The cross-cultural use of sample surveys: problems of comparability

Scheuch, Erwin K.

Historical Social Research, Vol. 18 — 1993 — No. 2, 104-138

The Cross-Cultural Use of Sample Surveys: Problems of Comparability

Erwin K. Scheuch*

Abstract: This article (first published in 1968) deals with the following problems of cross-cultural research: change in the identification of problem areas; question meaning and problems of verbal communication; equivalence of indicators; the respondent as a unit in design and analysis; the usage of »culture« in cross-cultural surveys; adminstrative and diplomatic problems; and some social effects of comparative social research.

1. Progress as a new problem identification

A review of publications based on cross-cultural surveys makes it at first appear doubtful that there has been definite progress in research methodology and practice.[1] Published research certainly does not show a neat pattern of continuous ascent to ever higher levels of methodological sophistication. However, I hope to show that nevertheless such progress exists in a specific - although somewhat frustrating - way.

In surveying surveys, one can point to some technological innovations. I think, however, that the major progress has been to increase the awareness of the real sources of difficulties. In this way, the use of surveys in cross-cultural comparisons influences our understanding of research techniques and methodology in general. This is then my main theme: progress as changes in the awareness of problem areas, as the spreading realization that an earlier identification of the sources of difficulties was much too simple.

Up to this day, the difficulties encountered in cross-cultural comparisons tend to be perceived as problems of research technology. However, it is now increasingly realized that the main problems are methodological in the more limited sense of this term.

* Originally published in: Stein Rokkan, ed., Comparative Research across Cultures and Nations, Paris/The Hague: ISSC/Mouton 1968, pp. 176-209. We are grateful to the first editors for kindly permitting us to republish this article.

[1] Compare S. Rokkan, S. Verba, J. Viet and E. Almasy, *Comparative Survey Analysis* (Paris, Mouton, 1968).

This change in the perception of problem areas is being extended also to the somewhat self-contained field of survey research. Cross-cultural surveys are beginning to be viewed as just one technical form of the method of comparison, and under this perspective difficulties in research technology often lead to the realization that actually both method and theory may be wanting. Regress to theory and method may still frequently permit the solution of problems that, understood merely as problems of research technology, were difficult to handle, and this in turn will advance substantive theory.

This orientation is just beginning to be acceptable - and perhaps this statement is more an expression of hope than a description of reality. In the cross-cultural use of surveys, the standards of technical perfection are already below the standards for good work within a particular country. And in the phase of interpretation it sometimes seems as though cross-cultural comparisons via survey research were less a source for new insight and more an empirical buttressing of pre-conceived notions about differences among social and political systems.[2]

However, this is merely a criticism of the behavior of researchers, and not an indictment of a method. Quite the contrary: if a better understanding of methodological problems in the cross-cultural use of surveys is achieved, this should lead to an immediate and dramatic advancement in our social science knowledge.

2. Changes in the identification of problem areas

The type of cross-cultural research we are discussing here, and its problems, are mainly a phenomenon of the period since World War H However, this should not lead to the frequent misperception that cross-cultural research as such is something new. In the 19th century sociology was largely comparative in orientation, relying both on historical and on cross-cultural comparisons. During the second half of that century a major empirical basis for sociological treatises was material collected by ethnographers, and while this material was cross-cultural by virtue of the subject matter, it remains significant that sociological theory was then understood to require empirical information from this inherently cross-cultural discipline. And what little social research proper was carried out then was likewise comparative in orientation.

When Le Play collected family budgets to analyze the structure of the family, he checked his findings in France against data from Germany as a matter of course;[3] as a matter of principle, Durkheim was not satisfied to test his

[2] Two examples are in my opinion G. A. Almond and S. Verba, *The Civic Culture* (Princeton, N.J., Princeton University Press, 1963) and D. V. McGranahan, *A Comparison of Social Attitudes Among American and German Youth*, *Journal of Abnormal and Social Psychology,* 41 (1946), pp. 245-256.

[3] F. Le Play, *Les Ouvriers européens,* 2e éd. (Paris, Marne, 1877).

theorems on anomie just by data from France: in his *Rules of Sociological Method* he declared the comparative method central to sociology. In this methodological orientation, Durkheim considered himself a direct descendant of John Stuart Mill; like Mill he maintained that the social sciences were essentially observational disciplines, and that in disciplines of that character, systematic comparisons were needed as an analogue of experimentation. Cross-cultural comparisons were of course not the only expression of this methodological orientation, but were widely considered to offer evidence of an especially convincing kind.

The present cross-cultural surveys are not a direct continuation of this early emphasis, and this becomes obvious in the current methodological discussions on comparative research. Comparative sociology, and especially comparative empirical research, had to be discovered all over again. The very development of empirical research as we understand it today seems partly responsible for this hiatus in tradition.

The Chicago school of social research (though not so much its founders, as, e.g., Robert Ezra Park) was fascinated with its own immediate environment. It placed an emphasis on the immediately *observable,* and developed techniques particularly suited to demonstrating within-culture variations. This social research was problem-oriented in the sense that 'muckraking' was. Quite understandably, a strong connection with social work came about, and in turn social workers were most influential in developing some of our techniques.[4] This is especially true for the prevailing rules of procedure in interviewing, the set of 'do's and don'ts', something which I would like to call the 'folklore of interviewing'. The same exclusive concern with one's own society was of course characteristic of consumer research and opinion polling - two major sources of survey technology.

In the familiar and unreflected environment of one's own society, a methodological reflection on the problems of relating sensory impressions to generalized statements appeared unnecessarily 'theoretical' (after all, the evidential meaning of indicators appeared obvious). Progress in research was then understood as a more reliable collection of observations. Given the further fact that empirical research was no longer carried out mainly by the scholar himself but that - for a variety of reasons - the collection and later the analysis of sensory impressions had to be delegated to relatively unskilled helpers, an emphasis on the codification of procedures was natural.

This emphasis was in my opinion a real progress. During the earlier prevalence of more 'fundamental' questions, the reliability and empirical validity of actual research procedure was without doubt treated too lightly. Nevertheless,

[4] Thus, the first methodological treatises on interviewing were prepared by social workers, as was the first bibliography of methodological writings on interviewing and field observation. Cf. B. V. Moore, *The Personal Interview. An Annotated Bibliography* (New York, Russell Sage Foundation, 1928).

in reflecting about current cross-cultural comparisons, it is important to realize that during the late twenties research methodology became largely the discussion of research technology. As such, it was later exported to other countries. This emphasis on technology is probably nowhere more pronounced than in such a highly codified procedure as survey research today.

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? These are the prevailing ways of identifying problems in surveys outside one's own culture.

Some progress has been made in answering such questions, and this experience seemed to call for fewer adjustments in our technology than was originally expected - e.g., one can conduct an interview in Europe pretty much in the same way one had learned to do in America. Actually, one of the most surprising experiences in doing cross-cultural survey work is that so much of the technology of the datacollection phase appears transferable.

Yet this happy experience obscures the fact that in a more strictly methodological sense the difficulties of cross-cultural surveys are indeed quite formidable. Genuine differences in social structures exert themselves even if they permit a partial transfer of one's technology. Becoming aware of this forces us to spell out assumptions we do not need to spell out when working within our own culture. Intuition and research folklore are not necessarily a great help in analyzing data not from one's own society. In handling technical problems in cross-cultural surveys - and especially in the secondary analysis of survey material - we have once more to change from research technicians into methodologists.

I now want to draw on some concrete examples in order to show how a better awareness of type of problems can help to turn the cross-cultural survey into the powerful tool for a generalizing social science that we have every reason to expect it to be.

3. Problem area I: Question meaning and problems of verbal communication

Question wording was the first problem that attracted attention as *the* supposedly major difficulty in international surveys. However, if one contrasts the discussion of the Time International Survey in the 1948 volume of the *Public Opinion Quarterly*[5] with the summaries of the 13th AAPOR conference in the

[5] D. Wallace, J. L. Woodward, E. Stern, M. Barioux, and H. Ylvisaker, 'Experiences in

same journal ten years later,[6] a shift in emphasis will be obvious. The first concern had quieted down, and standard techniques were developed that gave researchers the feeling that they were in control of the difficulty. The situation is, by the way, analogous to the decline in the excitement originally generated by Hadley Cantril's *Gauging Public Opinion,* when he reported dramatic examples of the effects of non-neutral question-wording.

The chief technique (especially for academic researchers) for controlling the correctness of translations has been borrowed from cultural anthropologists. Ideally, a foreign bilingual translates a questionnaire into his language and a bilingual from the researcher's home country prepares a retranslation; this version is then checked against the original wording. While this is undoubtedly a fine technique to check the ability of translators, it does little to control the chief problem in question wording: equivalence of meaning.

Commercial researchers tend to be more sophisticated here and to take less refuge in the comfort that they were able to achieve literally or idiomatically correct translations. By now, a commercial institute making an international survey quite often has a master questionnaire drawn up by the particular unit handling the client. The individual participating agencies then translate and modify (!) the instrument. If time and money permit, a general conference is called, where changes in questions are discussed that appeared necessary to achieve equivalence of meaning. Essentially, this is a qualitative procedure in which the individual researcher uses his best judgment as a criterion.

However, in too many cases there has been insufficient awareness of the source of difficulties when developing equivalent forms of questions. These difficulties arise from the relation of language to reality, from structural differences between societies, which are reflected in translation problems.

I. Let me point to some experiences which led me to postulate a number of problem factors; these are presumed to account for most of the difficulties experienced in question wording for international surveys. If we should succeed in identifying the main problem factors in international surveys, we should at the same time learn something more about the meaning of questions in general.

In 1954 in Italy and France, Gabriel Almond found the term 'Communist Party' referred to quite different political entities and had a different content for the respondents.[7] If the term 'communist' is used in questions asking for membership and voting, this difference in the real meaning of the term obviously does not cause problems; but if a question using the term 'communist' is used

---

the Time' International Study: A Symposium', *Public Opinion Quarterly,* 12(4), (Winter, 1948-49), pp. 709-722.

[6] 'Proceedings of the Thirteenth Conference on Public Opinion Research', *Public Opinion Quarterly,* 22(2), (1958), pp. 170-215.

[7] G. A. Almond, et al., *The Appeals of Communism* (Princeton, N. J., Princeton University Press, 1954).

as an indicator for a system of attitudes and beliefs, problems do certainly arise in a cross-cultural comparison.

Using a semantic differential, Hofstätter explored the meaning of the word 'lonesomeness'.[8] He could show that this English word did not correspond to the presumed German equivalent, 'Einsamkeit'; even between the American and the British usage there were strong differences in the meaning of this term. An attitude test that included the term 'lonesome' would undoubtedly have a different indicator meaning in a cross-cultural comparison. Ruth Benedict found that she had to use a number of Japanese terms in order to account for the English term 'duty'. This signified that duty in Japanese society is probably no abstract norm, but denotes specific obligations in a particular relationship.[9]

Conclusion 1. The concepts behind words are often delineated differently in different languages; the more abstract the concept, the greater the likelihood of differences.

To manipulate the surplus emotional meaning of words that are presumed to have a straight cognitive meaning is one of the prime skills in question phrasing, but becomes a major problem in cross-cultural work. When we adapted the Bogardus social distance scale to Germany, we found that the term 'neighbourhood' carried distance implications that were not equivalent to the cognitive equivalent 'Nachbarschaft'. Similar problems were encountered when using the term 'your community'. While in most societies the subject matter referred to by the term 'socialized medicine' can be denoted by a value-free term, in the US there seems to be no simple term left any more that does not tap party preferences.

Conclusion 2. Terms may have an emotional meaning only in one society and be quite technical in another.

Questions referring to 'fair play' are extremely hard to render in other languages; it is the combination of the concept 'fair' with 'play' that, e.g., makes it impossible to give an adequate rendering in German. The same is true for the words *esprit* or *Beamter.* Conversely, there is no general term for husband in Japanese. In a technical sense, problems arising from terms unique to a language or the absence of a term in another can usually be solved by word combinations; this is obviously not true for the emotional implications.

Conclusion 3. That certain words are unique to a language or altogether absent from it may signify that the phenomenon to which they refer is unique or absent. As Blood has pointed out, the discovery of unique terms and 'linguistic blanks' is an important substantive finding that should not be treated as calling for mere technical ingenuity.[10]

---

[8] P. R. Hofstätter, 'Die amerikanische und die deutsche Einsamkeit', *Verhandlungen des 13. Deutschen Soziologentages* (Köln, Westdeutscher Verlag, 1957), pp. 87-106.

[9] R. Benedict, *The Chrysanthemum and the Sword* (Boston, Houghton, 1946).

[10] R. O. Blood, Jr. and Y. J. Takeshita, 'Development of Cross-Cultural Equivalence of Measure of Marital Interaction for USA and Japan', *Transactions of the Fifth World Congress of Sociology, Washington, 1962,* Vol. 4 (Louvain, International Sociological Association, 1964), pp. 333-344.

In the International Citizenship Survey by Almond and Verba,[11] one of the questions was: 'Here are some important problems facing the people of this country. Which one do you feel is most important to you?' One of the choices offered was 'spiritual and moral betterment'. The combination of this with 'a country's problems' in the context of a political interview seems plain silly in Europe (and in a cognitive sense it may also be silly in the US, campaign promises of Mr. Goldwater notwithstanding). In the same survey respondents were asked to choose from the following statements in order to describe their feelings when going to the polls:

'1. I get a feeling of satisfaction
2. I do it only because it is my duty
3. I feel annoyed - it's a waste of time'.

The implication of this ordinal scale, that a hedonistic component is a very meaningful dimension of political participation - 'a feeling of satisfaction' - is in most European countries quite unwarranted; and in these countries joining the two statements 'I feel annoyed' and 'it's a waste of time' as equivalent expressions of displeasure is very problematical. To ask a Japanese respondent 'Where do you go on your vacation?' makes little sense since only the élite has the privilege of going on vacations.

Conclusion 4. Questions that presuppose the combined existence of factors may - however well worded - appear quite silly when transported into other societies. Questions have to be 'realistic' in wording.

In some of the developing societies (and sometimes among faculty members) direct questions are considered improper or even threatening. In a methodological study we could show that bifurcated vs. open-ended questions are differentially suited to the style of reacting to stimuli with uneducated vs. highly educated respondents. Closed questions, which in effect impose upon the respondent the conceptual orientation of the researcher, can be rather risky in cross-cultural comparisons. The very notion that a question is a stimulus to obtain cognitively used information from a 'subject' was found to be alien even in parts of Western European populations.

Conclusion 5. The format of questioning may carry cultural implications. Researchers in technologically and economically underdeveloped societies report that it is much more difficult to find words that are high in cognitive meaning than is true in Western industrialized societies.

Specifically, it is presumably very hard to find terms for social roles that do not carry undesirable status connotations. Thus, it is reported for Java that there are sixteen terms for the wife of a partner in conversation, depending on whether the questioner is younger or older than his partner, lower or higher in status, a stranger or an acquaintance.

[11] The volume by G. Almond and S. Verba, *op cit.,* is based on this project.

A somewhat extreme example of the status implications of presumably cognitive terms is again to be found in the Almond-Verba study.[12] In order to test partisanship, the respondents are asked how they would feel if their child married across party lines. Due to the high degree of association of parties with certain occupational and status groups in several European countries, the question has in effect the 'surplus meaning': 'How would you feel if your son or daughter married someone of lower, equal or higher status?' Consequently, adherents of higher status parties object much more strongly to their children's 'marrying down' than is true for the reverse. The degree of 'surplus meaning' which party labels carry is obviously a function of the closeness with which particular parties are associated with specific occupational groups. For an American it is probably difficult to perceive that party preference may be an indicator both for preference for a political content and for self-identification with occupational status groups.

> *Conclusion 6.* A language may be so constructed as to contain few terms referring to roles and group affiliation that would not carry strong status implications. In general, the cognitive loading of such terms appears to be lower, the lower the degree of cross-cutting group membership.

Stern and d'Epinay report from Switzerland that, while the written language is uniformly more or less 'High German', the spoken 'Schwitzerdeutsch' has strong local differences.[13] Accordingly, questionnaires are sometimes drafted in a sort of basic 'Schwitzerdeutsch' where the interviewer is then free to decide on the local idiom to be used (e.g., Züridеutsch or Berner Deutsch). This is not just an example of a widely used technique in multilingual nations (such as India); High German is understood by all and used by most respondents. However, High German is largely restricted to mass communications and public functions, and smacks of 'officialese'.

In a survey dealing with sexual behavior and attitudes we experienced major problems from the lack of a familiar language for this topic, a language that was neither vulgar nor medical. However, for the great majority a language dealing with sex in a matter-of-fact way is not available.

> *Conclusion* 7. The very choice of a language (or level of discourse) may have implications that are non-existent in another culture. Not in all societies is there a universal colloquial language, and even in countries where this is true, the language of colloquial discourse does not extend to all topics.

It is hoped that these generalizations may be useful as caveats. Hopefully, they are also suggestive of a better understanding of the question-and-answer process within a particular society, where due to group differentiation analogous problems arise - although undoubtedly much less dramatic in impact and con-

---

[12] *Ibid.,* p. 136 er *passim.*

[13] E. Stern and R. Lalive d'Epinay, 'Some Polling Experiences in Switzerland', *Public Opinion Quarterly,* 11 (4), (Winter, 1947-48), pp. 553-557.

sequently less visible. In actual cross-cultural survey work, experiences such as those mentioned above have naturally led to an emphasis on remedies for the problems arising from lack of comparability in wording.

II. In designing questionnaires for cross-cultural surveys, the emphasis has changed from insistence upon the correctness of literal translations to comparability of meaning. Usually, a purely qualitative procedure is used to achieve such comparability: the researcher in charge of a project informs his colleagues what he had in mind when phrasing a specific question, e.g. when he notices a non-literal translation of his original formulation. If time and money permit, the preferred procedure is a questionnaire conference after the first round of pretests. Of course, these qualitative checks are least effective if a question offers no problems in a literal translation. What is urgently needed as a routine procedure is a circulation of manuscripts in the analysis phase, affording cooperating researchers a chance to object to a questionable interpretation of previously unproblematical items. Unfortunately, the need for comparability of meaning is usually so far only perceived as a problem in the stage of design.

We think indeed that comparability of questions as a problem poses itself especially in the analysis stage, and needs to be tackled here, too. (If the later notion of 'equivalence of indicators' is accepted, concern with the problem at this stage does not amount to crying over spilt milk). However, so far there has been somewhat more interest in some devices which are to ensure comparability during the stage of design. Notable here is a suggestion by Osgood to check comparability of meaning by using the semantic differential.[14] Unfortunately, this is much too cumbersome a process to be used except for some key terms; even here it is, of course, only useful in checking the emotive connotations.

Somewhat greater has been the practical impact of suggestions to use nonverbal devices in international surveys. Such instruments as the scalometer are by now real fads in cross-national studies. Hadley Cantril's study, 'Hopes and Fears for Self and Country', is the latest example; here one chief instrument was graphic.[15] Some of these techniques are undoubtedly quite useful - but they are only a limited answer to our more serious difficulty: achieving equivalence of meaning. After all, visual stimuli are also a sort of language, and thus are not completely free from the problems mentioned before.

An example is Cantril's latest multi-nation study, where he uses as one stimulus a ladder.[16] This picture - or rather the device it refers to - was unknown to the Zulus. Subsequently, he employed with good success a picture of climbing successive terraces of a mountain, since ascending mountains was a widely shared experience.

[14] C. E. Osgood, 'On the Strategy of Cross-National Research into Subjective Culture', *Social Science Information,* 6 (1), (1967), pp. 5-37.

[15] H.Cantril and L. A. Free, 'Hopes and Fears for Self and Country', *American Behavioral Scientist,* 6(2), Suppl. (October, 1962), pp. 32.

[16] H. Cantril, 'A Study of Aspirations', *Scientific American,* 208(2), (1963), pp. 41-45.

It is true that many elementary errors are still committed when transposing questionnaires into other cultures. Experienced researchers, however, can usually handle intuitively the problems of developing an instrument that will be a good 'free translation', and will be reasonably easy to administer.

The latter may be a very misleading criterion. One of the major problems in survey work - magnified in cross-cultural comparisons - is the fact that questions are answered even if they mean quite different things to different respondents. This in itself does not invalidate surveys by any means. Obviously, it is not the faultiness of an instrument as such that causes worry, but the degree of faultiness as measured against the uses made of the responses. This simple consideration is widely overlooked: the frequent European critiques of comparative research via surveys advance unrealistic standards of the needed perfection of empirical instruments; the practitioners of survey research tend to rest assured with the experience that most of the predicted difficulties just did not materialize and that one almost always gets 'results'.

I do think that the meaning of the many difficulties we experience in question wording for cross-culturally comparative surveys is inadequately understood; I also think that there has been considerable progress in handling the practical problems. Beyond this advance in practices, I think there has been a major progress especially in research that is administered by commercial institutes: a willingness to depart from literal translations and to aim at equivalence in meaning. This I think is a major step - a step, however, that should be followed by the understanding that equivalence of meaning is only a special case of the general property of 'functional equivalence of indicator meaning'. This, however, presupposes a different outlook on the question-and-answer process in survey interviewing.

4. Problem area II: Equivalence of indicators

There is still, however, a good deal of uneasiness about putting equivalence of meaning before identical wording, especially in academic research. After all, haven't we learned what seemingly minor variations can do to the responses? Thus, the prevailing orientation is one of rather naive realism: if a question is not asked in the same way, we cannot compare results; if it has heen asked the same way, we can. This may be defensible folklore of research, but it certainly is not good methodology.

By now social scientists should have become accustomed to looking at questions as indicators - indicators that have a probabilistic relationship to a property one intends to measure. Indicators are interchangeable in terms of their functions, which are to express the property we want to ascertain. Hence, the criterion for maintaining that questions are comparable is not whether they are identical or equivalent in their commonsense meaning, but whether they are

functionally equivalent for the purposes of analysis. I see major progress in cross-cultural survey work in that this notion is now beginning to be more generally accepted. Hopefully this will translate itself into research *within* one country.

I. Functional equivalence may be an exciting concept, but it is of course somewhat difficult to apply in actual research.[17] As a matter of fact, one would either need to check this property empirically, or have to use an empirically supported substantive theory - preferably both, if they are available.
To give some examples:

1. As I mentioned above, several years ago in Germany we tried to adapt the Bogardus social distance scale. We found that, in Germany, the concept of neighborhood carries a different (less stringent) implication in terms of social intimacy. Thus, we had to try to find another item indicating the same position on the Seven-Step Distance Scale. Of course, the criterion we finally settled for - 'Have as a greeting acquaintance' - did not have exactly the same position on the latent continuum (in terms of distance from the end points), but it did have the same ordinal position, and the total scales were equivalent. A somewhat different problem was encountered some twenty-five years ago by Stuart Dodd in the Near East, when he found that the Bogardus social distance items did not cover the total range of possible enmities.[18] As you may remember, Dodd finally came up with an item T wish somebody will kill all these people' as an adequate measure for the total range of actually existing distances.

2. As just mentioned, Cantril had to use different versions of the scalometer in cultures as different as the Zulus and the United States.[19] He encountered another and more difficult problem with his verbal stimulus: to imagine one's best conditions and worst conditions, five years hence. The ability to project hope and fear into the future for definite time intervals was - not surprisingly - quite culture-specific.

In the US, five years indicated a planning span familiar to respondents in their everyday life; in some primitive societies, a season, or at most a year, was the timespace in which respondents normally operated. Cantril finally resorted to a more variable time-reference to achieve equivalence of function: a referent for fears and hopes.

In a cross-cultural survey comparing role differentiation between husband and wife in the United States and Japan, a battery of household tasks had to be translated from English to Japanese.[20] Of the eight items used in America, only

[17] 'Functional equivalence' as a concept was used in sociological theory to overcome what was felt by some to be a misplaced concreteness in some structural-functionalist work. For a discussion of the notion see R. K. Merton, *Social Theory and Social Structure,* 2nd ed. (Glencoe, 111., The Free Press, 1957), Chapter 1.

[18] S. C. Dodd, 'A Social Distance Test in the Near East', *American Journal of Sociology,* 41 (1935), pp. 194-205.

[19] H. Cantril and L. A. Free, *op. cit.*

[20] D. S. Yamahura and M. M. Zald, 'A Note on the Usefulness and Validity of the Herbst Family Questionnaires', *Human Relations,* 9 (1956), pp. 217-221.

one could be used literally, and one further after considerable rewording. For the six other items, the activities did not occur in Japanese households or meant different things there, and functional equivalents had to be found.

3. In a cross-cultural survey of the goals and methods of child-training it was found that rates of physical punishment varied among cultures and within societies among classes.[21] It was not possible to measure the amount of disciplinary activity using this question as the main indicator. Even a middle-class sociologist living in a homogeneous middle-class suburb and teaching middle-class students would be aware of within-culture variations enough to know that beating one's children and/or wife has considerably less emotional significance in the lower classes. Therefore, a battery of indicators had to be designed which would really measure the extend of disciplinary activities as a latent concept, rather than merely reporting the prevalence of some overt acts.

II. Logically the same problem of 'equivalence of meaning' has been discussed for many years under a different label - but the discussion has taken the same course I have just outlined for the debate on question meaning. I am referring here to the attempts to standardize internationally so-called 'background characteristics', i.e. questions that are routinely employed as independent variables in breakdowns (such as age, income, education, community size, etc.).

There has certainly been no lack of effort in several meetings of WAPOR, ISSC, and various international chains of commercial institutes. Recommendations were passed by international committees to standardize background variables. I have to report now that in this sense there has been little progress in international survey work. It has not even been possible to agree on common age-groupings.

Why these difficulties in agreeing upon common background characteristics - and why the lack of success even with such seemingly trivial conventions as common age-groupings? This time lack of cooperation is not due to the frequent idiosyncracies of any group of scientists. There is a better reason, and awareness of it can be viewed as progress. In discussing internationally the 'best' age breaks it became more and more obvious why age is really used, and when used why it is effective. Certainly the attribute, *age,* is usually not interesting as a physical characteristic; rather age-groupings are employed to denote powerful and general role differences - differences which can pattern behavior across many different situations. Often age has been used when 'stage-in-the-life-cycle' as a sequence of role configurations was meant. However, if physical age is used as an indicator for different social roles, then the same physical age will denote different roles and different stages in the life cycle in different cultures. Hence no agreement on a standardization.

---

[21] E. Devereux, *Family Authority and Child Behavior in West Germany and the US* (Ithaca, N. Y., Center for Child Development, Cornell University, 1963). Mimeo.

Age is just one clear example. 'Old* as a social definition is obviously related to a different physical age in various societies. Thus, 40 years of age will define a different position for a woman in Sicily than for a woman in southern California or Florida. Actually, the same relationship between physical and social age does not even hold within the same country. A 45-year-old worker is nearing old age; a physician of the same age has just about reached full maturity.

Another example is reported by Stern and d'Epinay.[22] They show that in Switzerland a town of 5,000 people is considerably more of a center of commerce and cultural activity - is in other words a different social unit - than is true for a town of the same size in France. If community size is used as an index of urbanization and relative access to central facilities, a town of 50,000 in Sicily may be equivalent to a town of 2,000 people in the midwestern United States.

In striving for standardization of background characteristics in the sense of identity of labels, researchers have been worrying about the wrong problem. Again there was the frequent danger in cross-cultural research of equating formal identity of procedures with equivalence of an indicator's meanings.

Actually some disagreement about the exact cutting points in a quantitative variable is not so important in the first place as long as the range of variability is comparable (except of course in dealing with very different societies). Much more difficult is the standardization of qualitative variables, or quantitative variables qualitively expressed - and if such standardization exists, the interpretation of the categories is difficult too.

A peasant in Europe is still something different from an American farmer; if one compares responses for both groups, much of what is done actually shows that similar labels refer to different groups, rather than demonstrating cross-cultural differences between the responses of otherwise comparable groups.

While it is a substantive finding of considerable interest to show that occupational groups do not exactly match their counterparts in other countries, this fact does not help much when employing occupation as a background variable. However, if a specific hypothesis exists as to *why* occupation should affect certain responses, we again can decide on functional equivalents. If one wants to test whether persons in occupations with easy access, little hope for upward mobility, and low job security show a lower preference for political movements that stand for social reform via evolution, one might take an unskilled worker in the US and compare him with a plantation worker in an underdeveloped country.

Consider even the seemingly comparable definitions of educational categories in the US and Europe. The last session of the ISA Subcommittee on Stratification and Social Mobility amply demonstrated the difficulties arising from the different social consequences attached to the same educational label.[23]

[22] E. Stern and R. D'Epinay, *op. cit.*

[23] K. Svalastoga, 'Gedanken zu internationalen Vergleichen sozialer Mobilität', in D. V.

As Mark Abrams pointed out during the 1962 World Congress of Sociology, college education means something vastly different in the US from what it means in England.[24] I am not referring here to real or imagined differences in the quality of education, but to its differing consequences for one's social position. A college education in England still denotes that one is a member of a select cultural minority which will interact frequently with other minorities across professional boundaries. In the US, the rarity of a college education is by now about as great as the rarity of a high school diploma was a generation ago. Thus in comparing the United States and England one might compare the college educated in England with the graduates of major private colleges in the US. A baccalauréat in France and a Matura in Germanspeaking countries have no exact counterpart in American education. Again, this is not meant as a reflection on quality, but refers to differences in the meaning of education. European educational systems just happen to be elitist, while American education is not.

Once these difficulties are properly conceptualized (in part I have only made the implications of practices explicit), many of the problems of insuring comparability become much easier to manage.

5. Problem area III: The respondent as a unit in design and analysis

Largely as a result of difficulties experienced in designing samples for underdeveloped countries, another problem area begins to emerge: assumptions about the respondent as a unit in survey research. These are implicit in the procedures that we customarily use. In making such assumptions explicit, technical problems in field work and in other stages of research once again turn into problems of methodology. Specifically we find ourselves discussing a long by-passed methodological problem, in part already posed by Durkheim: the character of data in sociological research as a basis for deriving general statements.[25]

I. In accordance with the historical context in which surveys developed, an important implication of survey research, or at least of prevailing survey design, is that all members of a population matter, are largely interchangeable as

---

Glass and R. König, eds., *Soziale Schichtung und soziale Mobilität* (Köln, Westdeutscher Verlag, 1961), pp. 284-302.

[24] M. Abrams, *The Political Division of the British Middle Class.* Paper presented at the Fifth World Congress of Sociology (Washington, D.C., 1962).

[25] Survey researchers appear prone to forget that the observables are not the facts of a social science discipline with theoretical intent. See Dürkheim's distinction between *faits sociaux* and *faits sociologiques* in E. Durkheim, *Les Règles de la méthode sociologique* (Paris, P.U.F., 1895), e.g., in the preface to the second edition.

units (i.e. as carriers of attributes), and exhibit the properties under investigation. That all units matter, and that individuals are the relevant units for the purpose of a study - this is by no means self-evident; in many societies and for many purposes this assumption is quite wrong.

When in Southern Italy, e.g., a rural woman is approached and asked to give her opinion, she not only often feels unable to do so, but also exhibits signs of feeling threatened when pressed to perform that peculiar operation that we call 'giving an opinion'. A frequent reaction to the question 'how do you feel about...?' is 'we here feel...'. In other words, the individual tends to report in terms of the prevailing opinions within his group; he is reluctant to dissociate himself mentally from group membership or to make a distinction between individual opinion and group consensus. Similar experiences have been reported from India. A salesman type of interviewer may still be able to extract some kind of answer as demanded by the questionnaire. Yet how are we to make use of such a response?

T. H. Marschall in 1950 suggested a model for what he called 'extension of citizenship' to ever more groups of a society, until finally all physical members were to be regarded as citizens in a substantive sense.[26] The degree to which this process has actually taken place in a society can be considered a (latent) limiting condition for the prevailing type of survey research.

Actually, it is a surprising phenomenon that, especially in the US, most people are willing to voice an opinion on nearly any imaginable topic, from resurrection to the effects of subsidizing medical care for the aged by social security funds, from weapons systems to the effects of a balanced budget. And equally surprising is that these responses are often treated as somehow referring to fact or deciding controversial contentions. 'Is there in your opinion life on Mars?' - such a question does not strike most respondents as strange, and as we know from some wellknown earlier studies, a majority is quite willing to have definite views on nonexistent consumer items or on an imaginary 'metallic metals act'.[27]

In societies with a legitimate stratification system, individuals define differently the range on which they feel competent to voice opinions. An extreme case appears to be France, for which Mark Abrams reports that in the middle classes only so-called serious questions, defined as involving fantasy on a rather cosmic scale are considered acceptable.[28] These same respondents are

[26] T. H. Marshall, *Citizenship and Social Class and Other Essays* (Cambridge, Cambridge University Press, 1950).

[27] Numerous examples of this kind are found in S. L. Payne, *The Art of Asking Questions* (Princeton, N. J., Princeton University Press, 1951).

[28] M. Abrams, *Problems of the Purposes and Procedures of Cross-National Comparisons.* Paper for a Round Table Conference on 'The uses of sample survey data in comparative cross-national research'. Paris, International Social Science Council, 1962, mimeo. Also D. Lemer, 'Interviewing Frenchmen', *American Journal of Sociology,* 62 (1956), pp. 187-194.

less willing to voice opinions on, e.g., government decisions that they define as subjects for the experts. The latent communality explaining the differential willingness to voice an opinion and to consider questions as reasonable seems to be the respondent's self-definition of competence.

Obviously we need more empirically substantiated propositions about relevance and reactions of our units in survey research. Such propositions about respondents' self-definition and behavior - and more generally a clarification of the very nature of data in survey research - are easier to generate in cross-cultural surveys, where differences in the contexts from which we draw our units of study (i.e. respondents) are greater and thus are more easily conceptualized. Though these differences are responsible for some of our greatest problems, they can at the same time contribute to the great advance which cross-cultural comparison can mean for survey research.

II. Whether the relevance we attribute to each unit - explicitly or implicitly in our procedures - holds cross-culturally is one of the chief problems underlying sampling for comparative research. This is a new problem-formulation, and one that has emerged from conducting surveys in the developing countries. These experiences are also relevant for research in industrialized and urbanized countries, although this fact as yet is rarely seen.

Sampling was one of the earliest problem areas to be recognized as such in comparative survey research. As is true for the other methodological aspects of cross-cultural surveys the problem definition has changed over the years, starting with concern for identical procedures. Again we meet the common-sense notion that comparability is assured only if the same procedures are applied in all countries. Yet on purely theoretical grounds this should strike us as a very peculiar notion. 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.

To insist on identical sampling procedures as a *condition* of comparability shows little confidence in samples as a tool of inference, and constitutes a misplaced trust in some of its concrete features. This trust is misplaced because concrete sample designs for human populations - at least by implication - represent adjustments of the model of the simple probability sample to specific topics and structural properties of the universe that is being studied. Rather than insisting on identical procedures, one should postulate that, if significant aspects of the universes to be studied differ, the designs should differ accordingly.

One important qualification is necessary. In terms of sampling theory, insistence on identical procedures is unnecessary and sometimes harmful. However, we have to consider losses in probability sampling - and we known that

the chances of losing persons within various population groups vary with different procedures. For various Western countries the usual rate of loss is between 10 percent and 20 percent of the original sample; in some subgroups, such as metropolitan populations, losses are often around 25 percent. These figures hold cross-culturally. Since such sizable groups of the sample are not covered, one needs to be reasonably sure that there are comparable types of losses; comparable procedures in the last stages of a multi-stage design may be one way of making comparability of losses more likely. Even then it makes absolutely no sense to insist routinely on uniform procedures.

The prevailing method of nation-wide probability sampling in the U.S. happens to be a type of area sampling, which in the final stages calls for rather cumbersome estimates of population composition (for stratification purposes) and the tedious procedure of 'prelisting.' If we compare conditions in the U.S. with those in some European countries, we realize that this sample design is well suited to handle some problems specific to the U.S.: unavailability of population lists (there is no compulsory registration), wide spacing of the census (at 10-year intervals), relative homogeneity of neighborhoods and specifically of the houses within these neighborhoods.

There is little reason to insist on this type of sample for Europe. There are laws forcing everyone to register his residence; the local administrations are infinitely more standardised; heterogeneity within apartment houses is high; and, in particular, there are often very efficient local statistical offices. In Germany and Holland, for example, there are excellent registers for all units in the population (in several German cities one IBM card per person contains all the data necessary for the usual stratification procedures). In France, registers have mainly been found deficient merely in cities between 10,000 and 100,000. In several European countries we have even had the choice between several sampling lists, enabling us to draw according to our needs a sample either of households or of individuals.

Some years ago I used an area sample in Germany partly to check which population groups would be under-represented in random samples from files. The area sample yielded only 2 percent of persons we would have missed in a conventional file sample, and it did so at a much higher cost and with higher sampling losses.

The area sample is just *one* practical design for coping with certain peculiarities of one's universe. There is certainly no reason to behave as if it were the 'natural' design for population samples - as is done sometimes in American texts.

Cross-cultural surveys have stimulated the fantasies of a few sample specialists. Some of their innovations in cross-cultural work might be used more widely in domestic research as well. In particular the work of the Indian Statistical Bureau in this immensely differentiated sub-continent has resulted in new designs (e.g., 'interpenetrating samples').[29] Deming's population sample

[29] P. C. Mahalanobis, 'Recent Experiments in Statistical Sampling in the Indian Statistical Institute', *Journal of the Royal Statistical Society,* 109 (1946).

for Greece and subsequent proposals based on similar experiences have been published in part and are reasonably well-known among sampling specialists.[30] One of the networks of commercial institutes, well-known for its insistence on random sampling, is presently using a revised version of Deming's master sample for survey work in South America. This design again presupposes only a minimum of available information. Repeatedly, another design has been proposed which makes very few assumptions about existing knowledge: the so-called random-path, or random-walk design. So far this design appears not to have found a continuous usage - probably because of the high demands it makes the interviewer's honesty. Essentially, this random walk design consists in sampling points from a grid. These points merely define the beginning of a cluster of households. A detailed prescription governs the way in which one proceeds along a chain of households and arrives at the one to be sampled. Since the distance between the starting point and the selected households is defined in number of households in between, different population density is automatically neutralized. This sample largely dispenses with the necessity of first stratifying by size and then prelisting.

In research practice (especially within international networks of commercial agencies) there have been repeated attempts to develop a sample that can be applied with only minor variations in all countries. Thus, one of the networks of commercial institutes works at present with a modified multistage area sample. There have also been recurring attempts to standardize the controls in quota samples. In these attempts there is a more or less intuitive awareness that uniformity of procedures is not the best strategy; but there is an insistence on using the same master plan. This insistence on some uniformity in principles of design with variability in procedures is advisable; it is essential in cross-cultural surveys tracing the importance of particular variables.

Ceteris paribus, the stronger the descriptive aspects of a cross-cultural survey, the less important is uniformity in the principles of sample design; the more analytical purposes are stressed, the more important is the identity of principles of design; identical procedures, however, are usually a misplaced concreteness.

The Almond and Verba project offers an excellent opportunity to exemplify what we consider a permissible - and often necessary - range in variability for cross-cultural surveys; it also makes obvious the conditions under which the need for uniformity of sampling principles occurs.[31] This study in five very different societies is characterized by the use of the same basic type of sample, a wide variety of designs, uniformity and diversity of procedures, and a very considerable divergence in the degree to which sample designs could actually be implemented.

---

[30] W. E. Deming, *Some Theory of Sampling* (New York, J. Wiley, 1955).
[31] G. A. Almond and S. Verba, *op. cit.,* Appendix.

The Almond-Verba study was actually carried out by commercial survey research institutes of the five countries studied. All these institutes adhered closely to their normal procedures - procedures that in the countries concerned were usually considered to represent good technical standards. These institutes were part of a major commercial chain, and in years of cooperative research they had agreed on considerable similarity of sampling designs. Consequently, there were both a basic similarity due to past cooperation and a divergence according to the traditions of particular institutes.

All institutes used a multistage, stratified probability sample that, up to the ultimate sampling stages, conformed essentially to the principles of area sampling. The number of stages in the area sampling phase differed, and so did the nature and variability of units.

To characterize, first, the variability of selection up to the smallest contiguous area: in Germany the psu's were the 30,000 communities, from which 100 communities were selected in one step. In Italy, the first stage units were the 92 provinces from which 13 were included in the sample as the universe for the second stage; here 49 communities were selected. In the United Kingdom three stages were used to arrive at the smallest contiguous area: from 630 parliamentary constituencies an (unknown) number was selected; from each of the remaining constituencies two wards were chosen, and from each of those one polling district. In the United States, the sample followed closely the standard procedures in the US versions of area sampling: first-stage units were the 3,000 metropolitan areas and non-metropolitan counties; second-stage units in urban areas were census tracts, while in rural areas 'localities' were used; in the third stage the sampling units were either blocks or segments (i.e. equivalent non-urban small geographical units); altogether, 491 small geographical units were thus selected. The most complicated design was employed in Mexico; although it is not quite clear what the actual procedures were, it seems that in some way 27 cities over 10,000 inhabitants constituted the sampling frame, from which in one step city blocks were chosen in a systematic fashion. There is no doubt that in all cases the design either conformed to the principles of probability sampling or came close to them, although the standard error of the procedures (or the cluster effects involved in multistage sampling) should differ dramatically. This cluster effect up to the smallest contiguous geographical area seems smallest in Germany, largest in Mexico and Italy - but the data reported do not permit even a rough calculation of size.

Quite different were also the procedures followed in the actual determination of respondents within these smallest geographical areas. In Germany, Italy, and the United Kingdom the register of electors was used, but in a different way in all three countries. In Germany the name of an elector was merely used to locate a household, within which the actual respondent was identified by using a table of random numbers. In the United Kingdom respondents over 21 years of age were immediately identified and respondents below voting age were

selected in a two-stage process akin to the one followed in Germany (instead of random numbers, the birthday procedure was used). In Italy the respondents above legal voting age were identified in a onestage process as in the UK; those below voting age were taken from households in which an older respondent had already been selected. Both in Mexico and in the United States, the selection of the psu's was preceeded by a 'prelisting' within blocks or segments. In the USA, dwelling units were prelisted within all blocks and selected first; a prelisting of all adults within a dwelling unit was then prepared and a table of random numbers applied to identify the actual respondent within the du. In Mexico selection within blocks proceeded immediately to the household as the next sampling unit, and within households the individuals were identified via random numbers. Thus, procedures in this second phase of sampling differed strongly by the sizes of the initial universes (communities or blocks), the number of steps taken to identify the respondent within the blocks, and the method for identifying respondents - two procedures sometimes being used simultaneously. The cluster effects were probably smallest in the UK and in Italy, probably largest in Mexico and the United States. Both in Mexico and in Italy, the chances of some of the respondents were more strongly affected by the size and character of households than was true in the other countries.

In this very summary description we have purposely left out a description of the - sometimes quite elaborate - stratification procedures on several stages of the sampling process. However, they undoubtedly tended to counteract some of the cluster effects and risky weighting operations. It still remains doubtful whether in view of the rather small sample size the reduction was generally sufficient to allow the use of the - anyhow a bit questionable - formula for the computation of standard errors that Almond and Verba list.[32] For highly skewed distributions we would guess that this mathematical expression may often lead to an underestimation of sampling errors, and this becomes quite relevant for some of the differences between countries when in several of them the $p < 25\%$. Much more important are differences in the degree to which the original design could be implemented. The rate of losses of sampled individuals is surprisingly high in most countries: in the USA the completion rate for the sample is a quite normal 83%; in both Germany and Italy completion rates drop to 74% which is (at least in Germany) rather high; completion rates are a disastrously low 60% both in Mexico and - most surprisingly - in the United Kingdom. It is well-known that losses affect the distribution of many study characteristics even if their effect on the marginal distributions for standard demographic checks are not very marked; but in the report of the studies we have no account of sampling losses, nor is any allowance made in interpretations for this partial coverage. If, e.g., the chance of inclusion in the sample is associated with participation in public affairs - which appears to be a reasonable assumption - then the highly different completion rates may account for some of the observed

[32] *Ibid., p. 525 et passim.*

differences. This effect is especially worrisome for England, which (especially in respect to its differences from Germany) is of crucial importance to the authors' main line of argument.

A number of aspects of the design are problematic: thus, nowhere is the effect of the different weighting procedures, such as the undersampling of rural areas in German and of metropolitan areas in Mexico, examined. Also, we find it questionable to rely in the interpretation sometimes on the size of the differences between percentages and sometimes on rank ordering; some of the arguments would not have looked so good if either type of usage had been adopted more consistently. There is one major and obviously disturbing aspect of the study design: the definition of the sampled population in Mexico. In all countries except Mexico the sample represents the population between 18 and 80 years of age, and only in this country does the sample refer to the population 21 years of age and older. Much more important is the restriction to an urban population in Mexico. It cannot be argued from the position that the authors chose to adopt that the rural population did not matter; if political institutions were really determined by basic population characteristics (as is maintained in this study) one can't very well eliminate part of the 'basis'.

A closer scrutiny of the Almond and Verba five-nation study offers an opportunity to demonstrate which aspects of sampling are to be considered in ensuring comparability in cross-cultural comparisons and which aspects do not matter, contrary to much of the folklore of survey research. In this study, the design and the actual operations in the sampling up to the smallest contiguous area differed among all of the countries - which was probably beneficial in as much as the various field work organizations adjusted well to the conditions in each country. Sampling errors were affected by these differences since some of the procedures implied strong cluster effects, especially with the (in view of these differences!) limited sample size.

The design and procedures for the selection of respondents within these areas were also different in all of the countries, again largely as an adjustment to the objective conditions. These differences during the second phase of the sampling process affect the composition of the samples more seriously than is true for the differences during the first place; e.g., the variations of within-household selections determine in part the distortions in age distributions. While these effects are by and large probably unimportant for the intended type of statements, the differences have serious consequences in England and Mexico, given the high rates of noncompletion. Here they are so high that it is questionable whether the samples for these two nations should be used at all in straight comparisons of marginal distributions between countries; the correlations between variables are obviously much less affected.

To conceptualize the above-mentioned difference in the consequences of variations in sampling design, let us introduce a distinction between the *representativeness* and the *scope* of a sample (a distinction which we believe

was first proposed by Zetterberg).[33] Representativeness of a sample is the quality of a partial enumeration to represent the relative size of properties in the universe sampled; scope of a sample is the quality of a partial enumeration to permit the intended inferences. Representativeness is the most important quality of a sample in descriptive investigations; scope is more important in investigations with analytical intent.

The Almond-Verba five-nation study strove for representativeness as a prerequisite of comparability - a property that was indeed achieved for most of the variables in most of the countries compared.[34] However, the intent of the study was also largely analytical - and in important instances the scope of the sample was not sufficient for this.

The most frequent empirical datum used in this study was a straight comparison of marginal distributions by nation. However, the second most frequent independent variable was education, which was used in 49 of the 140 tables and figures of this volume, i.e. in 35 percent of all cases. Such usually important demographic variables as sex, age, community size, occupation, and religion were used much less frequently, and social class not at all. On substantive grounds I think this was a mistake - though an emphasis on education follows logically from the authors' theoretical orientation.

Provided sample sizes and number of countries to be covered were a constant, Almond and Verba were confronted with a dilemma which they apparently did not anticipate: if demonstrating between-nation differences was a chief purpose, representativeness of the sample was a main requirement; if demonstrating the effect of education on citizenship participation was a main goal, merely a normal representative sample would not do. Probably the authors did not fully realize the differences between the educational systems of the US and European countries, and consequently did not provide in their sample design for the small numbers of persons with a higher education in Western Europe. A disproportionate stratification would have extended the scope of the sample and might even have permitted a better gradation than the constantly employed split: primary education versus secondary education and more. Had they anticipated the different distributions of educational categories in the various countries, it might have been possible to show that 2 percent or so of adults in Western European countries with a university degree differ very, very significantly from persons with secondary education; of the samples actually used, even some of the presumed effects of higher education in general are of questionable significance.

---

[33] H. Zetterberg, 'Notes on Theory Construction and Verification in Sociology', *Transactions of the Fifth World Congress of Sociology,* Lou vain, International Sociological Association, 1964, Vol. 4, pp. 155-184; *On Theory and Verification in Sociology,* 3rd enl. ed. (Totowa, N. J., Bedminister Press, 1965).

[34] G. A. Almond and S. Verba, *op cit.* Methodological appendix.

In extending sample surveys to countries with differing degrees of technical and economic development, and with markedly different social and political organizations, the main problem of sampling is still more important than the often overlooked difference between representativeness and scope. In cross-cultural surveys of countries with such marked differences, the question of the comparability of the samples themselves is preceded by the question: what are comparable populations? In Western countries we usually consider the total adult population the relevant universe to be represented - implying that for the purposes of study there is some reason to know the reactions of all these units. This may be open to question; it is often an outright unreasonable assumption for developing countries.

In research practice the coverage of total populations in developing countries is quite bad anyway. But instead of having the sample represent merely the more easily accessible parts of the population (e.g., in South America, the *permanent* city dweller), we should consider what parts of the population one really needs to know about. This often requires an explicit theory about the relation of the units to the dependent variables. Thus, one of the chief problems in comparative sampling seems now to relate sample design more explicitly to substantive theory.

Recently some colleagues from Venezuela argued that in one sense a simple probability sample implies a 'know-nothing' approach to the population under investigation, a studied posture of ignorance when in fact one knows better.[35] In such a random sample one relies on breakdowns to yield structural information, and controls for the effect of such structural factors on the dependent variables in the analysis stage. There is no good reason why one should not build this knowledge into the design at the beginning. To give an example: if population groups are not integrated into the money economy, they are just not relevant for all studies on this topic. If certain migrants are known to be a passive quantity in accepting whatever conditions exist in the polity, one might often concentrate on those groups which do exert an influence on these conditions.

Comparability of samples with regard to 'scope' can be a quite different matter from comparability with regard to the property of 'representativeness'. Comparability in scope may not only permit but even require different designs (e.g. in the Almond-Verba study a disproportionate stratification would probably have been unnecessary in the USA). And comparability of the scope of the sample may even require different delineations of the parent populations of samples in order to ensure comparability of units in analysis.

---

[35] Unpublished communication by F. Bonilla of M.I.T. Bonilla reported on a joint project between CENDES (Caracas) and political scientists from the Massachusetts Institute of Technology.

6. Problem area IV: The usage of 'Culture' in cross-cultural surveys

The proper units in cross-cultural survey research may be sometimes open to question. No less open to question may be the meaning of the very criterion that makes a research cross-cultural: the use of different cultural (or national) contexts. When conducting surveys in more than one country it is usual that differences in the distribution of properties are obtained; but how should these differences be interpreted - or, more precisely, what is the logic of accounting for these differences? There appears to be little progress as yet in becoming aware of the actual uses made in conducting surveys in different settings.

Cross-cultural comparisons were characterized in section 2 as a special case of the comparative method. Depending on the treatment of the units that are being compared (it is the treatment of units that is suggested as a distinguishing feature, not some imaginary 'essence' of the units) we may distinguish two versions of the comparative method. These versions differ in the goal of comparisons (i.e. in what is shown to be different from what) and in the logic of accounting for the differences that are found.

a. The goal may be to establish the different identities of the cultures (or other units) that are being compared. The individual characteristics that are being observed are treated as indicators of the configuration ***(Gestalt)*** of properties that establish the identity of different units, and/or are causing the differences.

b. In working with observational data (in the wide sense of the term), registration of the same data under different conditions is to serve as an analogue to the experiment. Experiments may be defined as the creation of different conditions that should lead to differences in the phenomena that are to be accounted for; in the comparative method, the differences in conditions are not created but selected and viewed as causative agents. Choosing different cultures (or nation-states) is an attempt to magnify the differences in conditions. This is a use of comparisons that J. S. Mill recommended as the procedure characteristic of generalizing social sciences. A main problem in this version of comparative method results from the lack of control over the conditions of differential observations: it is difficult to establish whether it is merely the set of factors deemed relevant by the researcher that really accounts for the observed differences in the dependent variables.

To give an obvious example of this difficulty, again from the Almond and Verba project: the authors observe that the frequency of political conversations in Germany is lower than that observed either in the USA or in the United Kingdom. They attribute this difference to differing 'political cultures'. An equally likely explanation would be (apart from the sampling problem in Britain) that the surveys were carried out at a time close to national election both in Britain and in the USA, but halfway between elections in Germany. If I

compare the 1959 data from Britain with the results for the same question in a nation-wide survey of ours just prior to the 1961 national election in Germany, the differences disappear.

We took four surveys spaced over a period of three months prior to national election and up to nine months afterwards, and we found drastic fluctuations in the frequency of political conversation at the four different points in time (fluctuations by 100 percent from the lowest value).[36] I conclude that the property of the collective that may actually have 'caused' the observed differences may have been the constancy or fluctuation of the degree of 'politicization' of societies at various, recurring points in time due to national elections or other important political events; and, as is frequently true in the Almond and Verba project, after a change in perspective there are relatively minor differences between the British and German polities.

Even if this explanation of the observed differences is not considered as plausible as that offered by Almond and Verba, it cannot be ruled out - demonstrating the difficulty in establishing an unambiguous relation between such complex units as 'political cultures' and the answers to specific survey questions.

A distinction along a second dimension may help to clarify further the accounting schemes (in Lazarsfeld's sense) used in cross-cultural comparisons, and specifically those employed in comparative survey research. In any scientific project, the emphasis may be 1) on establishing similarities or 2) on establishing differences (usually one or the other), and in any research there may be an attempt to accentuate either the similarities or the differences in the dependent variables. Obviously, both can be observed in any investigation, and unfortunately it is largely a matter of judgment whether the observed differences are explained as a consequence of measurement error or seen as evidence of differences in the realities studied; this is nothing peculiar to the social sciences. Comparisons across cultures usually provide a more acid test whether one or the other of the interpretations is more plausible - regardless of the original intent.

The intention of demonstrating either similarities or differences takes a different form whether a) the units (or cultures) are treated as configurations or b) they are treated as conditions different with respect to causative agents:

a. If an emphasis on similarities is combined with the use of different cultures as units, such cross-cultural studies may take the form of testing the applicability of conceptual schemes (e.g., Marion Levy)[37] or - in research with a substantive orientation - of establishing cultural universals (e.g., Murdock's

---

[36] E. K. Scheuch, 'Die Sichtbarkeit politischer Einstellungen im alltäglichen Verhalten', in E. K. Scheuch and R. Wildenmann, eds., *Zur Soziologie der Wahl* (Köln, Westdeutscher Verlag, 1965).

[37] M. Levy, *The Structure of Society* (Princeton, N. J., Princeton University Press, 1952).

work on the incest taboo).[38] Research in cultural anthropology is usually of this character. A typical problem in this type of research is to deal with both between- and within-culture variation, the latter being a property that cultural anthropologists tend to treat a bit lightly anyway.

p. Cultures may still be treated as units while the intent is to show the differences between those units. This was characteristic of the traditional attempts to construct typologies of societies (e.g., the 'older' historical school of German social science), and this orientation is now reasserting itself again in some empirical studies. Examples of this form of cross-cultural comparison are Upset's more recent work[39] and, to a degree, the Almond and Verba study also. The crucial problem in this research is to relate the differences in observables (e.g., survey questions) unambiguously to the defining criteria *(konstituierende Prinzipien).* These criteria are treated as parts of a unique configuration of elements which, when applied individually (i.e. each one in succession), can be used to characterize an aspect of all the societies considered. Thus, all societies relate the individuals either as active or as passive elements to processes on the level of the polity; it is in combination with other variables that the unique type 'parochial-participant culture', is established. Usually the configuration of system properties (culture) is considered causative of the observed phenomena - a use of 'culture' that is reminiscent of the 'ideal type' in Max Weber's writings.

Sociology and political science - when choosing the format of the natural sciences - intend to be generalizing sciences; and yet the time-space coordinates within which empirically found 'laws' are valid remain (actually or potentially) a matter of perennial dispute. In lieu of the experiment, observation of the same factors in different contexts is the obvious strategy, either to confirm empirically the generality of the statements or to specify the time-space coordinates of their applicability. Characteristic of this approach is that the units of analysis (i.e. the units to which statements of in variance refer) are of a lower order than that of global systems such as polities or societies; predominantly, the statements to be tested refer to individual behavior. In the attempt to specify or to establish generalities, 'cultures' (or nation-states) are serving as opportunities for observation under different conditions, but the problems of relating the observations to their contexts are somewhat different.

y. Since the mid-fifties attempts have been increasing to verify the generality of ahistorically phrased propositions. If the same proposition about relations between factors can be shown to hold in all cultures or countries, this is equi-

[38] G. P. Murdock, *Social Structure* (New York, Macmillan, 1949); B. B. Whiting, ed., *Six Cultures: Studies of Child Rearing* (New York, J. Wiley, 1963); see also J. W. M. Whiting, *et al., Field Guide for a Study of Socialization in Five Societies* (Cambridge, Mass., Laboratory of Human Development, 1954).

[39] S. M. Lipset and R. Bendix, *Social Mobility in Industrial Society* (Berkeley, University of California Press, 1959); also S. M. Lipset, *Political Man* (Garden City, Doubleday, 1960).

valent to demonstrating the invariance of a statement regardless of assorted third factors. That this logic conforms to Mill's 'Canons of Proof, especially to the 'method of agreement', is evident. The process of explanation is usually simple, if indeed a proposition can be shown to hold cross-culturally. Especially if, in the study design the method of agreement is followed, there is no further need to specify the differences between the various contexts in which the statement of invariance between factors was studied. Some specification of this sort is specifically necessary for the 'method of concomitant variation'; this strategy is sometimes recommended as the best one for observational disciplines, but if cultures or countries are the settings in which the variation are observed, in view of the complexity of this context the explanatory process becomes quite difficult.

8. Using different cultures or countries as conditions for specifying the time-space coordinates under which statements will hold is a most promising but in practice often a very difficult strategy. Whether the logic of such comparisons follows the 'method of difference' or the 'method of concomitant variation', it is necessary in both cases to relate observed differences to specific aspects of the different cultures and to exclude other aspects as irrelevant. Specifying a finite set of factors as limiting the applicability of a proposition is more conclusively done by following the 'method of difference'; the 'method of concomitant variations', however, is especially well suited to survey research in internally differentiated collectivities. Regarding modern nation-states, it is unrealistic to assume that specific phenomena are completely absent; sample surveys indeed show that most responses occur everywhere, although with widely different frequencies. It is then the task of the survey researcher to relate the differences in response frequencies to the differing strength of factors in the collectivity (culture or nation-state).

To summarize my conception of the different use of 'culture' (or nation-state) in cross-cultural research, the following four-fold table might be helpful. (The distinctions - especially between ***a*** and ***b*** - have been stimulated by a speech by Alex Inkeles).

Treatment of 'Culture '

Purpose of Comparison	a. 'Culture, as an entity and a unit in analysis	b. 'Culture' as a set of conditions for units in analysis
1. Show similarities	(a) Identify universals (e.g., Murdock)[40]	(y) Demonstrate generality of propositions (e.g., Whiting)[41]

[40] G. P. Mordock, *op. cit.*
[41] J. W. M. Whiting, *op. cit.*

2. Show differences	(β) Distinguish between societies (e.g. Thurnwald[42], Redfield[43])	(δ) Specify time-space coordinates of propositions (e.g., Inkeles)[44]

Any of these strategies may be used in survey research. This classification scheme was proposed with the assumption that problems in comparative survey research are different in each of these cases. Accordingly, it is suggested that providing the explanatory processes in the design for the anticipated analysis, and choosing the cultures to be compared, be done with some such scheme in mind. A consciousness of the logic of accounting for differences is of prime importance in choosing the very units that make cross-cultural comparisons cross-cultural: the cultures of nation-states in which surveys are conducted.

Should the purpose of a comparison be to demonstrate similarities, a comparison across very different conditions is obviously the strongest test. What kinds of cultures are most suitable is a function of the use of 'culture'. Should the culture (or nation-state) be a unit in analysis, the choice of industrial societies (even if they are as different as the USA and the USSR) would probably increase the likelihood of the researcher's coming to grief; choosing very complex societies but widely differing by type of economic development would be the strongest test of a proposition's generality.

Whenever the demonstration of differences is the focus of cross-cultural comparison, it is preferable to select cultures (or nation-states) that are not too dissimilar in too many aspects at the same time. Otherwise the problem of relating differences in the dependent variables to system features becomes unmanageable, or is solved in the usual misleading manner of referring to presumably evident differences.

Whether the evident differences between cultures or nations that exhibit different values for the dependent variables are the true causative factors, remains very often open to question. References to these differences, more often than not, unwittingly appeal to stereotypes that the researcher shares with the unsuspecting readers in his own society. In matching e.g. the United States with some developing country, it may be plausible to explain the lower scores for achievement motivation among inhabitants of a developing country as causing the low stage of technical and economic development there; after all, the USA with its high scores has a much higher level of development. However, whether the higher averages for achievement motivation are a primary cause - or any cause at all - of the high level of technical and economic development is far

---

[42] R. Thurnwald, *Die menschliche Gesellschaft in ihren ethno-sozialogischen Grundlagen* (Berlin/Leipzig, Walter de Gruyter, 1931-35), 5 vols.

[43] R. Redfield, *The Papers of Robert Redfield,* Park Redfield, ed. (Chicago, University of Chicago Press). Vol. I, *Human Nature and the Study of Society* (1962). Vol. II, *The Social Uses of Social Science* (1963).

[44] A. Inkeles, 'Industrial Man: The Relation of Status to Experience, Perception and Values', *American Journal of Sociology,* 66(1), (July, I960), pp. 1-31.

from obvious to anyone who does not share the opinion that good living is the consequence of drive and hard work (especially if achievement is conceived, as by McClelland, as the motivation to get ahead of one's proverbial neighbor).[45] Thus, cross-cultural comparisons may actually serve to confirm stereotypes even about one's own society, rather than to overcome the culture-blindness of research.

In demonstrating differences, the strategy would differ again by the manner in which references to culture are introduced into the process of explanation: invoking culture as a unit is in most cases facilitated by choosing such nations or cultures that are not too internally differentiated. In specifying time space coordinates for statements or different forms of a general tendency, a choice of internally differentiated cultures or nations is - *ceteris paribus* - the optimal strategy. A good example is Alford's strategy of limiting his secondary analysis of the influence of social class on political behavior to English-speaking and British-influenced polities with high levels of technical and economic development.[46]

The classification system introduced here is meant as a heuristic device. Actual research projects may sometimes be fitted into these categories only with difficulty, or may fit into more than one of the cells. The latter case, however, may help to call attention to some ambiguity inherent in the reasoning process. As an example, I want to refer once more to the Almond and Verba study. This study is meant to demonstrate differences existing in different cultures and at the same time to prove one general proposition: a stable democracy is only possible when the citizens participation is, both mentally and in terms of actual behavior, non-parochial (types p and *y* of our classification). Accordingly, the authors first postulate that the different nation-states are units, in the sense that they are *entities,* and invoke them as accounting for the differences in the frequency of responses. Simultaneously, the character of the polities is postulated to be a direct consequence of the differences in response frequencies to various indicators for the trait: non-parochial, fully participating citizen, *and* manifest differences of the polities are referred to as proof for this argument.

Although cross-cultural comparisons are crucial for any generalizing social science, the optimal strategy is affected by the type of theorizing. It is usual to think of theorizing as one distinct orientation leading to one distinct class of statements about relations between factors. This won't do for our needs, and we have to be aware of different goals in theory which lead, or should lead, to different strategies for the testing of theory. To simplify, I suggest that we distinguish between three types of orientation in constructing 'general theories': a) the main purpose may be the identification of basic and simple factors underlying all the manifest diversity (e.g., Homans); or b) general factors may

[45] D. C. McClelland, *The Achieving Society* (New York, D. Van Nostrand, 1961).
[46] R. R. Alford, *Party and Society* (Chicago, Rand McNally, 1963).

be identified with the primary intent of accounting for all the manifest diversity (applied theory; probably even Parsons); or c) general theoretical statements may be advanced as distinguishing features of a specific type of society and polity (the 'industrial man' of Alex Inkeles)[47].

Cross-cultural research has been most often linked to the first (=a) kind of theoretical orientation. Especially in cultural anthropology, this took the form of the well-known search for 'universals' of human existence. By force of habit and precedent, this orientation of a theoretically oriented cross-cultural comparison is often looked upon as characteristic of cross-cultural comparisons in general. Survey research, however, is probably not often a very useful tool in this orientation, and accordingly it should by implication not treat this orientation as a model.

I believe that survey research finds its best use in connection with the second (=b) and third (=c) orientations in theorizing - provided it is accepted at all as a tool for testing elements of general theories. In this context I cannot spell out the consequences that I think are implied in this usage - with one exception. Obviously much of the diversity among societies can be accounted for by differences in the frequencies of demographic characteristics or other objective factors that themselves are not system properties. Before invoking the manifest differences between societies as proof of the 'causative effects' of the theoretically postulated main dependent variables, one should obviously control for the effects of these objective factors. Usually the design of a cross-cultural survey is not suitable for such controls. To implement the need for such controls might have rather far-reaching consequences for sample design - perhaps too far-reaching, so that we may have to rely here largely on secondary analysis. However, another consequence is the need to be more specific than is now usual in the design stage in identifying variables as dependent or independent; and there are fewer technical problems preventing us from doing just this.

7. Problem area V: Administrative and diplomatic problems

A critical analysis of many cross-cultural studies will show them to be methodologically less impressive than good studies done *within* a particular culture. This is certainly not due to the professional qualifications of those involved, since the same researchers have often done neater work in their own countries. Reviewers usually have not been too worried about this state of affairs; certainly these projects have not by and large been subjected to such rigorous methodological discussion as were, e.g., *The American Soldier,* the

[47] Examples of these three strategies in theorizing are: A. Inkeles, *op. cit.;* T. Parsons, *Towards a General Theory of Action* (Glencoe, 111., The Free Press, 1962); and G. Homans, *Social Behavior in its Elementary Forms.*

Kinsey Report, *The Academic Mind;* even the national election polls get more critical appraisal. The intrinsic interest that cross-cultural studies demand has probably shielded them from much scrutiny.

Why doesn't our cross-cultural research conform to more rigorous methodological standards? As outlined above, the methodological and theoretical problems are greater here than in research within one country. However, no less important are, in my opinion, the practical administrative and diplomatic problems involved in such research. Often these difficulties cause comparative research to fall short of optimal standards. There is agreement among most of those who have undergone this experience that the diplomatic problems involved in cross-cultural research are formidable. Cooperation is at times a rather trying experience, and in general problems in research administration multiply when one leaves one's own culture. Unfortunately it is somewhat hard to generalize from experiences; the morphology of cross-cultural studies is so rich that it is a bit hard to identify conditions. This is one suggestion for a typology of research situations:

1. One man has control over funds, design and execution of the study, and in essence employs cooperating researchers in other countries as his helpers. In this case there is obviously no problem in decision-making, and in preventing a heterogeneity of interests from adversely affecting a study. The value of the research will then largely depend on the degree to which the central decision-maker is able to elicit comments and to understand objections from his collaborators. In practice the procedure is remarkably different whether one is sub-contracting with commercial research agencies or with academic institutions. Academic institutions will exert their intellectual independence whether one likes it or not and even though formal control remains in the hands of one person.

Some examples of this type of arrangement of straight sub-contracting are the Buchanan-Cantril study, Lerner's investigation on elites, and the Almond and Verba international citizenship survey.[48] As Mark Abrams remarked recently, such an arrangement usually leaves the client more satisfied than his co-operators.

2. An international organization is established that employs nationals from various countries and has a supervisory body representing various nationalities. The UNESCO Institute for Social Sciences in Cologne (1951-1961) was a main example. Although this institute was usually not considered a great success, similar institutions have been founded or are being planned now in various parts of the world. A major asset of such a solution is the permanent staff that may develop into a truly cross-cultural body itself; a major problem is

[48] W. Buchanan, H. Cantril, *et al., How Nations See Each Other; A Study in Public Opinion* (Urbana, University of Illinois Press, 1953); D. Lerner and M. Görden, *European Leaders Look at World Security* (Cambridge, Mass., M.I.T. Center for International Studies, 1960), mimeo; G. A. Almond and S. Verba, *op. cit.*

decision-making and the staffing of positions at the director's level. The wisdom of supervisory bodies for such institutions has so far not been very impressive, and the usual personality problems in cooperative work between professors have often been compounded by national (and amateur) politics.[49]

3. *Ad hoc* committees may be appointed to see through a particular project from inception to publication; they may even have in effect the character of a temporary international institute built around one project. Schachter and Rokkan's cross-cultural project on teachers' attitudes may be considered a prime, and largely successful, example of this form of organization.[50] Many of the dysfunctional aspects of permanent institutes are avoided, but perhaps at the cost of lack of technical facilities. The rise of those 'costs' depends obviously on the degree to which a definite structure can be developed. I think it is one of the major advantages of this type of cooperative effort that it can be abandoned if it does not work - which is much harder in the case of an international institute (cf. the long agony of the UNESCO Institute in Cologne).

4. International working groups bear a certain resemblance to these *ad hoc* committees, except that they are characteristically organized around a certain research area instead of a specific research project. This usually has the consequence that the various members develop their own projects within the context of the group's program, and also have to provide funds for their research individually. Examples are the International Committee for the Social Science Study of Leisure, the International Statistical Institute and the various research committees of the International Sociological Association. The latter example also demonstrates the strengths and weaknesses of this type of organization: some of the research committees of the ISA are very productive while others merely exist. Such committees are slower in dying than *ad hoc* groups; they can stimulate a variety of research and serve clearing-house functions as probably no other organizational forms can do. Whether such committees work or do not work depends largely on the personalities of those that find themselves in such settings. European social scientists appear now to develop better skills in teamwork than they displayed so far, and this acquisition may enhance the usefulness of this form of internationally cooperative research.

---

[49] The group dynamics of such multi-national groups has been described by W. Baur, 'Sozialpsychologische Probleme einer multidisziplinären Studiengruppe', *Kölner Zeitschrift für Soziologie und Sozialpsychologie,* 14 (3), (1962), pp. 476-486.

[50] Some publications on organization and methodology resulting from this project: H. C. Duijker and S. Rokkan, 'Organizational Aspects of Cross-National Social Research', *Journal of Social Issues,* 10(4), (1954), pp. 8-24; E. Jacobsen, 'Methods Used for Producing Comparable Data in the OCSR Seven-Nation Attitude Study', *Journal of Social Issues,* 10(4), (1954), pp. 40-51: S. Rokkan, 'An Experiment in Cross-National Research Co-operation', *International Social Science Bulletin,* 7(4), (1955), pp. 645-652; and S. Schachter, 'Interpretative and Methodological Problems of Replicated Research', *Journal of Social Issues,* 10(4), (1954), pp. 61-68.

5. Most cross-cultural research that is not a one-man enterprise is conducted via successive bilateral agreements. In this way it is still occasionally possible to span the world, and even penetrate presumably Iron Curtains at a time when they were definitely more 'iron' than now. Gordon Allport's cross-cultural investigation of 'Youth's Outlook on the Future' is probably the most impressive example.[51] However, such bilateral agreements are usually limited in scope to two-country comparisons. As such, they serve a most useful function and are the most versatile solution to expand cross-cultural comparisons into a normal form of research. To stimulate such an expansion, better clearing-house facilities than are currently available would be most useful.

There is a continuous increase in the importance of international scholarly associations which is in turn most helpful in increasing the quantity of cross-cultural comparisons. Unfortunately there is little reason to believe that this will lead to an automatic concurring rise in quality. Many shortcomings could be reduced were the researchers more aware of the stages in the research process where close and real cooperation is especially important. I think that the phase in which a research concern is being translated into a tentative design is where closer cooperation than now exists is necessary. Quite naturally, researchers are hesitant to communicate when their own notions are still in a somewhat nebulous form; but it is exactly then that discussion with a colleague from another culture might be most helpful. Approaching another culture with an already formulated design and/or instrument, often amounts to imposing a conceptual scheme appropriate to one culture onto all others: I think of all areas of cross-cultural research, a number of prejudice studies are most often guilty of this.

In this realm of the administration and diplomacy of cross-cultural survey research, there has been little progress, and there exists certainly little awareness of the best strategies. It seems that every researcher wants to commit the same errors all over again. Obviously I could not give here a sort of sociology of the administration of cross-cultural research; I merely wanted to suggest some reflection on this, so that eventually we may have some benefit from cumulative experience.

8. Some social effects of comparative social research

Progress in the methodology of cross-cultural survey research, additional facilities to make use of such surveys, increased financial support for such studies - these come about when societies have an increased and immediate relevance for each other. Thus it does not seem much of a guess to forecast that cross-cultural comparisons will rapidly gain favor. Within this trend, surveys will remain the most facile tool for accomplishing such comparisons. I tried to

[51] J. M. Gillespie and G. W. Allport, *Youth's Outlook on the Future: A Cross-National Study* (Garden City, N. Y., Doubleday, 1955).

emphasize that such comparisons are a powerful tool for such a largely observational science as ours. But this tool is more difficult to handle than is apparent - and this paper was meant as a plea not to commit the same errors over and over again. Provided we learn from the experience of others, provided we make more often proper use than improper use of this tool - what will be some of its chief benefits?

I personally expect major benefits for methodology - not just for the methodology of such comparisons, but equally for an improvement of research within one's native setting. Unfamiliar research settings can stimulate inventiveness, and will force us to become conscious of seemingly self-evident assumptions in research. Specifically, there should be a better understanding of the interrelation between the type of theorizing one prefers, and the research operations.

Undoubtedly, increased information about other societies will have in some way (and I am afraid it needs to be phrased that vaguely) important social effects. I personally believe that in many cases such surveys will just be used to buttress existing misconceptions with numbers, much as travel abroad often strengthens previously held stereotypes. On balance, however, information will be substituted for beliefs.

For better or worse: decision makers in the field of foreign policy and consumer goods marketing already base their decisions partly on cross-cultural surveys. The media of mass communication are just about to be affected by such research.[52]

A few remarks on some other effects for the social scientists. It is a point too obvious to belabour here, that the substantive findings will both change and enrich our sociological knowledge. It will be an additional and less obvious benefit to make us aware how ethnocentric sociologists and political scientists have been during the last decades. We are just beginning to realise that sociology has largely become the sociology of the particular society the sociologist happens to be living in. Recently Robert Marsh computed that in the US of all Ph. D. theses in sociology between 1950 and 1960 only 12 percent used material from societies other than their own.[53] And this still over-estimates the degree to which real use (as opposed to haphazard quoting) of the literature from other societies is made. By now, our very conceptual schemes themselves are quite often translations of stereotypes from the mass media into sociologese. Examples are the 'egalitarian family' or the way democracy has been defined in some recent political science studies.

---

[52] E. K. Scheuch, 'Sozialer Wandel und Sozialforschung,' *Kölner Zeitschrift für Soziologie und Sozialpsychologie,* 17 (1), (1965), pp. 1-48.

[53] R. M. Marsh, 'Training for Comparative Research in Sociology', *American Sociological Review,* 27 (1962), pp. 147-149; cf. also E. Sibley, *The Education of Sociologists in the United States* (New York, Social Science Research Council, 1964).

Hopefully, cross-cultural surveys will become a rather common-place activity, and this will provide the experience that one's own society is not the sole earthly manifestation of a predestined ideal. Even by just looking habitually at marginals and cross-tabulations from other societies, these other societies may appear more common-place, and one's own society less self-evident. Ethnology in the 19th century was largely characterized by its approaching the non-Western societies as exotic. Exotism is the frame of mind with which many social researchers now tend to look at other societies.

The use of cross-cultural surveys as a relatively commonplace activity, as a matter of course - this I expect to change the style of thinking of social scientists, to affect their very conceptual apparatus. I expect that this availability of cross-cultural surveys will serve to help more of our colleagues to become alienated towards their own society in a professional capacity. This conjecture is based on my private and distinctly unsystematic sample of social researchers, and I shall continue to believe so, until someone does research on the effects of cross-cultural survey research on the person performing it.