2008; Pennell et al. 2010), space is typically either treated as an independent variable – as its properties might influence, enable or constrain human (inter)action – or it is seen as a frame of reference in which social (inter)action takes place. Thus, in these cases, the research interest is not space itself but what happens within space. Alternatively, research aims at comparing the spatial units which are then treated as cases, i.e. entities in their own (Verba 1968; Scheuch 1989 [1968]; Rokkan 1970). Looking at sur-

* Nina Baur, Technische Universität Berlin, Department of Sociology, Fraunhoferstraße 33-36 (FH 9-1), 10587 Berlin, Germany; nina.baur@tu-berlin.de.

[1] Please note that there is a much-discussed theoretical and terminological difference between *cross-cultural*, *cross-societal*, *cross-national*, *multi-national* and *international comparative survey research*, as these definitions work with different spatial concepts and aim at different types of comparison (Rokkan 1970). For the purpose of this paper I treat them synonymously, as I want to focus on the principal methodological problems that all of these types of study types have in common.

vey methodology from a spatial point of view, the *research process* typically consists of the following steps (Baur et al. 2014, in this HSR Special Issue).

1.1 Defining a Frame of Reference

Researchers typically start with defining a given space as a *frame of reference*. Typically, they use topological projections of physical space on maps (e.g. Graph 1).

Graph 1: Defining a Frame of Reference (World Map of Physical Space)

This world map uses the so-called Robinson projection of three-dimensional space into two-dimensional space (see also Baur et al. 2014, in this HSR Special Issue). Source: E Pluribus Anthony [Public Domain], via Wikimedia Commons <http://upload.wikimedia.org/wikipedia/commons/1/17/BlankMap-World-noborders.png> (accessed 19 February 2014).

1.2 Defining Units of Analysis

Researchers then divide the frame of reference into *units of analysis*. These *spatial units* (i.e. *territories*) are conceptionalized as "*(target) populations*". Note that the concept of statistical population implicitly treats space as a container, as populations have to be defined with regard to contents, temporally and *spatially*, i.e. they address a specific set of people being located at a specific time in a specific spatial unit demarcated by borders from other spatial units.

One could divide the world map into areal units of equal size. However, this is not typically done. Instead, most researchers use *politically defined administrational units*, e.g. *nation states, regions, cities or quarters*, as Graph 2 exemplifies.

As researchers in research practice rarely reflect their terminology, it is often not clear, if the aim of demarcating these territories is to compare political states (as in "cross-national research"), cultures ("cross-cultural research") or societies ("cross-societal research") (Rokkan 1993 [1970], 9). Further, "the prefix

cross- stresses the objects of comparison while the prefix *inter*- [and *multi*-] relates to a characteristic of the research organization" (Rokkan 1993 [1970], 9).

Graph 2: Dividing the Frame of Reference into Spatial Units: Example of Political Borders of Nation States in 1985

Source: Several unknown authors [GNU Free Documentation License 1.2], via Wikimedia Commons, <http://upload.wikimedia.org/wikipedia/commons/e/e0/BlankMap-World-1985.png>, (accessed February 19, 2014).

1.3 Sampling Units of Analysis

Next, researchers select one or more units of analysis for detailed analysis. Sometimes, researchers simply select all spatial units. However, in cross-cultural survey research, typically several nations are selected, as illustrated in Graph 3. Sometimes the selection process depends on practical issues, e.g. countries willing to participate in data collection.

Alternatively, countries are chosen using theoretical indicators as suggested by the case study method, i.e. a number of dimensions is defined that are used to classify the territories (Scheuch 1968; Verba 1968; Rokkan 1970; De Baus 2008; Rohlfing 2009; Pennings 2009). For example, Rokkan (1993 [1970], 44) suggests four dimensions for comparing nation states:

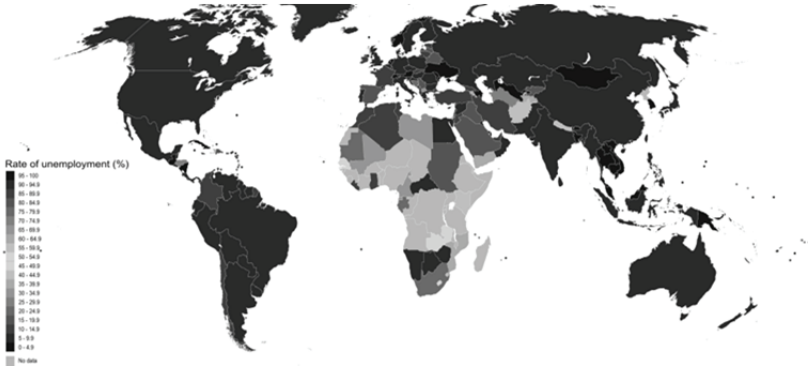
1) "Centre – Periphery" ("One standardized national language or several?"),
2) "State – Church" ("Establishment of national Church vs. alliance with a supranational Church vs. establishment of competing secular agencies?") as cultural dimensions and
3) "Urban – Rural" ("Protection of urban vs. rural products against foreign competition: the tariff issue."),
4) "Owner – Worker" ("Protection of rights of property vs. equalization of economic conditions through union and/or state action.") as economic dimensions.

Another, newer classification is that of *Welfare Regimes* suggested by Esping-Andersen (1990). An alternative classification scheme is selecting countries

because they are similar in data collection infrastructure and respondent behaviour, making it easier to collect comparable data. For this reason, many newer comparative research design focus e.g. in European or OECD countries.

Regardless which classification scheme for territories is used, the dimensions and typologies should be theoretically fruitful, i.e. relevant for the research question. Then, using this classification the spatial units are typically sampled using the most-different and/or the most-similar cases design, if the number of spatial units under examination is small (Behnke et al. 2010, 200-10; Hering and Schmidt 2014). If a larger number of territories is to be analysed, one can alternatively use random sampling designs (Delmelle 2009; Heeringa and O'Muircheartaigh 2010).

Graph 3: Selection Comparison and Classification of Spatial Units: Example for Unemployment Rates in Various Countries around 2010



Source: Jolly Janner [Public domain], via Wikimedia Commons, <http://commons.wikimedia.org/wiki/File:World_map_of_countries_by_rate_of_unemployment.png?uselang=de>, revision on 31 December 2013. If possible, the IDU (International Definition of Unemployment) rate is used. Fort further details on methodology see: <http://en.wikipedia.org/wiki/List_of_countries_by_unemployment_rate>.

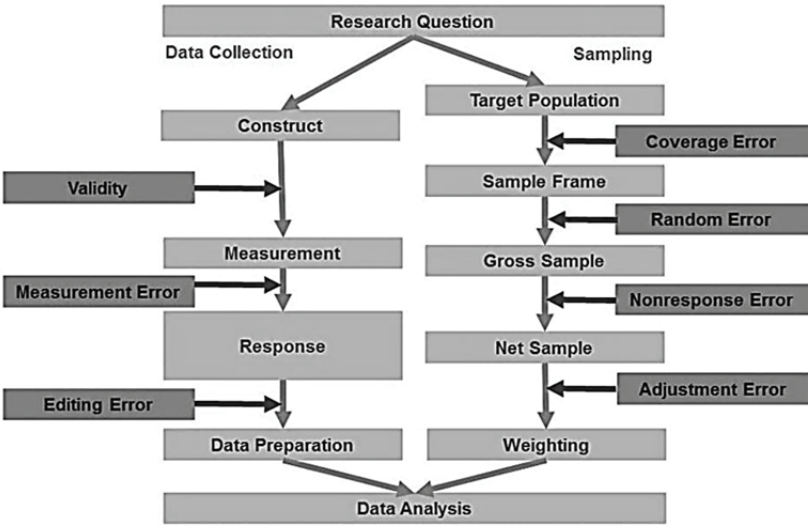
1.4 Sampling and Collecting Data for Each Unit of Analysis

Researchers sample and collect data for each unit of analysis in order to describe what is happening within this spatial unit as frame of reference. For example, Graph 3 illustrates the unemployment rates in various countries around 2010. Other examples of cross-cultural surveys are the Eurobarometer, the World Value Survey, PISA, ISSP, ESS, LIS and EVS.

In modern survey methodology, this part of the research process is conceptionalized by the model of the *Survey Life Cycle* (Graph 4), which stresses that sampling and data collection are entwined and have to be processed parallelly. Survey methodology also tries to integrate various errors that can occur into one model (Groves et al. 2009).

Specifically, during the *sampling process*, a sampling frame has to be defined. In most cases, this is not identical with the target population, i.e. there can be a coverage error. When a random sample is drawn from the sampling frame, a random error can occur. As humans cannot be forced to participate in surveys, due to noncontacts and refusals, there is usually a difference between people in the gross sample and people in the net sample. Finally, researchers have to decide, if and how to weight data, which implies that there might be an adjustment error.

Graph 4: Survey Life Cycle and Total Survey Error (TSE)



Source: Own Creation, based on discussion in Groves et al. (2009), Rippl and Seipel (2008) and Pennell et al. (2010).

For *data collection*, researchers have to define the constructs to be measured and they have to operationalize them, i.e. define, how to measure them. If they err, data are not valid. As people can respond differently to the same survey question, there might be measurement error. Finally, during data preparation, a processing error might occur.

These various errors are often related in the sense that – if you reduce one error type – another error increases. For example, in order to minimize measurement error, it is typically recommended to ask many different and detailed questions, a classical example being a psychological test for diagnosis of mental disorders, which usually takes more than one hour to answer and is very precise. However, these kinds of questionnaires could not be used in surveys of the general population, as respondents typically are willing only to spend only a limited time on answering survey questions. If the survey is too long, they

will either outright decline to participate in the survey or drop out during processing the survey – which in turn results in unit or item nonresponse. Therefore, researchers typically ask fewer questions in surveys, which increases the likelihood of measurement error. In this example, there is a trade-off between measurement error (due to short questionnaires) and nonresponse error (due to long questionnaires).

The other error types are likewise related. Therefore – while traditionally, these various errors were handled individually – modern survey methodology tries to incorporate them into the concept of the *Total Survey Error (TSE)* (Groves et al. 2009). This means that researchers should take into account *all* errors and try to minimize the error as a whole. The TSE is supposed to indicate the overall likelihood of an error. Note that this also means that survey methodology assumes that in research practice it is not possible to "objectively" measure reality. Instead, there might always a difference between what truly happens and what the data convey.

1.5 Major Methodological Challenges

Starting from the above model of the research process, cross-cultural survey methodology addresses two problems arising from this way to handle space methodologically:

1) *The Problem of Equivalence:* Apart from the general problem of the TSE, cross-cultural surveys face an additional problem. In order to compare data from different nations/cultures, researchers have to be sure that the sampling and data collection is "spatially neutral" – or in the words of survey methodology: data (and even more important: analysis results) have to be equivalent. In other words: Researchers have to be sure that the various errors occur *in the same way* in *all* spatial units. Otherwise, if one observes a difference, one can never be sure, if this is a methodological artefact or if there are actual "real" differences.

   Unfortunately, it has been known for almost a century that the assumption of "spatially neutrality" is *not* true. On the contrary, these various errors vary spatially. E.g., nonresponse differs widely between countries. Therefore, since the 1920s survey methodology has evolved to the field of cross-cultural survey methodology which systematically analyzes how the various error types vary internationally and – drawing on these findings – tries to develop recommendations how to methodologically handle the problem of equivalence (Verba 1968; Scheuch 1989 [1968]; Rokkan 1970). Researchers have started systematizing these findings since the 1950s (Scheuch 1989). Unfortunately, although many of the problems of spatial comparison have been known for a long time, many findings have been either repeatedly lost with each generational change within science or have been only rarely applied in practical research, or, as Erwin Scheuch (1993 [1989], 172) put it:

> The real problem is not the methodology per se, but it is methodological in its consequences: what can be done to make methodological advances and practical experiences in comparative research more cumulative? Or phrasing the question both more realistically and more depressing: how can we make knowledge in this area cumulative at all?

Thus, in the first part of this paper I will summarize the current state of the debate for readers who might not yet be familiar with this field.

2) *The Problem of Demarcation:* In the second part of the paper I will discuss a problem which is also well-known to survey research but rarely discussed: As stated in section 1.1, when defining a target population, spatial units are defined. However, how these units are defined and what consequences this has on statistical results is seldomly discussed, or in Scheuch's (1993 [1989], 179) words:

> Usually, the nation-state is the geographical frame for sampling. Subsequently, the data collected in various nation-states are used as though that would also mean that the comparisons were cross-cultural and cross-societal as well. Whether the three meanings the geographical sampling frame could have – namely, nation, culture and society – do in fact coincide is in any particular case a substantive issue.

2. Equivalence

In cross-cultural research, sampling and data collection are typically organized separately for each spatial unit (such as quarters, cities, regions or nations) mainly for practical reasons: Very often spatial units (like nation states) differ in the legal and administrative requirements which have to be fulfilled to collect data, and often the availability of data collection infrastructure (including interviewers proficiency in the local language) varies locally. Finally, crossing borders in order to conduct research often involves increased costs (e.g. travel or telephone costs).

When comparing such spatial units later, data, methods and results have to be equivalent, and the major problems to be solved are still the same that Scheuch (1993 [1968], 107) identified five decades ago:

> Thus, the methodological problems of cross-cultural surveys were, and are still, seen primarily as problems of comparability only in a narrow sense. How to design a sample that in its concrete form can be used in all countries to be studied? How to standardize field work procedures internationally? How to ensure that questions are well translated? How to standardize categories, e.g. for background data?

2.1 Equivalence of Samples

In cross-cultural survey research, the ideal procedure for sampling within each spatial unit is still *random sampling* (Delmelle 2009; Heeringa and

O'Muircheartaigh 2010; Rippl and Seipel 2008, 63-8, 75-95; Bachleitner et al., 29, 44-59, 75-82, 89-91, 97-100, 119-25, 137-41, 151-61; Häder and Lynn 2007; Hoffmann 2007), as random samples allow for generalization using inferential statistics later. Ideally, the sample should represent the population. However, in sampling practice, there is almost always a slight difference in the distribution of the population and the *gross sample*. As long as this difference varies randomly between samples, it can be captured by the concept of *random error*. Although researchers never know for sure if such an error really occurred in the specific sample and how large it is, they can compute the likelihood of such an error, using either *significance tests* or *confidence intervals*.

However, it is frequently forgotten that significance tests and confidence intervals do not make generalizations from the random sample to the target population but to the *sampling frame*. In most cases, the sampling frame is not identical with the *target population*, i.e. there can be a *coverage error*, and this error is linked to the *mode of data collection* (Skjåk and Harkness 2003; Fu and Chu 2008; Rippl and Seipel 2008, 63-8, 75-7, 97-110; Pennell et al. 2010).

Imagine, for example, a researcher's target population are all adults currently living in Germany. The most common modes of data collection in Germany are postal, telephone, face-to-face, and online surveys. When conducting postal and face-to-face interviews, two traditional ways of creating a sampling frame would be getting addresses by using registry data or the random route procedure – automatically excluding all persons without a permanent home address from the sampling frame (e.g. the homeless, vagrants, illegals and persons staying long-term in public institutions such as hospitals and prisons). In case of telephone surveys, nowadays researchers would need a dual frame of combined landline and mobile phone numbers – excluding all persons without a phone from the sampling frame. When conducting online-interviews, researchers would either need a list of emails in order to contact respondents directly, or access them via a website or social network – all persons without access to the internet are automatically excluding from the sampling frame, as well as all persons to whose email researchers do not have access to or who do not use the website/social network the researchers uses for sampling.

Unfortunately, while coverage rates already vary widely for the different modes of data collection *within* one spatial unit, they vary even more widely *across* spatial units (Fuchs and Busse 2009). For example, as Table 1 illustrates, in 2011 statistically every inhabitant of countries in the Commonwealth, Europe or the Americas had at least one mobile phone, while in Africa only every second person owned a mobile phone and only about 14% of all individuals could be accessed via landline, meaning that at least one third of the African population would drop out of the sampling frame in telephone interviews. In online surveys, the situation is even more devastating: Every third European and nine out of ten Africans did not have access to the internet in 2011.

What makes the situation worse is that even within world regions and nation states access to specific technologies varies widely spatially (Fuchs and Busse 2009). For example, in Germany – which as a European country has relatively high IT coverage – in 2013 only one in five inhabitants of the city states Berlin, Bremen and Hamburg, but one in three of the more rural regions Sachsen-Anhalt and Mecklenburg-Vorpommern did not use the internet at all (Initiative D21 2013, 20).

Table 1: Percentage of Population Having Access to Information Technology in Various World Regions (2011)

World Region	Landlines	Mobile Telephones	Individuals using the Internet
Africa	14%	54%	12%
Arab States	96%	96%	30%
Asia and Pacific	14%	77%	26%
Commonwealth of Independent States	26%	147%	41%
Europe	42%	120%	69%
The Americas	29%	101%	53%

Notes: Regions in this table are based on the ITU BDT Regions, see: <http://www.itu.int/ITU-D/ict/definitions/regions/index.html>. If numbers are higher than 100%, this means that some persons have more than two or more items. Source: ITU World Telecommunication/ICT Indicators database, <http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx> (accessed February 16, 2014).

Now, this would be no problem at all, if those who drop out of the sampling frame shared the same characteristics (concerning the research question) with those within the research frame. Unfortunately, methodological research has shown that very often, this is not the case. For example, in the case of the internet, respondents can only have access, if companies or the state provide the necessary infrastructure for access which is often not the case in lowly populated or very poor areas. Further, respondents themselves have to be able to afford this technology (Baur and Florian 2008). Apart from extreme poverty, the most common reasons for non-usage are anxiety about security and data privacy, lack of knowledge or having someone else who handles the technology (Initiative D21 2013, 70-1).

The difference between the target population and sampling frame is not the only reason for bias (i.e. non-random errors): As in democracies, humans cannot be forced to participate in surveys, in almost every survey a large number of persons in the *random sample (= gross sample)* drops out due to *noncontact* or *refusals* (Couper and de Leeuw 2003; Holst 2003; Billiet et al. 2007; Blom et al. 2010; Stoop et al. 2010; Pennell et al. 2010). Again, noncontact and non-response rates vary widely spatially. For example, in the 2nd wave of the European Social Survey (ESS) in 2002, between 0.8% (Poland) and about 10% (Austria) of the persons in the gross sample could not be contacted (with Germany ranging in the middle with 6%). Of those contacted, between about 15%

(Hungary, Slovenia) and about 55% (Switzerland) refused to participate in the survey (again with Germany ranging in the middle with about 30%). As not all persons willing to participate were eligible for the sample, *response rates* ranged from about 33% (Switzerland) over about 50% (e.g. Germany) to about 80% (Greece) (Billiet et al. 2007).

As a result of *nonresponse*, the *net sample* in almost all surveys is much smaller than the *gross sample*. This does not matter as long as the reasons for nonresponse are random and response behaviour of those participating does not differ from those not participating. In this case (or if researchers use a complex sampling design such as a layered or stratified sampled), researchers have to decide, if and how to *weight data*, which implies that there might be an *adjustment error* (Behnke et al. 2010).

Unfortunately, as for coverage, this assumption is often not true, as nonrespondents often systematically differ from respondents, as analyses of reasons for nonresponse indicate. Typical reasons for nonresponse are that respondents (1) as a principle do never participate in surveys, (2) are not interested in this particular study, (3) have privacy issues with, or (4) a negative attitude towards this particular study/research institution, or (5) that simply the timing is bad. Again these reasons may vary spatially. For example, in the 2002 wave of the ESS, only one in four Ukrainians but three out of five Norwegians and Luxembourgians refused to participate in the study indicating not to be interested (Billiet et al. 2007). If the reason for nonresponse has to do with the research question, then it is very likely that data are *biased*.

So, in summary, bias due to coverage, noncontact and nonresponse varies widely across spatial units. To make matters more complicated, these error types depend on and are linked to *data collection modes*, and this connection between the various errors and data collection modes varies across spatial units, too (Skjåk and Harkness 2003; Fu and Chu 2008; Rippl and Seipel 2008, 63-8; 75-7, 97-110; Pennell et al. 2010) – and these links change over time, as usage behaviour of these technologies may change over time. As a consequence, it has to be decided for each research question separately, if it makes more sense to use the same or different sampling strategies and data collection modes for the spatial units under question. Therefore, Scheuch's (1993 [1968], 119) statement concerning sampling is as true as ever:

> Whatever the sampling procedure: either a sample is representative and will permit inferences as to the composition of the sampled population (usually a nation-state) or it is not. There is no one sample design that insures representativeness under all conditions; it can usually be achieved through a variety of designs. Of course there is often an optimal sample for any specific purpose and population – but only in terms of relative cost.

2.2 Equivalence of Concepts

Data collection and sampling are linked via the survey mode, as each mode needs a specific type of sample (e.g. telephone numbers for telephone surveys, addresses for postal surveys) and questionnaire type (e.g. one cannot use images and pictures in telephone interviews). Therefore, the questionnaire is typically developed parallel to sampling. During the process of developing instruments and data collection, a number of additional errors can occur which – again – can vary spatially and this threatens equivalence.

In order to assess the errors that can occur, it is important to keep in mind that survey researchers try to achieve comparability by standardizing the whole process of data collection and analysis (Baur 2006, 2009b): Every interviewed person is asked exactly the same questions in a given order, and they can only choose from a given number of answers. Not only is the questionnaire standardized, but also the interview situation and the way of recruiting possible interview partners. By using statistics, data analysis is standardized too, as the results will be the same regardless of who does the analysis and there are good practice-rules for each of these steps of the research process. Now, the question addressed by equivalence is: "What is the same?"

The first question to be asked in every survey is therefore, what *concepts* researchers want to measure and how to measure them equivalently (Friedrichs 1980, 73-89; Kromrey 2002, 72-7, 111-47; Rippl and Seipel 2008, 57-77, 94-5). The problem is that – even if the theoretical concept is clear to the researcher – it can mean very different things to different respondents in different societies. For example, such a seemingly simple thing as a "family" or "household" can mean anything from a pair of same-gender persons sharing a flat to a nuclear family consisting of husband, wife and kids to a large and extended tribal family. Therefore, even if two persons actually have the same attitude towards their children, researchers might get very different answers, if they ask respondents how important family is to them or who is part of their household, depending on which of the above definitions of family respondents use (for other examples see Warner and Hoffmeyer-Zlotnik (2009) or Baur (2009b)). As a rule of thumb, "[t]he concepts behind words are often delineated differently in different languages; the more abstract the concept, the greater the likelihood of differences" (Scheuch 1993 [1968], 109). Therefore, the first steps of survey development are typically semantic and dimensional analysis.

Semantic analysis aims at choosing the survey languages and assessing the meaning of concepts in each of these languages (Kromrey 2002, 72-7, 111-47; Pennell et al. 2010). Amongst others, things to consider are the emotional meaning of words, unintentional humour, things that cannot be expressed in some languages and political and status implications of words (Scheuch 1968; Baur 2006, 2009b). A very important aspect of semantic analysis is the *translation* process, for which complex procedures have been developed in order to

ensure equivalence. Typically, this involves translation from one language to another by one team of translators, extensive pretesting and back-translation by a different set of translators (Behr 2009; Harkness et al. 2010a; Willis et al. 2010; Dept et al. 2010).

Dimensional analysis starts from the observation that very often concepts are latent variables and therefore better measured not in their totality but split up in different indicators that can actually manifest. Then a question is phrased for each indicator. Therefore, dimensional analysis asks which aspects are part of the concept and how these indicators can be measured best (Kromrey 2002, 72-7, 169-86; Burzan 2008).

In fact, finding the right indicators for theoretical concepts is a research strand on its own within survey methodology which might require decades of collective effort, as the example of social class ("Soziale Schicht") illustrates. This concept is classically measured by at least three combined indicators: educational level, income and occupational prestige.

Now, the first question to answer is: Whose educational level, income and occupational prestige? For example, in Germany until the 1970s, women were mostly denied access to higher education. Therefore, highly educated men typically married women with lower educational levels, the cliché being the doctor marrying a nurse or a manager marrying a secretary. As institutions promoted the male breadwinner/female housewife model, middle-class women often became housewives after marriage (Baur 2008a). Thus, classically social class was measured for the household, using the man's educational level, income and occupational prestige. But what of today, when many German women are highly educated and career-oriented?

The second question to answer is how to operationalize each of these subconcepts comparably. For the example of social class, the major problem with income (Burzan 2008, 21-46) is item nonresponse as this is a sensitive question in many countries (Warner and Hoffmeyer-Zlotnik 2003). In contrast, with education (Burzan 2008, 71-88), national institutions have to be kept in mind during operationalization, which results in different suggestions for measuring this concept, e.g. ISCED (UNESCO 2003), CASMIN (Brauns et al. 2003) and the Hoffmeyer-Zlotnik/Warner-Matrix (Hoffmeyer-Zlotnik and Warner 2005).

Occupational prestige is even more complicated to measure, and there are several ways to do this. For one option, (1) a research team actually conducted cross-cultural surveys earlier that asked respondents to assess the prestige of specific occupations. For each country, these individual answers were aggregated, assigning each occupation an average prestige rating for that country, e.g. the Treiman-Prestige Scale (Ganzeboom and Treiman 2003). (2) Drawing on these earlier results, researchers can now ask respondents in every new survey what occupation they have. (3) After data collection, these open-ended questions are coded, e.g. using the International Standard Classification of Occupations (ISCO), resulting in a nominally scaled number for each occupa-

tion (Hofmann 2003; Geis and Hoffmeyer-Zlotnik 2001; Ganzeboom 2005). (4) Then, the next step is to transfer each occupational code into a number on a scale for occupational prestige, using the results from the Treiman-Prestige Scale.

The example of measuring social class illustrates that developing concepts for cross-cultural survey research is an intricate process where theoretical and methodological discussion are linked and often prior research is necessary to develop concepts later on. This example also shows why the survey life cycle has to be thought of as a process: Even when starting to develop the questionnaire, one has to decide how to analyze data afterwards and which problems to solve at which stage of the research process – in the case of measuring occupational prestige, a lot of analytical work is transferred from data collection (and the respondent) to data analysis (and researcher) in order to improve equivalence.

In the example of social class, the same set of indicators is used for each spatial unit. However, what makes dimensional analysis even more complicated is that it might be more adequate to use different indicators for each spatial unit (Beckers 2010; Davidov and Siegers 2010; Braun and Johnson 2010). For example, if one would like to measure prejudices against ethnic minorities, a classical indicator is whom people would not like as their neighbours. Now, due to the specific immigration history and history of racial conflicts, it makes sense to ask for different ethnic groups for each country. For example, in Germany, not wanting to live next to Jews or Turkish people would be a likely indicator for racism, while in the UK, one would probably better ask about not wanting to live next to Indians or Bangladeshi.

It also might be recommendable to compute indicators differently for each spatial unit (Beckers 2010; Davidov and Siegers 2010; Braun and Johnson 2010). For example, the Otte-Scale contains a set of five indicators for measuring "modernity" in West Germany. One of those indicators is religiosity, with non-religious persons typically being more modern than religious persons. The reason is that during the new social movements in the 1960s, religiosity became culturally strongly associated with traditionalism. However, in East Germany, this indicator does not work at all – on the contrary: During Socialism, religiosity was politically punished. Therefore, there was a high percentage of religious persons who propagated the democratic movement which in turn means that in East Germany, religiosity became culturally associated with modernity (Otte and Baur 2008).

In order to avoid a researcher's prejudice when entering concept development, survey research has a preference for answering these questions *empirically* (and not purely theoretically), the above study by Otte and Baur (2008) being an example. Using factor analysis, we could show which indicators were suitable for measuring modernity in different German regions.

So far, I have discussed conceptualizing and defining individual concepts. However, social research typically aims at linking these concepts. For example, researchers are very often interested in establishing *causal links between differ-*

ent concepts (Opp 2010). In order to do so, all individual variables that might influence the dependent variable have to be included into the survey. This raises the next question for cross-cultural survey research: Are the same causal relations to be expected in all spatial units? Due to very different institutional background, this is often not to be expected, meaning that it may make sense to ask a whole different set of questions for different spatial units (Baur 2004).

Blossfeld et al. (2004, 2006) demonstrate how equivalence can still be achieved in such cases: In the Globalife-Project, they chose a set of countries, using the most-dissimilar-cases-design. For each country, a country expert was chosen who first conducted a thorough case study of the institutional background of the specific country and then set up a theoretical concept of the causal relations to be expected for this country. Based on this model, a suitable data set for this country was selected and analyzed, resulting on country-specific models. In a final step, these country-specific models were treated as case studies, compared and systematized on a theoretical level. The result was a very concise study of what had an effect on individual careers in different career stages and in different strata of the social structure. This study also demonstrated which processes where global and which processes had country-specific effects.

2.3 Equivalence of Instruments

After having decided which concepts to measure with which indicator(s), the next step in conducting a survey is to develop *survey questions* and the actual *questionnaire* (especially deciding on *question order*, *questionnaire length* and *layout*), keeping in mind not only the general rules for questionnaire design, but also the rules for the specific *survey mode* (Behnke et al. 2010, 97-103, 107-12, 223-32; Harkness et al. 2003, 2010b; Blair and Piccinino 2005).

The specific problem to be solved in cross-cultural survey research (Pennell et al. 2010; Braun and Johnson 2010; Smith 2003; Johnson et al. 2005; Saris and Gallhofer 2007; Rippl and Seipel 2008, 63-8, 75-7, 110-14) for every *single question* in the survey is that every respondent in every spatial unit has to understand the same question in exactly the same way and be inspired to answer truthfully. The ideal is that persons with the same "objective" value also give exactly the same answer in the survey (Schwarz et al. 2010) in any language (Pan et al. 2010), e.g. that any person with two biological children gives the answer "two" when asked about the number of children or that any homosexual man answers that their partner is male when asked about this. Now, the whole reason why questionnaire design is so complicated is that this is very hard to achieve, as "[t]he format of questioning may carry cultural implications" (Scheuch 1993 [1968], 110).

This can result in *social desirability* operating differently in different cultures (Johnson and van de Vijver 2003). For example, cultures emphasize con-

cepts like honor, modesty and self-enhancement differently, which might influence the way questions are answered (Uskul et al. 2010). A well-known result is that, in some countries, questions on income are more sensitive than those on sexuality or politics. In other countries, the reverse effect can be observed. Respondents might either refuse to respond to sensitive topics (*item nonresponse*), drop out of the interview (*unit nonresponse*) or – even worse – lie to the interviewer.

Even if questions are not sensitive*, response styles* (Yang et al. 2010) differ widely between cultures: While in some countries, it is impolite to give a negative answer, in other countries it is expected to answer questions critically.

What makes matters even more complicated is that – even if respondents have exactly the same attitude towards an issue, share the same response style and are not affected by social desirability, their answers may vary depending on cultural context. This is often the case with attitudes, as these require assessing one's own perspective relative to one's *cultural and institutional contexts*. An example I have discussed in more detail in Baur (2006, 2009b) is the attitude towards income inequality. When answering the question "Do you agree or disagree that differences in income in [your country] are too large?", people compare their perceived ideal inequality with their perceived actual inequality, which differs largely among countries and therefore answers to this questions are not really comparable without knowing the context (Schwarz 2003; Braun 2003, 2006; Braun and Harkness 2005; Friedrichs and Nonnenbacher 2010).

Finally, not only cultural context but also local *media discourses* (Baur 2006, 2009b; Stoop 2007; Kepplinger 2008; Roessler 2008) influence response behaviour which is especially hard to handle, when comparing longitudinal surveys cross-culturally, as then current media discourses may overlay the general cultural pattern. This effect is even harder to handle, if media discourses occurred or changed during fieldwork.

It becomes clear from this list of problems that it is highly unlikely that exactly the same questions can be asked in all spatial units. Instead, the aim is to ask questions that are *functionally equivalent* (Braun 2006).

2.4 Equivalence of Field Work

Even if researchers have succeeded in drawing equivalent samples and developing equivalent questionnaires, one final challenge remains: Data now have to be collected, and *styles in conducting field work* vary widely culturally and between institutions (Koch and Blohm 2006; Koch 2002; Harkness et al. 2003; Pennell et al. 2010), although there are hardly any findings on this as comparing fieldwork is especially difficult, the reason being that fieldwork is often delegated to market and opinion research companies which do not want to face recourse claims.

Better documented is the fact that, despite all efforts of standardizing the interview situation, not only respondent but also *interviewer behaviour* varies cross-culturally and is highly related to the interview language (Kleiner and Pan 2006; Jäckle 2009; Pennell et al. 2010). However, at least for telephone interviews, this can be better controlled by supervision during the interview (Kleiner and Pan 2006).

2.5 Survey Methodology, Harmonization and Process Quality

As the discussion in this section has illustrated, it is extremely hard to gain comparable data on different spatial units. In order to come as close as possible to the goal of equivalence, cross-cultural survey research typically combines three measures:

1) *Survey Methodology* has become a research field of its own. As the examples given in this section illustrate, it aims at systematically identifying and analyzing the methodological problems that might arise during the research process and at developing solutions to avoid or (if this is not possible) handle these problems.
2) *Harmonization:* An important step for handling these problems is to develop procedures on how to organize the research process in order to make data and results as comparable as possible – a process called harmonization of surveys. Now, there are several ways of harmonization, and which one is most suitable depends on the specific research question (Granada et al. 2010):

   As always, it is best to think about what one wants to compare and how to achieve this at the beginning of the research process. One version of doing so is *ex-ante output harmonization*. Here concepts, variables and categories to be included in the final data set are defined (at the beginning of the research process). The exact way of how to measure these concepts and how data collection is organized may be decided upon individually by the research team responsible for each spatial unit – as long as the results are the same.

   The alternative is *input harmonization*. Here the same study with the same strict rules for all steps of the research process are defined. Note that – as explained above in detail – this may not always make sense because results may be worse (not better) than with ex-ante output harmonization.

   Now, harmonizing data at the beginning of the research process is not always possible, the two main reasons being (a) that sometimes, researchers only realize during data analysis that it would be interesting to compare data with other spatial units, and (b) that often – especially in longitudinal research – researchers conduct a secondary analysis of data that have been already collected (by somebody else). In this case, the option of *ex-post harmonization* remains. This means that researchers try to choose similar data sets and variables and make them comparable by data preparation. Unfortu-

nately, this is only possible, if all relevant concepts have been measured in all spatial units in a minimally equivalent way.

In order to leave the option of ex-post harmonization open for as many studies as possible, survey methodology has been developing so-called *socio-demographic and socio-economic standards* (Wolf and Hoffmeyer-Zlotnik 2003; Körner and Meyer 2005). They contain a set of key variables that are needed for almost any study (e.g. gender, age, social class, ethnicity) and recommendations how to measure them. The first recommendation is that all surveys worldwide should use these standards and thus keep the possibilities for future comparison as flexible as possible. The second recommendation is that – even if no standards exists – any survey should (after having defined central concepts) assess, if the same concepts have already been successfully used in other surveys and – if this is the case – copy them.

In order to avoid as many mistakes as possible, survey instruments should be thoroughly *pretested* (Prüfer 1996; Goerman 2006; Goerman and Caspar 2010). Additionally, *evaluation of survey instruments* has become a continuous process designed as a set of linked feedback-loops during the whole research process (Esposito and Rothgeb 1997; Prüfer et al. 2004; Braun and Johnson 2010). In order to ensure that researchers do not forget any important steps of ensuring *process quality*, standard procedures have been developed that work like check-lists of things not to forget (Lyberg and Stukel 2010; Rasinski 2008; Blair and Piccinino 2005).

Another prerequisite for future harmonization (and for validity assessment) is that the research process always and definitely needs to be *documented* (Mohler et al. 2010) and should be stored in a major *archive* such as the Data Archive for the Social Sciences at GESIS[2] or the Research Data Centres by the RatSWD.[3]

3) *Assessing Validity*: As elaborated above, it is almost impossible to achieve perfect equivalence. Moreover, in practical research despite researchers' best efforts, some things almost always go wrong. Therefore, the final step of any survey is assessing validity. As I have already stated above, traditionally, these various errors were handled individually, using concepts like reliability, validity, objectivity (Rammstedt 2010) and various sampling errors. In contrast, modern survey methodology tries to integrate them in the concept of the *Total Survey Error (TSE)* (Groves et al. 2009).

[2] <http://www.gesis.org>.
[3] <http://www.ratswd.de/publikationen>.

3. Demarcation

So far, I have discussed problems of cross-cultural survey research that have been handled within survey methodology. However, there is one problem that – although well-known to survey methodologists – is rarely discussed: Survey researchers prefer random sampling as a sampling strategy, as it allows for making use of inductive statistics as a strategy for generalizing results. However, a key element of this sampling strategy is defining a so-called target population with regards to contents, temporally and spatially. This means, that in research practice, cross-cultural survey research treats space as absolute space, i.e. as a container (Baur et al. 2014, in this HSR Special Issue): The sampling population is addressed as a specific set of people being located at a specific time on a specific spatial unit demarcated by borders from other spatial units. As discussed in section 1.2, for cross-cultural research, the world map is divided into areal units. In order to do so, in research practice typically politically defined administrational units are used, e.g. nation states, regions, cities or quarters. These units are often of very different geographical size and population. For example, compare the relative sizes of China, Germany and Luxembourg.

In this section, I will discuss what decisions we implicitly make during *demarcation* (Friedrichs 1977, 344-7; Carter 1980, 233-76). Namely, we choose (1) a level of aggregation and (2) a territory as frame of reference. We have to handle (3) the problem of fuzzy boarders and ambiguous allocation and (4) the problem of changing frames of reference. Using the example of changing frames of reference, I will continue to discuss what methodological problems arise from this problem which in geoinformation science is called *MAUP (Modifiable Areal Unit Problem)* (Wong 2009; Fotheringham et al. 2007, 236-40).

3.1 The Problem of Choosing the Right Level of Aggregation

The first problem that needs to be addressed during demarcation is that space is a multi-level phenomenon (Baur et al. 2014, in this HSR Special Issue): During demarcation, we basically *aggregate time-space-coordinates to spatial units*. The size of the spatial unit that the time-space-coordinates are aggregated into can vary, ranging from a room to a building, quarter, city, region, nation, or to the whole globe (Friedrichs 1977, 344-7). Which level of aggregation is adequate depends on the research question.

Therefore, from a methodological point of view, various distinct research traditions only vary in the level of aggregation of the spatial unit, i.e. target population. These traditions are (1) the analyses of social spaces ("Sozialraumanalysen") and social accounting analyses ("Sozialberichterstattung") (Mardorf 2006; Heineberg 2006, 150-9), (2) the comparison of indicators in different quarters of the same city within urban sociology (e.g. Koell 1986;

Häußermann and Kapphan 2004) (3) the comparison of different regions within a country such as Germany (e.g. Albrecht 2010; Wingerter 2009; Holst and Schupp 2009) and (4) international and cross-cultural comparative research.

3.2 The Problem of Choosing the Right Territory (Spatial Unit) as Target Population

Another problem to be solved during demarcation is which spatial units should be chosen as a frame of reference when dividing up physical space into spatial units (i.e. territories). E.g., Germany is politically divided into 16 federal states ("Bundesländer"). In most studies on regional disparities within Germany, Germany is demarcated into territories using these political-administrative regions and borders. As it is well known that respondent behaviour in surveys may vary widely between these regions, often they are also used as a stratification criterion when drawing stratified random samples. Note that these regions vary largely in sizes, with Bremen being the smallest region both geographically and population-wise, and Bavaria being the largest geographical region, while Northrhine-Westphalia has the largest population.

There are good reasons for sticking to this conventional way of demarcation, one of the most important reasons being that – as almost every study demarcates by using political-administrational borders – target populations and therefore results become more easily comparable.

A second important reason concerns a practical issue: Very often the production of natural data and process-produced data (Baur 2009a) follows the logic of these political-administrational borders, e.g. administrations collect data for "their" region which in turn might be either re-analysed or helpful as a base for research-elicited research. For example, when using public registries for sampling, researchers typically have to sample for each region separately. Also, as illustrated in section 2.4, legal and practical issues to consider for field work will very likely vary between politically regions. E.g., in Germany, every federal state has their own data protection law and procedures for using sensitive data – any study sticking within the borders of one political region is thus much easier to practically organize.

Finally, as Chan-Tack 2014 (in this HSR Special Issue) states, spatial units should be ecologically valid, i.e. relevant to the social processes under examination. Even today, it makes sense to assume that political-administrational regions fulfil this requirement at least partially, as typically these territories are stabilized by political and administrational institutions.

While in most cases, it makes sense to stick to the tradition of demarcating territories by using political-administrational regions, it is still worth to think about alternatives and what their implications would be.

An alternative would be demarcating Germany by using geographical regions and landscapes. Note that in this case, Germany would not be divided

into 16 but instead into four territories, and these territories would also not be equally distributed over the map with each federal state sharing a border with several others. Instead, the territories would run from north to south, starting with the very large and flat North German Lowlands, followed by the hilly Midlands, followed by much smaller and even hillier foothills of the Alps and ending with a small but mountainous stretch of the Alps.

Yet another possibility for territorial demarcation is using language. In this case, the continental West Germanic Languages can be divided into three major dialectical ranges: Dutch, Frisian and German, and the German dialectical range can be divided into three sub dialectical ranges: Low Saxon, West Central German and North Upper German.

Now, the interesting thing concerning demarcation is that these dialectical ranges do not correspond at all with political-administrational borders. Like in the case of landscapes, the territory of the Federal Republic of Germany (FRG) would be divided from North to South, albeit with different borders.

Additionally, the external borders do not follow political borders: Dutch dialects are both spoken in the nation states of the Netherlands and the FRG. German is spoken not only in the FRG but also in Switzerland and parts of Austria, Denmark and Italy. Finally, there are regions where German is not an official language but still a part of the population either speaks German as their native language or is bilingual.

The various ways of demarcating territories discussed so far have in common, that all spatial units are distinct and border each other. Each time-space-coordinate (and the people/things located there) is assigned to one (and only one) spatial unit. This need not always be the case. If for example, one uses wildlife parks, tourist regions or urbanization as criterion for defining territories and is interested in comparing these territories, there might as well be blank spaces on the map.

3.3 The Problem of Fuzzy Boarders and Ambiguous Allocation

So far, I have discussed demarcation as a process in which a number of theoretical decisions have to be made. The next steps seem rather simple: Each time-space-coordinate (and the people/things located there) is assigned to one (and only one) spatial unit. It is important that these allocations are distinct, as this is a typical prerequisite for almost all used statistical analysis procedures.

For example, with Hierarchical Linear Modelling (HLM) (Friedrichs 1977, 350-61; Pötschke 2006), researchers can analyze the interconnectedness and relative influences of different levels of aggregation. For example, they might want to know, if the national, regional or urban context is more important for an individual's employment and career changes. In order to be able to calculate these relative influences, it has to be clear for each individual, which (1) com-

munity, (2) region and (3) nation she lives in, and these three aggregation levels have to be distinct.

However, both in social reality and in research practice, allocations are often ambiguous and borders are often fuzzy. For example, some German regions (aggregation level 2) are also cities (aggregation level 1) – the so-called city-states of Berlin, Bremen and Hamburg thus belong to both spatial units.

This ambuiguity does not only apply vertically (for the aggregation level) but also horizontally for the frame of reference. For example, it is common in urban sociology and urban planning to compare quarters of a city. In Germany, it is possible for researchers to draw on public administrational data in order to calculate differences (e.g. concerning employment and career changes) between quarters and see how they correlate with other indicators for those quarters (e.g. voting behavior). Often, these data stem from different data sources (e.g. different administrations), and researchers also might combine administrational data with survey and other kinds of data (also maybe from different sources). These types of combinations of data sources depend on the target population always being the same. However, when looking into detail, it becomes clear that this is not always true. For example, in Berlin, the lower part of the street "Christburger Straße" belongs to the administrative district "Prenzlauer Berg" and to the voting district "Friedrichshain", meaning that – if one study uses the first and a second one uses the second definition of spatial units, the same persons will belong to different spatial units in these two studies, making them hard to compare.

3.4 The Problem of Changing Target Populations

To make things even more complicated, target populations (i.e. territories) chosen as frame of references for data collection may and do change over time. This is especially true to for political-administrational borders, as the German history in the last 200 years exemplifies.[4]

During the Congress of Vienna, the German Confederation was founded in 1815 by 39 principalities and city states. As Graph 5 illustrates, it covered a much larger territory than Germany in its present borders – a large part of these territories today belong to other states, e.g. Austria, the Czech Republic, Italy, Latvia, Lithuania, Poland, Slovakia and Slovenia. The German Confederation

[4] Note that the following examples only discuss the geographical region of Middle Europe and exclude German colonialism. Although German colonialism started later than that of other European states and while there were fewer colonies, Germany still had a colonial phase of territorial expansion in the 19th century, ending with World War I. The legal state of these territories (in comparison to German regions within Middle Europe) varied and is unclear which makes discussing them complex. The most important former German colonies today are parts of various African States, Venezuela, Papua New Guinea and Samoa.

was dissolved in 1866, and some of its former members remained in the North German Confederation.

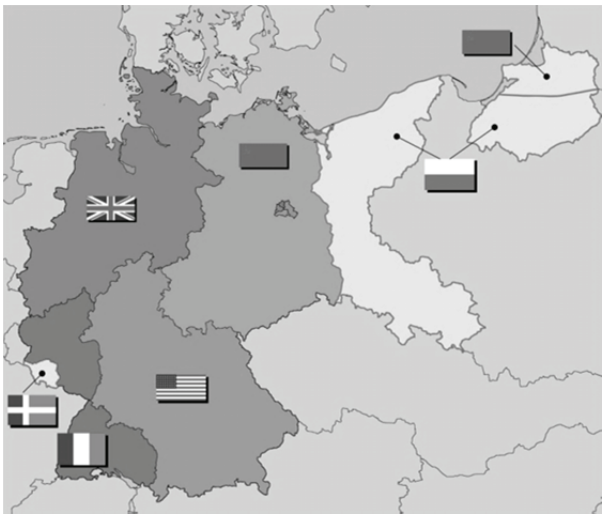
In 1870, a new German Confederation was founded for a short time, and then, in 1871, the German Reich was founded, which covered a much smaller geographical region than the German Reich. During World War I (1914-1918), there were many and continuous border shifts between the fighting parties.

Graph 5: Map of German Confederation 1815-1866

Source: Putzger – Historischer Weltatlas, 89th edition, 1965, modified by ziegelbrenner [GFDL, CC-BY-SA-3.0 or CC-BY-2.5], via Wikimedia Commons, 17 July 2006.

Then, in 1919, the Republic of Weimar was founded which (due to territorial losses during the war and as a result of the consecutive peace treaty), again was smaller than the territory of the German Reich. When the National Socialists came to power in 1933, Germany's territory at first stayed the same until 1937.

Graph 6: Occupation zone borders in Germany, 1947



Notes: The territories East of the Oder-Neisse line, under Polish and Soviet administration/annexation, are shown as light grey as is the likewise detached Saar protectorate. Berlin is the multinational area within the Soviet zone. Source: Based on map data of the IEG-Maps project (Andreas Kunz, B. Johnen and Joachim Robert Moeschl: University of Mainz) – <www.ieg-maps.uni-mainz.de>, modified by 52 Pickup [CC-BY-SA-2.5], via Wikimedia Commons, 10 August 2007.

However, when Germany started World War II it also started with territorial expansion by conquering many surrounding states, reaching its highest level of territorial expansion in 1942.

After being defeated in 1945, all areas east of the Oder-Neisse line were first administrated and then annexed by Poland and Russia, and the Saar region was likewise detached and declared a protectorate. The rest of Germany was divided up into an American, English, French and Russian Occupation Zone, and the capital (Berlin) was divided into four zones (Graph 6).

In 1949, the American, British and French Occupational Zones were merged and the Federal Republic of Germany (FRG) was founded. Several months later, the Russian Occupational Zone became the German Democratic Republic (GDR). The Saar Protectorate became a separate state in 1947 and only rejoined the FRG in 1957. These political processes resulted in a demarcation of German territories as illustrated in Graph 7.

Finally, after the fall of the Berlin Wall in 1989, Germany was unified, which resulted in the in the same external demarcation as shown in in Graph 7 but in a different internal demarcation – Germany today has 16 Federal Regions ("Bundesländer").

Graph 7: Federal Republic of Germany (FRG) and German Democratic Republic (GDR), 1949-1989

Notes: The area of the FRG is marked dark grey, the area of the GDR is marked in light grey. Berlin was divided into West and East Berlin (also marked in dark grey) with West Berlin belonging to the FRG and East Berlin belong to the GDR. Saarland only became part of the FRG in 1957. Source: Created by WikiNight [GFDL], via Wikimedia Commons, 17 July 2006.

3.5 Methodological Consequences of Demarcation

So far, I have illustrated that in order to define their target population, researchers have to demarcate space concerning aggregation level and type of territory. The borders of the so-defined spatial units may be fuzzy, ambiguous, overlap, and change over time without the name of the territory changing.

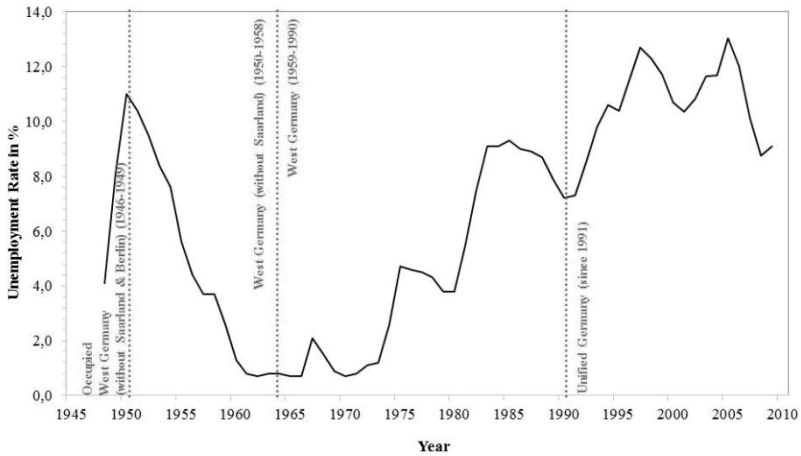
The problem is that these territorial changes (including resulting changes of the target population) may well distort statistical results. While Chan-Tack 2014 (in this HSR Special Issue) shows this, using cross-sectional data, I argue that the problem is augmented in longitudinal research, as the following example illustrates:

As is often done in social research, Graph 8 shows a long time series – in this case of the West German unemployment rate. In Germany, the standard interpretation of this graph is (Baur 2006, 2009b):

Due to the disruption of the German state after World War II, the economy completely broke down which – among other things – resulted in an enormously high unemployment rate. Then, with the founding of the FRG, there was a currency reform and the German Mark ("Deutsche Mark") was introduced, which started stabilizing the economy. In subsequent years, there was a series

of reforms inspired by ordoliberalism (the German version of neoliberalism) (Baur 2008a), which resulted into the "Wirtschaftswunder" in the 1960s. Then – due to the oil crisis and/or the expansion of the welfare state (there are alternative interpretations here) (Baur 2001, 2008a) – unemployment started rising again in the mid-1970s, rocking high during recessions and afterwards dropping slightly but never to its former lower level, until finally in the mid-1990s, unemployment rates reached insufferably high levels and approached the rates before 1950.

Graph 8: Unemployment Rates in West Germany, 1948-2009 including Border Changes



Source: Own Calculation based on data by DISI and iab.

Then, in 1997, the absolute number of unemployed persons hit the mark of 4 million people, which to German media and politicians points to a national trauma, as these were the numbers of unemployed persons during the World recession in the 1930s which led to Hitler's election in 1933. Thus, these numbers of unemployed persons were both seen as a weakness of the German economy and as a danger to democracy (Baur 2006, 2009b). The result were a series of reforms that retrenched the German welfare state, the most prominent being the "Hartz Reforms" which seriously reduced unemployment benefits for long-term unemployed in order to force them to take up any employment opportunity available (Baur 2008a, 2008b).

As I have discussed in detail in Baur (2004) on a general level and in detail for unemployment rates in Baur (2001), there are many methodological problems with computing such time series. The one problem I want to go into in more detail in this context is that during the time span under discussion (1948-

2009), there were three border shifts and thus changes of the spatial definition of the target population.

Especially unification in 1990 resulted into a massive increase of Germany's population size by 18% and of the geographical size by 43%, meaning that today, about 15% of Germans live in East Germany (which covers about 30% of Germany's territory) and 85% live in West Germany (which covers about 70% of Germany's territory).

Now, although the above time series treats both parts of the country as being part of the same target population since 1991, it is questionable, if this makes sense: Although the two parts of the country slowly assimilate, they still have two very different historical pasts, and even today, life styles and attitudes vary widely (Otte and Baur 2008; Baur 2013).

An alternative would be to treat each part of the country separately, but again, this is problematic from a methodological point of view, as both regions have by now shared a common history of almost 25 years, and within this time people have not only been sharing a common history but have moved back and forth between the two parts of the country.

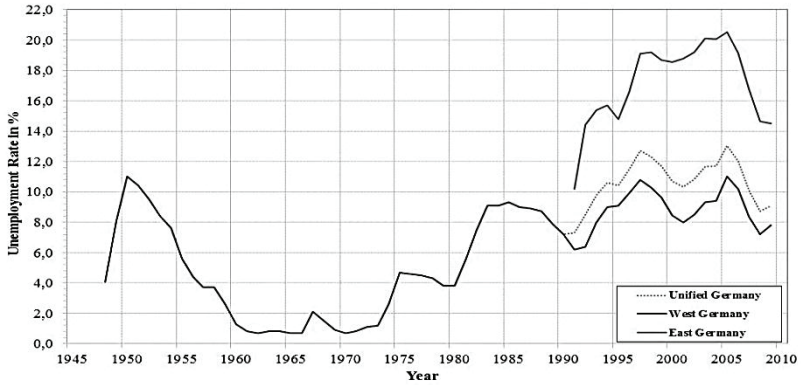
How to optimally solve this dilemma is a question for future research. For the moment, the most common good practice in survey research is to compute separate models for each territory, i.e. in this case (1) a model for unified Germany, as already illustrated in Graph 8, (2) a model for West Germany and (3) a model for East Germany.

The first thing to keep in mind with these kinds of calculations is that (as data collection takes time) there is a time-lag between territorial change and first available data. In the case of unemployment data provided by the Federal Agency of Unemployment, the time-lag is one year – as unification took place in 1990, the first data for unified Germany are available for 1991.

The second thing to consider in the case of Germany is that there are almost no reliable data available for East Germany due to the time of GDR, as most data were distorted by political processes. This means that – while it is possible to compute time series for West German unemployment rates for the time-span from 1948 up to today, comparable rates can only be calculated for East Germany form 1991 onwards.

Now, what are the results when computing separate models for each German territory? As Graph 9 illustrates, results draw a massively different picture: When looking at unemployment rates for unified Germany (Graph 8), observers get the impression that Germany as a whole suffered from a slacking economy and high unemployment rates – which, as stated above, led to the retrenchment of the German welfare state.

Graph 9: Unemployment Rates in Germany, 1948-2009 with Separate Calculations for East and West Germany after 1990



Source: Own Calculation based on data by DISI and iab.

When regarding West and East Germany separately, it becomes clear that unemployment rates stabilized in West Germany in the mid-1980s, dropping in some regions well below 6%. In contrast, the East German economy broke down in the 1990s and unemployment rocketed – in some regions well over 20% (Baur 2013). Now, these numbers differ not only widely, but disaggregating territories in this way would probably also have led to completely different political conclusions, as in neither case the retrenchment of the welfare state would have seemed a sensible option (Baur 2013): In West Germany, it was simply unnecessary. In East Germany, it did not help: The underlying idea of the Hartz reforms was a neoliberal notion that people were unwilling to work and thus had to be forced to take up less popular jobs by lowering benefits. However, the high rates of unemployment in East Germany imply that there are simply no jobs available – and where there are no jobs available, people cannot be forced to take them. It would therefore have been much more sensible to work on developing the East German economy.

4. Conclusion

In this paper, I have discussed two methodological problems that arise in cross-cultural survey research: the problem of equivalence and the problem of demarcation.

I have shown that the *problem of equivalence* cannot be solved but at least methodologically handled within survey methodology, using the model of the Total Survey Error (TSE). Still, current methodological research illustrates that it is almost impossible to gain comparative measurements from different spatial

units or even time-space-coordinates, as there is always a difference between measurement and reality (ideally captured by the TSE as a number). In other words: If results from different time-space-coordinates differ (e.g.: if respondents answer questions differently), researchers can never be absolutely sure, if this is really due to a difference in reality or in measurement. As this problem cannot be completely solved, in recent years survey research increasingly aims at methodological reflection and harmonization.

The discussion on equivalence also points to a methodological blind spot of theoretical concepts using relational concepts of space (Baur et al. 2014, in this HSR Special Issue) and mostly qualitative methods: Although these concepts and research traditions face exactly the same methodological problems, they have not yet systematically addressed them in methodological research. Facing them instead of avoiding them is a task for future research.

Another issue to be solved for future research is the *problem of demarcation*, which in geoinformation science is called *MAUP (Modifiable Areal Unit Problem)*. There are several suggestions for addressing this problem. However, none of these suggestions solves the problem completely or in all contexts.

For example, Chan-Tack 2014 (in this HSR Special Issue) suggests *collecting spatially sensitive data.* However, in survey research, secondary analyses become increasingly more common due to survey costs but also in order to conduct longitudinal research (Baur 2009a). As sometimes, data were collected decades ago, it is not always possible to collect these spatially sensitive data retrospectively.

Other concepts, e.g. Schräpler's (2009, 2011) suggestion for using spatial densities, compute *spatially sensitive indicators* that can make do without the concept of a sampling population. The problem of this approach is that researchers ideally need data for all time-space-coordinates in the research frame. For example, if the research frame was Germany, one would ideally need data on all people living in that geographical area (including the information where they are located, e.g. where they live). Apart from the legal problems in obtaining such data, in social research, this is normally only possible, if spatially sensitive public administrational data (Baur 2009a) are available. However, the whole point of surveys is that for most research questions and most spatial units such data do not exist – and thus research-elicited data have to be collected. And for practical reasons (time, costs, participants' willingness to cooperate) one typically only collects data on a sample, e.g. of 1,000 persons as representative of the German population. Although for most research questions this sample size is enough, these data would not suffice to compute spatially sensitive indicators.

Finally, Nosek and Netrdová 2014 (in this HSR Special Issue) suggest *alternating between these different perspectives of analysis*. Like Schräpler's suggestion, this approach has its limits in cases when there are not enough data for computing indicators.

Thus, it remains an open question for future research, if there are other options available for handling the problem of demarcation and in which cases which of these option is the best solution. It is very likely that – similar to the problem of equivalence – there will not be one best way but several good options and that researchers have to choose one of these options, depending on the goals of the specific study.

References

Albrecht, Susanne. 2010. Erwerbstätigkeit und Lebensbedingungen. In *Deutschlandatlas*, ed. Dirk Hänsgen, Sebastian Lentz and Sabine Tzschaschel, 37-60. Darmstadt: WBG.

Bachleitner, Reinhard, Martin Weichbold, and Wolfgang Aschauer. 2010. *Die Befragung im Kontext von Raum, Zeit und Befindlichkeit*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Baur, Nina. 2001. *Soziologische und ökonomische Theorien der Erwerbsarbeit*. Frankfurt a. M.: Campus.

Baur, Nina. 2004. Wo liegen die Grenzen quantitativer Längsschnittsanalysen? *Bamberger Beiträge zur empirischen Sozialforschung* 23. Bamberg.

Baur, Nina. 2006. Not Dancing with Memories. In *Dancing with Memory*, ed. Rosemary Block et al. Sydney.

Baur, Nina. 2008a. Konsequenzen des Verlusts des ganzheitlichen Denkens. In *Arbeitsmarkt und Sozialpolitik*, ed. Olaf Struck and Hartmut Seifert, 189-229. Wiesbaden: VS Verlag für Sozialwissenschaften.

Baur, Nina. 2008b. Regionale Ungleichheiten der Arbeitsmarktchancen. *zur debatte* 7: 26-7.

Baur, Nina. 2009a. Measurement and Selection Bias in Longitudinal Data. *Historical Social Research* 34 (3): 9-50.

Baur, Nina. 2009b. Memory and Data. In *Sociology of Memory*, ed. Noel Packard, 289-312. Newcastle: Cambridge Scholars.

Baur, Nina. 2013. Deutsche Wirtschaftsregionen. *SozBlog*. <http://soziologie.de/blog/?p=1223> (accessed March 25, 2013).

Baur, Nina, and Michael Florian. 2008. Stichprobenprobleme bei Online-Umfragen. In *Sozialforschung im Internet*, ed. Nikolaus Jackob et al., 106-25. Wiesbaden: VS Verlag für Sozialwissenschaften.

Beckers, Tilo. 2010. Comparing Comparisons for Cross-Validation (CC-CV). In *Komparative empirische Sozialforschung*, ed. Tilo Beckers et al., 363-89. Wiesbaden: VS Verlag für Sozialwissenschaften.

Behnke, Joachim, Nina Baur, and Nathalie Behnke. 2010. *Empirische Methoden der Politikwissenschaft*. Paderborn et al.: Schöningh

Behr, Dorothée. 2009. *Translationswissenschaft und international vergleichende Umfrageforschung.* Bonn: GESIS

Billiet, Jaak, Achim Koch, and Michel Philippels. 2007. Understanding and Improving Response Rates. In *Measuring Attitudes Cross-Nationally*, ed. Roger Jowell et al., 113-38. London et al.: Sage.

Blair, Johnny, and Linda Piccinino. 2005. The Development and Testing of Instruments for Cross-Cultural and Multi-Cultural Surveys. In *Methodological Aspects in Cross-National Research. ZUMA-Nachrichten Spezial* 11., ed. Jürgen H. P. Hoffmeyer-Zlotnik and Janet Harkness, 13-30. Mannheim: GESIS.

Blom, Annelies G., Annette Jäckle, and Peter Lynn. 2010. The Use of Contact Data in Understanding Cross-National Differences in Unit Nonresponse. In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, ed. Janet A. Harkness et al., 335-54. Hoboken: Wiley.

Blossfeld, Hans-Peter, Erik Klijzing, Melinda Mills, and Karin Kurz, eds. 2004. *Globalization, Uncertainty and Youth in Society*. London/New York: Routledge.

Blossfeld, Hans-Peter, Sandra Buchholz, and Dirk Hofäcker, eds. 2006. *Globalization, Uncertainty and Late Careers in Society*. London: Routledge.

Braun, Michael. 2003. Communication and Social Cognition. In *Cross-Cultural Survey Methods*, ed. Janet Harkness et al., 57-68. New York: John Wiley & Sons.

Braun, Michael. 2006. *Funktionale Äquivalenz in interkulturell vergleichenden Umfragen. Mythos und Realität*. Mannheim: GESIS

Braun, Michael, and Janet Harkness. 2005. Text and Context. *Methodological Aspects in Cross-National Research. ZUMA-Nachrichten Spezial* 11, ed. Jürgen H. P. Hoffmeyer-Zlotnik and Janet Harkness, 95-108. Mannheim: GESIS.

Braun, Michael, and Timothy P. Johnson, 2010. An Illustrative Review for Detecting Inequivalences. In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, ed. Janet A. Harkness et al., 375-94. Hoboken: Wiley.

Brauns, Hildegard, Stefanie Scherer, and Susanne Steinmann. 2003. The CASMIN Educational Classification in International Comparative Research. In *Advances in Cross-National Comparison*, ed. Jürgen H. P. Hoffmeyer-Zlotnik and Christof Wolf, 221-44. New York et al.: Kluwer.

Burzan, Nicole. 2008. *Quantitative Forschung in der Sozialstrukturanalyse*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Carter, Harold. 1980. *Einführung in die Stadtgeographie*. Berlin et al.: Borntraeger.

Chan-Tack, Anjanette M. 2014. The case for spatially-sensitive data: How data structures affect spatial measurement and substantive theory. *Historical Social Research* 39 (2): 315-46. doi: 10.12759/hsr.39.2014.2.315-346.

Couper, Mick P., and Edith D. de Leeuw. 2003. Nonresponse in Cross-Cultural and Cross-National Surveys. In *Cross-Cultural Survey Methods*, ed. Janet Harkness et al., 157-78. New York: John Wiley & Sons.

Davidov, Eldad, and Pascal Siegers. 2010. Comparing Basic Human Values in East and West Germany. In *Komparative empirische Sozialforschung*, ed. Tilo Beckers et al., 43-63. Wiesbaden: VS Verlag für Sozialwissenschaften.

De Vaus, David. 2008. Comparative and Cross-National Designs. In *The Sage Handbook of Social Research Methods*, ed. Pertti Alasuutari et al., 249-64. Los Angeles et al.: Sage.

Delmelle, Eric. 2009. Spatial Sampling. In *The Sage Handbook of Spatial Analysis*, ed. A. Stewart Fotheringham and Peter A. Rogerson, 183-206. London et al.: Sage.

Dept, Steve, Andrea Ferrari, and Laura Wäyrynen. 2010. Developments in Translation Verification Procedures in Three Multilingual Assessments. In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, ed. Janet A. Harkness et al., 157-76. Hoboken: Wiley.

Esping-Andersen, Gøsta. 1990. *The Three Worlds of Welfare Capitalism*. Cambridge/Oxford: Polity/Blackwell.
Esposito, James L., and Jennifer M. Rothgeb. 1997. Evaluating Survey Data. In *Survey Measurement and Process Quality*, ed. Lars Lyberg et al., 541-72. New York et al.: John Wiley & Sons.
Fotheringham, Stewart A., Chris Brunsdon, and Martin Charlton. 2007. *Quantitative Geography*. Los Angeles et al.: Sage.
Friedrichs, Jürgen. 1977. *Stadtanalyse*. Reinbek: Rowohlt.
Friedrichs, Jürgen. 1980. *Methoden empirischer Sozialforschung*. Opladen: Westdeutscher Verlag.
Friedrichs, Jürgen, and Alexandra Nonnenbacher. 2010. Welche Faktoren erklären Kontextinformationen? In *Komparative empirische Sozialforschung*, ed. Tilo Beckers et al., 468-97. Wiesbaden: VS Verlag für Sozialwissenschaften.
Fu, Yang-chih, and Yun-han Chu. 2008. Different Survey Modes and International Comparisons. In *The Sage Handbook of Public Opinion Research*, ed. Wolfgang Donsbach and Michael W. Traugott, 284-93. Los Angeles et al.: Sage.
Fuchs, Marek, and Britta Busse. 2009. The Coverage Bias of Mobile Web Surveys across European Countries. *International Journal of Internet Science* 4 (1): 21-33.
Ganzeboom, Harry B.G. 2005. On the Cost of Being Crude. In *Methodological Aspects in Cross-National Research. ZUMA-Nachrichten Spezial* 11, ed. Jürgen H.P. Hoffmeyer-Zlotnik, and Janet Harkness, 241-58. Mannheim: GESIS.
Ganzeboom, Harry B. G., and Donald J. Treiman. 2003. Three Internationally Standardized Measures of Comparative Research on Occupational Status. In *Advances in Cross-National Comparison*, ed. Jürgen H. P. Hoffmeyer-Zlotnik and Christof Wolf, 159-94. New York et al.: Kluwer.
Geis, A. J., and Jürgen H. P. Hoffmeyer-Zlotnik. 2001. Kompatibilität von ISCO-68, ISCO-88 und KldB-92. *ZUMA-Nachrichten* 48: 117 -38.
Goerman, Patricia L. 2006. An Examination of Pretesting Methods for Multicultural, Multilingual Surveys. In *Conducting Cross-National and Cross-Cultural Surveys. ZUMA-Nachrichten Spezial* 12, ed. Janet Harkness, 67-80. Mannheim: GESIS.
Goerman, Patricia L., and Rachel A. Caspar. 2010. Managing the Cognitive Pretesting of Multilingual Survey Instruments. In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, ed. Janet A. Harkness et al., 91-116. Hoboken: Wiley.
Granada, Peter, Christof Wolf, and Reto Hadorn. 2010. Harmonizing Survey Data. In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, ed. Janet A. Harkness et al., 315-34. Hoboken: Wiley.
Groves, Robert M., Floyd J. Fowler, Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2009. *Survey Methodology*. Hoboken: John Wiley & Sons.
Häder, Sabine, and Peter Lynn. 2007. How Representative Can a Multi-nation Survey Be? In *Measuring Attitudes Cross-Nationally*, ed. Roger Jowell et al., 33-52. London et al.: Sage.
Harkness, Janet, Fons van de Vijver, and Timothy P. Johnson. 2003. Questionnaire Design in Comparative Research. In *Cross-Cultural Survey Methods*, ed. Janet Harkness et al., 19-34. New York: John Wiley & Sons.

Harkness, Janet, Ana Villar, and Brad Edwards. 2010a. Translation, Adaption and Design. In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, ed. Janet A. Harkness et al., 117-40. Hoboken: Wiley.

Harkness, Janet A., Brad Edwards, Sue Ellen Hansen, Debra R. Miller, and Ana Villar. 2010b. Designing Questionnaires for Multipopulation Research. In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, ed. Janet A. Harkness et al., 33-57. Hoboken: Wiley.

Häußermann, Hartmut, and Andreas Kapphan. 2004. Berlin. In *An den Rändern der Städte*, ed. Hartmut Häußermann, Martin Kronauer and Walter Siebel, 203-34. Frankfurt a.M.: Suhrkamp.

Heeringa, Steven G., and Colm O'Muircheartaigh. 2010. Sampling Designs for Cross-Cultural and Cross-National Survey Programs. In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, ed. Janet A. Harkness et al., 251-68. Hoboken: Wiley.

Heineberg, Heinz. 2006. *Stadtgeographie*. Paderborn. Schöningh.

Hering, Linda, and Robert F. Schmidt. 2014. Fallstudie. In *Handbuch Methoden der empirischen Sozialforschung*, ed. Nina Baur and Jörg Blasius. Wiesbaden: Springer VS Verlag für Sozialwissenschaften.

Hoffmann, Hermann. 2007. Kombinierte Stichproben für Telefonumfragen – Ansätze in Europa. In *Mobilfunktelefonie*, ed. Siegried Gabler and Sabine Häder, 47-58. Mannheim: GESIS.

Hoffmeyer-Zlotnik, Jürgen H. P., and Warner, Uwe. 2005. How to Measure Education in Cross-National Comparison. In *Methodological Aspects in Cross-National Research. ZUMA-Nachrichten Spezial* 11., ed. Jürgen H. P. Hoffmeyer-Zlotnik and Janet Harkness, 223-40. Mannheim: GESIS.

Hofmann, Eivind. 2003. International Statistical Comparisons of Occupational and Social Structures. In *Advances in Cross-National Comparison*, ed. Jürgen H. P. Hoffmeyer-Zlotnik and Christof Wolf, 137-58. New York et al.: Kluwer.

Holst, Christian. 2003. The Validity of Income Measurements in Comparative Perspective. In *Advances in Cross-National Comparison*, ed. Jürgen H. P. Hoffmeyer-Zlotnik and Christof Wolf, 367-88. New York et al.: Kluwer.

Holst, Elke, and Jürgen Schupp. 2009. Situationen und Erwartungen auf dem Arbeitsmarkt. In *Datenreport* 2008, ed. Statistisches Bundesamt et al., 122-8. Bonn: Statistisches Bundesamt et al.

Initiative D21 (2013): *D21 – Digital-Index*. URL: <www.initiatived21.de> (accessed February 16, 2014).

Jäckle, Annette. 2009. Dependent Interviewing. In *Methodology of Longitudinal Surveys*, ed. Peter Lynn, 93-112. Chichester: Wiley.

Johnson, Timotyh P., and Fons van de Vijver. 2003. Social Desirability in Cross-Cultural Research. In *Cross-Cultural Survey Methods*, ed. Janet Harkness et al., 195-206. New York: John Wiley & Sons.

Johnson, Timothy P., Young Ik Cho, Allyson Holbrook, Diane O'Rourke, Richard Warnecke, and Chávez Noel. 2005. Cultural Variability in the Effects of Question Design Features on Respondent Comprehension. In *Methodological Aspects in Cross-National Research. ZUMA-Nachrichten Spezial* 11., ed. Jürgen H. P. Hoffmeyer-Zlotnik and Janet Harkness, 65-78. Mannheim: GESIS.

Kepplinger, Hans Matthias. 2008. Effects of News Media on Public Opinion. In *The Sage Handbook of Public Opinion Research*, ed. Wolfgang Donsbach and Michael W. Traugott, 192-204. Los Angeles et al.: Sage.

Kleiner, Brian, and Yuling Pan. 2006. Cross-Cultural Communication and the Telephone Survey Interview. In *Conducting Cross-National and Cross-Cultural Surveys. ZUMA-Nachrichten Spezial* 12, ed. Janet Harkness, 81-90. Mannheim: GESIS.

Koch, Achim. 2002. 20 Jahre Feldarbeit im ALLBUS. *ZUMA-Nachrichten* 51: 9-37.

Koch, Achim, and Michael Blohm. 2006. Fieldwork Details in the European Social Survey 2002, and 2003. In *Conducting Cross-National and Cross-Cultural Surveys. ZUMA-Nachrichten Spezial* 12, ed. Janet Harkness, 21-52. Mannheim: GESIS.

Koell, Johannes. 1986. Von Volkszählungsdaten zu städtebaulichen Kartierungen. *Städte und Gemeindebund* 41 (8): 347-50.

Körner, Thomas, and Iris Meyer. 2005. Harmonising Socio-Demographic Information in Household Surveys of Official Statistics: Experiences from the Federal Statistical Office Germany. In *Methodological Aspects in Cross-National Research. ZUMA-Nachrichten Spezial* 11, ed. Jürgen H. P. Hoffmeyer-Zlotnik and Janet Harkness, 149-62. Mannheim: GESIS.

Kromrey, Helmut. 2002. *Empirische Sozialforschung*. Opladen: Leske + Budrich.

Lyberg, Lars, and Diana Maria Stukel. 2010. Quality Assurance and Quality Control in Cross-National Comparative Studies. In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, ed. Janet A. Harkness et al., 227-51. Hoboken: Wiley.

Mardorf, Silke. 2006. *Konzepte und Methoden von Sozialberichterstattung.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Mohler, Peter et al. 2010. A Survey Process Quality Perspective on Documentation In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, ed. Janet A. Harkness et al., 299-324. Hoboken: Wiley.

Opp, Karl-Dieter. 2010. Kausalität als Gegenstand der Sozialwissenschaften und der multivariaten Statistik. In *Handbuch der sozialwissenschaftlichen Datenanalyse*, ed. Wolf Christof and Best Henning, 9-38. Wiesbaden: VS Verlag für Sozialwissenschaften.

Otte, Gunnar, and Nina Baur. 2008. Urbanism as a Way of Life? In *ZfS* 37 (2): 93-116.

Pan, Yuling, Ashley Landreth, Hyunjoo Park, Marjorie Hidsdale-Shouse, and Alisú Schoua-Glusberg. 2010. Cognitive Interviewing in Non-English-Languages: A Cross-Cultural Perspective. In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, ed. Janet A. Harkness et al., 91-116. Hoboken: Wiley.

Pennell, Beth-Ellen, Janet A. Harkness, Rachel Levenstein, and Martine Quaglia. 2010. Challenges in Cross-National Data Collection. In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, ed. Janet A. Harkness et al., 269-98. Hoboken: Wiley.

Pennings, Paul. 2009. Qualitative Comparative Analysis (QCA) and Related Techniques. In *Methoden der vergleichenden Politik- und Sozialwissenschaft*, ed. Susanne Pickel et al., 365-88. Wiesbaden: VS Verlag für Sozialwissenschaften.

Pötschke, Manuela. 2006. Mehrebenenanalyse. In *Methoden der Politikwissenschaft*, ed. Joachim Behnke, Thomas Gschwend, Delia Schindler and Kai-Uwe Schnapp, 167-79. Baden-Baden: Nomos.

Prüfer, Peter. 1996. Verfahren zur Evaluation von Survey-Fragen. *ZUMA-Nachrichten* 20 (39): 95-116.

Prüfer, Peter, Margrit Rexroth, and Floyd Jackson Jr. Fowler, eds. 2004. QUEST 2003. *ZUMA-Nachrichten Spezial* 9. Mannheim: GESIS.

Rammstedt, Beatrice. 2010. Reliabilität, Validität, Objektivität. In *Handbuch der sozialwissenschaftlichen Datenanalyse*, ed. Christof Wolf and Henning Best, 239-58. Wiesbaden: VS Verlag für Sozialwissenschaften.

Rasinski, Kenneth A. 2008. Designing Reliable and Valid Questionnaires. In *The Sage Handbook of Public Opinion Research*, ed. Wolfgang Donsbach and Michael W. Traugott, 361-63. Los Angeles et al.: Sage.

Rippl, Susanne, and Christian Seipel. 2008. *Methoden kulturvergleichender Forschung*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Roessler, Patrick. 2008. Agenda-Setting, Framing and Priming. In *The Sage Handbook of Public Opinion Research*, ed. Wolfgang Donsbach and Michael W. Traugott, 205-17. Los Angeles: Sage.

Rohlfing, Ingo. 2009. Vergleichende Fallanalysen. In *Methoden der vergleichenden Politik- und Sozialwissenschaft*, ed. Susanne Pickel et al., 133-52. Wiesbaden: VS Verlag für Sozialwissenschaften.

Rokkan, Stein. 1970 [1993]. Cross-cultural, cross-societal and cross-national research. Reprinted in *Historical Social Research* 18 (2): 6-54. URN: <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-32835>.

Saris, Willem E., and Irmstraud Gallhofer. 2007. Can Questions Travel Successfully? In *Measuring Attitudes Cross-Nationally*, ed. Roger Jowell et al., 53-78. London et al.: Sage.

Scheuch, Erwin K. 1993 [1968]. The cross-cultural use of sample surveys. Reprinted in *Historical Social Research* 18 (2): 104-38. URN: <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-32817>.

Scheuch, Erwin K. 1993 [1989]. Theoretical implicatons of comparative survey research. Reprinted in: *Historical Social Research* 18 (2): 172-95. URN: <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-3279>.

Schräpler, Jörg-Peter. 2009. Verwendung von SGBII-Dichten als Raumindikator für die Sozialberichterstattung am Beispiel der „sozialen Belastung" von Schulstandorten in NRW – ein Kernel Density Ansatz. *Statistische Analysen und Studien NRW* 57: 3-28.

Schräpler, Jörg-Peter. 2011. Konstruktion von SGBII-Dichten als Raumindikator und ihre Verwendung als Indikator im Rahmen der Sozialberichterstattung am Beispiel der „sozialen Belastung" von Schulstandorten in NRW – ein Kerndichte Ansatz. *AStA Wirtschafts- und Sozialstatistisches Archiv* 5 (2).

Schwarz, Norbert. 2003. Culture-Sensitive Context Effects: A Challenge to Cross-National Surveys. In *Cross-Cultural Survey Methods*, ed. Janet Harkness et al., 93-100. New York: John Wiley & Sons.

Schwarz, Norbert, Daphna Oyserman, and Emilia Peytcheva. 2010. Cognition, Communication and Culture. Implications for the Survey Response Process. In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, ed. Janet A. Harkness et al., 177-90. Hoboken: Wiley.

Skjåk, Knut Kalgraff, and Janet Harkness. 2003. Data Collection Methods. In *Cross-Cultural Survey Methods*, ed. Janet Harkness et al., 179-93. New York: John Wiley & Sons.

Smith, Tom W. 2003. Developing Comparable Questions in Cross-National Surveys. In *Cross-Cultural Survey Methods*, ed. Janet Harkness et al., 69-92. New York: John Wiley & Sons.

Stoop, Ineke. 2007. If it bleeds, it leads. The Impact of Media-Reported Events. In *Measuring Attitudes Cross-Nationally*, ed. Roger Jowell et al., 95-112. London: Sage.

Stoop, Ineke, Jaak Billiet, Achim Koch, and Rory Fitzgerald. 2010. *Improving Survey Response*. Chichester: Wiley

UNESCO. 2003. International Standard Classification of Education, ISCED 1997. In *Advances in Cross-National Comparison*, ed. Jürgen H. P. Hoffmeyer-Zlotnik and Christof Wolf, 195-220. New York et al.: Kluwer.

Uskul, Ayşe, Daphna Oyerman, and Norbert Schwarz. 2010. Cultural Emphasis on Honor, Modesty, or Self-Enhancement: Implications for the Survey-Response-Process. In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, ed. Janet A. Harkness et al., 191-202. Hoboken: Wiley.

Verba, Sidney. 1968. The uses of survey research in the study of comparative politics. Reprinted 1993 in *Historical Social Research* 18 (2): 55-103 <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-32826>.

Warner, Uwe, and Jürgen H. P. Hoffmeyer-Zlotnik. 2003. How to Measure Income. In *Advances in Cross-National Comparison*, ed. Jürgen H. P. Hoffmeyer-Zlotnik and Christof Wolf, 307-24. New York et al.: Kluwer.

Warner, Uwe, and Jürgen H. P. Hoffmeyer-Zlotnik. 2009. Das Konzept des ‚privaten Haushaltes' in international vergleichenden Umfragen und dessen Messung. In *International vergleichende Sozialforschung*, ed. Birgit Pfau-Effinger et al., 207-28. Wiesbaden: VS Verlag für Sozialwissenschaften.

Willis, Gordon B., et al. 2010. Evaluation of a Multistep Survey Translation Process. In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, ed. Janet A. Harkness et al., 141-56. Hoboken: Wiley.

Wingerter, Christian. 2009. Arbeitsmarkt. In *Datenreport 2008*, ed. Statistisches Bundesamt et al.: 109-21. Bonn: Statistisches Bundesamt et al. <https://www.desttis.de/DE/Publikationen/Datenreport/Downloads/Datenreport2008.pdf?__blob=publicationFile> (accessed January 28, 2014).

Wolf, Christof, and Jürgen H. P. Hoffmeyer-Zlotnik, eds. 2003. *Advances in Cross-National Comparison*. New York et al.: Kluwer.

Wong, David. 2009. The Modifiable Areal Unit Problem (MAUP). In *The Sage Handbook of Spatial Analysis*, ed. A. Stewart Fotheringham and Peter A. Rogerson, 105-23. London et al.: Sage.

Yang, Yongwei, Janet A. Harkness, Tzu Yun Chin, and Ana Villar. 2010. Response Styles and Culture. In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, ed. Janet A. Harkness et al., 203-26. Hoboken: Wiley.