SBORNÍK PRACÍ
FILOSOFICKÉ FAKULTY BRNĚNSKÉ UNIVERSITY
1970, G 14

I. GAĎOUREK

SOME RECENT TRENDS IN TECHNOLOGY OF SOCIAL RESEARCH

University of Groningen

Twenty years is a long time in the history of a modern science. This is even more true with reagard to sociology, still striving for recognition. Badly affected by the war and the authoritarian regimes that hampered its development, it has witnessed a rapid development in the West since the early fifties through to our days.

Several terms which are in daily use by sociologists today (such as "social role", "social network", "social system") were hardly known at the time when Professor Bláha retired and sociology ceased to exist as an independent discipline in Czechoslovakia. It is a thrilling mental enterprise to compare the state of sociology as a science some twenty years ago with its present state. Several "readers" make such a comparison possible, especially with regard to the sociological theory and to the various fields of application.[1]

Nowhere, probably, is the change more striking than in the way the sociological investigations are carried out. In the period between the wars, when Professor Bláha conducted his stimulating research-projects, the measurement theory was practically unknown, the modern mass-survey as based on a division of labour (between research-designers, interviewers, coders, statisticians and program-writers) yet in its cradle, and the statistical evaluation confined to a

[1] One of the first attempts at stock-taking of sociological progress after WW II was G. Gurvitch, W. E. Moore (eds.) *Twentieth-Century Sociology*, New York, 1945, Paris, 1947, with contributions among others of F. Znaniecki, P. A. Sorokin. In 1959, *Sociology Today* appeared under the editorship of R. K. Merton, L. Broom and L. S. Cottrell, Jr. It presents contributions by some thirty American (U. S.) authors and pretends to cover the entire field. About ten years later, in 1968, the indefatigable Talcott Parsons in cooperation with some 24 other authors makes another attempt at overlooking the field in *Knowledge and Society*. It is, of course, chiefly the sociology as it is seen and practised in the United States (S. N. Eiserstadt of the Hebrew University in Jerusalem is the only non-U. S. contributor to the volume). While the above mentioned works lay emphasis on theory and conceptual frame-work, an interesting inventory of research-results has been compiled by B. Berelson and G. S. Steiner who published their *Human Behavior. An Inventory of Scientific Findings* in New York, in 1964. This covers the entire field of behavioral sciences, including psychology and cultural anthropology and, obviously, lacks a unitary theoretical frame of reference.

single statistical test (usually a Pearsonian r) between two variables. With the exception of studies of local communities, the post-mail questionnaire formed *the* research technique of those days and written documents (including fiction) were widely used, though systematic content analysis was not applied at all or in a most rudimentary form. Social investigation then was rather descriptive and problem finding than evaluative and hypotheses testing.

This is no place to pass a final judgment on the work of the older generation of sociologists. Yet even superficial confrontation leads to the recognition of their inventivity, their originality of thought, and efficiency of work: much has been achieved with often scarce or insufficient means.

On the other hand, important developments in the methodology and the techniques of research took place that make old ways of investigation look rather obsolete.

There is, to begin with, a growing emphasis on evaluation, especially statistical evaluation, and on testing of functional or causal hypotheses. This requires measurement scales that make the application of statistics possible and computational devices (such as high-power computers) which allow a control of a large number of potential intervening variables. To make the picture round: the development of scaling techniques has consequences for the data-collection process. Not single stimuli but entire batteries are used to measure the latent dimensions of the subjects, each corresponding to certain formal rules of the scaling model under consideration.

We shall briefly deal with these fields in the reverse order, describing some trends in the technology of the data-collection, the measurement theory, the data-evaluation, and the model construction in sociology.

Data Collection

Observation of human conduct, of the material environment of groups and of individuals, together with a mail-questionnaire and verbal interviews are, since long, our main source of knowledge about the social and cultural phenomena. To these the direct analysis of symbolic vehicles, the direct interpretation of the mass-communication stream, should be added.

The content analysis, as the latter is called in the technical language of the social research, can be indentified neither with the observation (as the perception of the symbolic meanings is immediate and in this sense different from a *post hoc* interpretation of the perceived behavior) nor with the interview or the questionnaire (since the element of administering of stimuli is missing in content analysis; moreover, while interacting subjects form the object of the interview technique, the content analysis is focussed on the product of this interaction — various cultural vehicles that exist relatively independently of both their producers and their consumers).

This is equivalent to saying that material aspects of communication are hidden and subconscious, as are the motives and intentions of the producer of the communication. By making the hidden relationships between the form and the content, between the social and psychological aspects of either the sender or the receiver and the communication explicit, content analysis attains at a form of social perception *sui generis*.

There is another trend in data-collection than a mere emphasis on content analysis (quite understandable in the period of cultural isolation owing to the war and of the expansion of television). Lately we came to better realize that each of the basic perception techniques (observation, interview, questionnaire and content analysis) exists in two basic variants at least: one destined for exploration and one designed for evaluative research. Differences are so pervasive that we can better speak of eight instead of four basic forms of data collection. They are dictated by different goals which are set for the research-worker (inducing hypotheses, finding new general terms or research tools in explorative investigations; testing hypotheses in evaluative or experimental research) and result in two different attitudes of the data-collector.

In the first case, he lets himself be impressed by his subject of investigation, his frame of reference is not rigid but allows the facts speak for themselves, modifying even the theme of the study itself. In the second case, the research-organization resembles one huge punching machine leaving distinct, unequivocal mark for every registered feature of the subject (there being a large number of subjects systematically selected according to a premeditated sampling plan). The reliability of registration is the main goal; this is striven for by a rigid operational translation of the theoretical concepts in detailed procedures of perception; by a standardization of perception either by mechanical means (film, band-recording) or by the division of labor (trained teams of observers, interviewers or coders working with detailed, standardized schedules); and, as already mentioned above. by the selection of a probability sample from the universe under study.

Better to grasp the distinction between the two kinds of techniques, the reader may compare for himself the observation-techniques of a tourist or a casual traveller with the techniques as developed for the observation of small group interaction in laboratories (one-side screens, synchronized devices for recording interactions, a coda-book designed for the benefit of observers, film and/or band-recording equipment, etc.).[2] The latter techniques also can be used in the field-research, in the study of committees or of various working groups.

With regard to the stimulus-response techniques (questionaire, interview) the two distinct approaches can be summarized as show the scheme on p. 90.

Another insight that is gaining ground is the recognition of the demands that a certain scaling-technique lays upon the design of a questionnaire. We are free no longer to choose deliberately the form of items but have to consider the data-model we wish to apply.

If, for instance, we wish to work with a deterministic model of Guttman-type, monotone items should be selected. This means that a stimulus should cut the sample into two or more parts, each with a cumulative intensity of the property measured. "Are you between 170 cm and 180 cm tall?" is an often quoted example of a nonmonotone (to be avaoided!) item. A negative answer makes it impossible to place the respondent on a monotone latent dimension of length (the person can be taller than 180 cm or shorter than 170 cm). Some open attitude questions or questions being adorsed instead of answered in the way indicating intensity (for instance: "Are you against or for the American intervention in Vietnam?") illustrate the case in interview-research.

Even more demanding are the judgment methods. The clumsy word of

---

[2] A very early manual is R. F. Bales, *Interaction Process Analysis,* Cambridge, Mass., 1950.

	EXPLORATORY APPROACH	STANDARDIZED, STRUCTURED APPROACH
Function:	the collection of information about others or about the respondent's frame of reference	classifying respondents in the system of categories
Form:	real questions; 'probing'; no response categories suggested	questions with printed alternative responses; statements for adorsement (with categories expressing the degree of adorsement); subjects to choose from or to compare
Registration:	laborious and time consuming shorthand or band recording	simple and efficient; checking or marking
Elaboration:	a classification-system of themes	assigning of address numbers making mechanical sorting or computer evaluation possible
Respondent's role:	spontaneous	passive
Satisfaction and involvement:	high	low
Reliability:	low	high
Validity:	high	low

"searchingness" which C. H. Coombs coined for an ingenious classification of the techniques for sorting and comparing of stimuli, suggests the varying amount of information which each technique implies.[3]

Suppose we wish to ascertain the occupational prestige of ten jobs in a given population. Obviously, several alternative ways of administering the stimuli are available. We may present the names of the stimuli on cards and ask the subjects to pick up one, two, etc. to nine, to which they ascribe prestige. In addition to these nine possibilities we also can instruct the subjects to select two, three or i-stimuli that are high (or low) and then let them order the selected stimuli according to their relative prestige. The latter method reaches its limit in the method of rank-order (when all stimuli are ordered in a hierarchy of high down to low prestige jobs). Moreover, as has been shown long before, the ranking can be performed with a higher precision if the stimuli are administered all in pairs of two and systematically compared with each other. This method of paired comparison offers a maximal amount of information extracted from the respondents while no double work (no "redundancy" in terms of Coombs, *op. cit.* p. 36) is implied. Yet other techniques are possible: that of triads in which all combinations of three stimuli are analyzed by the method of paired comparisons; or we adminster all possible tetrads or higher order series to be treated accord-

[3] We draw on Warren S. Torgerson: *Theory and Methods of Scaling,* New York, John Wiley and Sons, 1958. Clyde H. Coombs: *A Theory of Data,* New York, 1965, is also a must for every serious student of the measurement theory in the behavioral sciences.

ingly. The potential information raises only slowly in these latter cases, at the costs of the steeply increasing redundancy.

By applying the elementary formulae of the information theory to his concept of searchingness, Coombs makes it clear that varying degrees of information about a certain atribute can be extracted from the subjects as one uses the alternative ways of administering the stimuli.

As k stimuli can be picked up out of a series of p stimuli in $\binom{p}{k}$ different ways and the number of ways a subset of size k can be ordered is $k!$, the total number of possible orderings k out of p is

$$k!\binom{n}{k} = \frac{p!}{(p-k)!}\,. \tag{1}$$

The number of ways an individual can pick up k stimuli is, of course, given by the formula

$$\binom{n}{k} = \frac{p!}{(p-k)!\;k!}\,; \tag{2}$$

this is much less than when we order the data.

By applying the simple formula of the information theory concerning the amount of information (I), we get for the equally probable, independent n stimuli

$$I = {}^2\log n. \tag{3}$$

Thus we have to apply logarithms to the base two to formulae (1) and (2) to get the channel-capacity for each sorting method:

$$C_{\text{order } k:p} = \binom{n}{p}\,{}^2\log\frac{p!}{(p-k)!} \tag{4},$$

where $\binom{n}{p}$ is the number of presentations.

$$C_{\text{pick up } k:p} = \binom{n}{p}\,{}^2\log\frac{p!}{(p-k)!\;k!}\,. \tag{5}$$

We see that ordering increases the amount of information by $k!$ bits ("bit" is a term contracted from "binary digits" and denoting a unit of information in the sense of the foregoing formulae).

Applying these elementary formulae, we get for the method "pick up 2 items out of ten"

$${}^2\log\frac{10}{(10-2)!\,2!} = 45 \times {}^2\log 45 = 45 \times 5{,}5 \text{ bits},$$

where 5,5 bits of information is won by each presentation. The method of rank-order gives us maximally 22.8 bits, that of paired comparisons 45 bits, that of triads maximally 310.2 bits of information per presentation for the equal number of stimuli. The latter is, of course, redundant. A simple measure of redundancy of method x in terms of paired comparisons is

$$R_{x:\text{order } 1:2} = \frac{C_x}{C_{\text{order } 1:2}} - 1\,. \tag{6}$$

By using the corresponding numbers for the method of triads, still following Coombs (*op. cit.* p. 37—8), we get

$$R_{order\ 2:3} = \frac{310.2}{45} - 1 = 5.89 \text{ bits.}$$

These are interesting attempts at formalization of what used to be a very clumsy field of methodology. Yet we have to bear in mind that the formulae give rather the upper limits of potential information than its real amount. The basic assumptions (equal probabilities of choice, and independence) are rarely met.

Moreover, the experience learns that "searchingness" is not the only selection criterium for items, as it runs counter to "feasibility". The higher the "searchingness" of a method, i. e. the higher amount of information is extracted from the respondent on a given matter, the greater amount of bother and bore the method entails. While it is possible to put a few cards on the table in the respondent's home and ask him to pick up a few or to order a few, the ranking methods meet with frustrations and opposition, already, while the use of some fifty occupations in the research into social stratification by the method of paired comparisons surpasses the limits of the psychological possible. The method of triads can be used with highly motivated respondents only when the number of stimuli is not too large (some vocational guidance tests). Ranking of all combinations of four or more stimuli is, probably, only of theoretical interest.

The concept of communication-channel that is referred to in the foregoing section finds a wider application. Recently, Johan Galtung from Oslo University, examines the entire process of data-collection through the eyes of an information theorist.[4] According to him, the respondent *sends* communication that passes through several channels (those of perception, recording, coding, punching, and tabulation or evaluation) before it is *received* as a processed datum. At each step of data-collection and data-processing a certain amount of "noise" distorts the communication, traditionally referred to in terms of error (thus: coding-error, punching-error, computational-error). Traditionally, the research-worker vaguely assumed that the reliability of his data was determined by the largest error-margin of one of the steps or he did not consider the total error at all. As Galtung shows, the total amount of information-noise exceeds the error-margin of the separate steps. The product rule of statistics applies here, and we have to construct a matrix of transmission probabilities, corresponding to a simple Markov-chain,[5] to obtain the total amount of noise.

Unter the reasonable assumption that the single steps are statistically independent of each other, the information transferred amounts to the product of the probability matrices of the single steps

$M = M_1M_2M_3M_4 \ldots\ldots M_n$. (n += the number of steps in the process of data collection)

---

[4] J o h a n G a l t u n g, *Theory and Methods of Social Research,* Oslo, Universitetsforlaget, 1967, pp. 174—178.

[5] A very useful handbook is J. G. K e m e n y, J. L. S n e l l: *Finite Markov Chains*, New York, D. van Nostrand and Co., 1960.

The transmission probability matrix of each step is given by:

$$\begin{array}{ccc} & & \text{signal received} \\ & & \begin{array}{cc}\text{yes} & \text{no}\end{array} \\ \text{signal sent} & \begin{array}{c}\text{yes}\\ \text{no}\end{array} & \begin{vmatrix} 1-P_i & p_i \\ q_i & 1-q_i \end{vmatrix} = M_1. \end{array}$$

Suppose that $p_i = q_i$ (which is not unreasonable to assume) and that all probabilities are of the same value (which is unlikely), namely .50, the usual error-margin tolerated in the survey-research. Then the product of $M_1M_2 = M_1^2$, that of $M_1M_2M_3 = M_1^3$, etc.

$M_1 =$	.95	.05	$M_1^2 =$	.905	.095	$M_1^3 =$	.87	.13	$M_1^4 =$	.84	.16
	.05	.95		.095	.905		.13	.87		.16	.84

Thus in no more than four steps the amount of information-noise is three times as high as the expected error in the single steps in this specific example (when $p_1 = q_1 = .05$).

The obvious consequence of this would be to limit the number of steps in the collection and the processing of data and to keep the error-margin per step as low as possible. In our experience it is especially the coding of data which accounts for a high percentage of noise. When separate coding sheets and coding books are used the error soars as high as .10 to .15, per response, owing to the slackened attention of the coders (especially when the number of subcategories is high). The translation of verbal materials into numbers generates thus much more noise than mere punching of number on Hollerith cards. A remedy should, therefore, be sought in pre-coded interview-sheets which can be used for direct punching (there being a code-number printed together with every subcategory of response). Mark-sensing cards (which are Hollerith cards on which responses are marked by special graphite pencils) that eliminate punching are less urgent. There also is their limitation owing to the uneconomic use of the space of the cards and the relatively high frequency of error in marking.

By eliminating the error of coding, one also limits the error in recording (thanks to the use of pre-coded response categories). The number of steps is cut down; it is, anyhow, lower than the seven steps mentioned by J. Galtung (*op. cit.* p. 174): (1) manifestation; (2) perception; (3) recording; (4) coding; (5) transfer; (6) punching; (7) tabulation. Of these, (1) and (7) can be neglected in practice, there being few mistakes made by computers or tabulators, and in the process of "sending a signal" (which is uncontrollable, anyhow). Transfer (5) is eliminated by using interview-schedules directly for punching. Thus the number of steps is limited to two or three: (a) perception; (b) recording and coding; and (c) punching. Of these, punching is a source of minor errors when the usual technique of control-punching is used. It strikes us, however, that Galtung does not mention one important source of noise in survey-research, namely the sampling error.

Our discussion of the progress in data-collection would be incomplete without a reference to the newer theory of sampling. The latter has been involved in a polemics about the application of statistical tests in sociological research. H. Selvin, R. Freedman and others raised the problem whether statistical inference may be drawn from samples that have not been randomly selected. R. McGinnis, M. J. Brouwer seem to come to a qualified positive answer: samples of subjects can be meaningfully compared even though the subjects are not randomly allotted to the experimental and the control groups, *if we multiply controls.*[6]

As any student of statistics knows, it is especially the field of multivariate and multivariable analysis, allowing the simultaneous control over a countless test-variables, that expanded a good deal ever since the high power computers began to be used in the 'fifties.

Leslie Kish raised a less general, but even more salient, point: the distortion of samples owing to the clustering effect.[7]

By drawing multi-phase samples from the population (for instance, by drawing randomly communities and then selecting inhabitants from the community-files; by randomly sampling the newspaper-issues and then selecting sentences for content-analysis; by sampling schools and drawing the pupils) we are inclined to forget that the variable studied may be correlated with the criterium for the selection of the primary sampling units (communities, newspaper-issues, schools). It then makes a difference whether we choose 100 persons out of ten communities each or quotas of ten persons out of 100 communities to obtain 1,000 respondents (the latter, being, for obvious reasons less biased).

The "clustering bias" is a nuisance when samples are used for descriptive purposes (the percentages, the means and other parameters of the population); it also hampers the application of statistical tests, most of which are designed for random samples only. Fortunately, we can estimate the degree of distortion by computing the variance of the characteristic studies *between* the primary sampling units. The resulting coefficient of clustering is of importance for the sampling-design: we use it to calculate the optimal relation between the number of primary sampling units and the total number of individual observations. Theoretically, as many primary units as possible should be used. But we know that the more scattered the individuals, the larger the costs (travelling, detection) and the energy bestowed on the collection of data per individual. Thus we get a simple set of equations with the known degree of precision required, the total size of the sample, the stratification of sample (into the primary or secondary units) as unknowns, and the correlation of clustering and the costs per observation unit as constants (to be ascertained empirically). We solve it by minimizing the costs (or labor) factor per individual observation. The application of linear programming to the planning of the social surveys is an important step towards the rationalization of research that used to be governed by heuristic principles and the rules of thumb, up to now.

---

6 See for the discussion between Selvin, Gold, Beshers, and McGinnis in the *American Sociological Review*, 22 and 23 (1957 and 1958). R. Freedman, M. J. Brouwer and R. J. Mokken discussed the same matter in *Sociologische Gids*, V (1958) and VI (1959) in English and Dutch.

7 Leslie Kish, *Survey Sampling*, John Wiley and Sons, 1965; see also his contribution to D. Katz, L. Festinger: *Research Methods in Behavioral Sciences*, New York, 1953.

The problem of measurement of human behaviour (i. e. its psychological and sociological dimensions) has been attacked succesfully by three different disciplines.

Factor-analysis, after the original setbacks and vain attempts to find general factors determining the behaviour as manifested in tests, came of age: L. L. Thurstone's *Multiple Factor Analysis* appeared in 1947, and was followed by publications of other writers who freed this technique of some of its shortcomings and gave it a sound mathematical base.[7a] Computers made some older but unpractical techniques (such as H. Hotelling's Principal Components Analysis) easy to perform; new techniques were designed for computers to standardize the rotation and free it from its by origin arbitrary nature (e. g. Henry F. Kaiser's Varimax-method; see "Computer program for varimax rotation in factor analysis", in *Educational and Psychological Measurement, 19,* (1959), pp. 413—420).

The relation of factor-analysis to the measurement theory can be sought in the problem of validity or — to use a newer term — dimensionality. By trying to reduce a larger number of test-scores to a limited number of "common factors", by posing the question of the rank (and "space") of the correlation-matrix, the factor analysis gives its own contribution to the core-problem of the measurement theory in the social science — that of dimensionality.

It is the lasting merit of Louis Guttman, Samuel Stauffer, P. Lazarsfeld and a few others, that they formulated the problem independently of factor-analysis.[8] Like the analysts, the protagonists of the scaling techniques started by looking for unity in diversity. Common factor theory was replaced by the theory of unidimensionality. The latter tried to tie together a number of stimuli of different connotation and of gradual order of "popularity". This order is basically of a monotone nature, i. e. the items of a scale evince a cumulation of positive responses: those passing a difficult test (of low "popularity") must necessarily pass an easier test.

The test of scalability of dichotomous items is based on the finding that we obtain $n + 1$ response-patterns if the items form a scale, as contrasted with 2 patterns if the items are independent, without any functional unity (this means a ratio of 8: 128 for a 7-items scale, and 9: 256 for an 8-item scale; Coombs calls this degree of contrast between the scalable and the non-scalable patterns the "vulnerability" of the scale).

Years of work with the scales of Guttman-type have taught us (1) that their construction is only justified when there is a difference between the manifest

---

[7a] I draw chiefly on Harry H. Harman: *Modern Factor Analysis.* The Univ. of Chicago Press, 1964 (Third Edition). Also see R. Cattell (ed.): *Handbook of Multivariate Experimental Psychology*, Chicago, Rand McNally and Co., 1966. Also W. W. Cooley, P. R. Lohnes: *Multivariate Procedures for the Behavioral Sciences,* John Wiley and Sons, 1962, containing computer programs in FORTRAN.

[8] The publication of the fourth volume of a series *The American Soldier* under the title *Measurement and Prediction* (Princeton University Press, 1950) containing articles by the above mentioned authors, formed a milestone in the thinking of many a research-worker. See for a comprehensive treatment of modern scaling techniques W. S. Torgerson, *Theory and Methods of Scaling,* New York, 1958; C. H. Coombs, *The Theory of Data,* New York 1964 (second printing 1967).

and latent aspects of the items used; (2) that the difficulty to find a similar empirically relevant scale that still satisfies the criteria of scalability (combined coefficients of reproducibility of Guttman and Green) is so great as to raise the question of the principal soundness of the approach.

With regard to the first point we wish to emphasize that "scaling" is already a phase of empirical research; if a number of seemingly different items can be proved to indicate one (latent) dimension, our knowledge of the empirical world is enhanced. In all other cases scales are merely gadgets for play. Countless scales can be designed by using, for instance, questions about time-interval: Did you see the doctor (or: have a drink, read a book, visit a restaurant), this week? this fortnight? this month? these six months? this year?

Answers are likely to form a strict cumulative, monotone pattern. Yet little is won, for, semantically, the items used only differ with regard to the time-category used; and there is no gain of information from those categories forming a monotone scale.

Empirical unidimensional scales are difficult to obtain to the extent that we may doubt their existence in certain fields of research on attitudes, values, or norms. If after repeated substitution of items one battery is found roughly satisfying the criteria of scalability, is there a reasonable likelihood that the sample of items is representative for the universe of contents?

It is this and similar sceptical questions that made several social investigators abandon the search for unidimensionality. Some of them turned to the factor-analysis of items (those of the Likert-type or the items of semantic-differential which seem very appropriate for the purpose) reasoning that unidimensionality cannot be attained if more common factors are at stake, and expecting that factor-analysis might help to identify those factors.

Yet others tried to generalize the insights won by the scaling to multi-dimensional measurement models. Several alternatives have been designed, some of them also applicable to non-metric data.

Unfortunately some of them (for instance several techniques of cluster-analysis or L. Guttman's facet-analysis) do not provide scores for analysis of external relationships with other variables; they give us insight into the internal structure of data, which is considered as a goal in itself.[9] Though of highly formal nature

[9] Torgerson and Coombs (see note 8) themselves developed new techniques in this field: J. B. Kruskal in two articles in *Psychometrica, 29,* of 1964, gives a lead in non-metric multidimensional scaling, followed by a Dutch scholar E. E. Ch. I. Roskam in his doctoral dissertation on *Metric Analysis of Ordinal Data in Psychology,* VAM, Voorschoten, 1968; he presents "models and numerical methods for metric analysis of conjoint ordinal data in psychology". Relatively independently of Kruskal c. s., L. Guttman and J. C. Lingoes developed "Non-metric Factor Analysis: A Rank Reducing Alternative to Linear Factor Analysis", *Multivariate Behavioral Research,* February 1967. For Guttman's earlier — and more detailed — thinking on the subject see his "A new approach to factor analysis: the radex", a contribution to P. F. Lazarsfeld's, *Mathematical Thinking in the Social Sciences,* Glencoe, Ill., 1954.

There are several approaches to cluster-analysis published in articles in various journals under different names: linkage- or pattern-analysis of L. L. McQuitty, hierarchical grouping of J. H. Ward, agreement-analysis of H. E. Watson, profile-analysis by J. Nunnally, profile-similarity by F. M. du Mas, etc. They were recently summarized and mutually compared in a Dutch socilogical dissertation of E. J. Bijnen. *Cluster Analyse. Overzicht en evaluatie van technieken* (Cluster Analysis: A Survey and Evaluation of Techniques), Tilburg, 1969.

(Guttman often uses a matrix of correlations as a basis for his non-metric analysis!) they serve rather a descriptive purpose, grouping or re-grouping the objects studied. In so far the variables and the construction of scales are a matter of deliberate choice, we prefer the unidimensional or at least — homogeneous — tests that can be correlated with predictors and themselves be submitted to multivariate causal analysis.

This term "homogeneous tests" brings us to the third approach (next to factor-analysis and to scaling) to quantification in social and behavioral sciences: that of a classical test-theory. It was developed independently of scaling. Jane Loevinger, L. J. Cronbach and some other test-constructors, ever since item-analysis was introduced in the field, looked for safeguards of validity of their tests. The measures of homogeneity as developed by them approach the values of reproducibility as measured by the techniques of Guttman and Green.

We may go even further and raise the question of concurrence in the three fields. In our view it is a pity that research is carried out in the parallel, yet isolated fields, while attempts at mutual comparison and synthesis are scarce. The functional relationship of the terms "common factor", "unidimensionality", "reproducibility", "homogeneity" should be examined both mathematically and empirically. It is true that L. Guttman has constantly sought alternatives to the factor-analysis. His unidimensional scales can be viewed as counterparts to Spearman's vain effort to find a single common factor battery of tests; his simplex, circumplex and radex are all attempts to find alternatives to multiple factor theories of L. L. Thurstone and the post-Thurstone generation of factor-analysts. He emphasises the element of *order* that he wishes to preserve even at the costs of the data-reduction principle to which the factor analysis adheres. Yet, as we remarked already with regard to Guttman's smallest space analysis (SSA I—III; it has been programmed by Lingoe in FORTRAN language for the electric computer[10]), the functional justification of these techniques, the delineation of their position in the research-projects and the quest for explanation, have not been spelled out, as far as we know.

In our view a more promising approach is that of measurement in its proper sense. B. J. White and E. Saltz made an interesting effort to compare several approaches to reproducibility. In addition to those employed by sociologists, L. J. Cronbach's phi- and alpha-coefficients prove to be useful measures, easily to compute and allowing a generalization to the whole battery of tests.[11] They even can be used as tools in the searching for suitable items to which we intend to apply Guttman's and Green's tests, since they are relatively independent of the marginal totals.

In the meantime, factor-analysis (and discriminant analysis) should not be frowned upon by sociologists. Scales can be constructed by them consisting of items that are highly loaded with a common factor. Though the scoring is based on other principles than that of order (especially as derived from the various popularities, i. e. marginal totals) it gives a good insight into the underlying dimensions of a battery of items.

---

[10] See *Behavioral Science,* volumes 10 through to 13 (1965 to 1968) for a short description of these programmes.

[11] We draw on B. W. White and E. Saltz, "Measurement of Reproducibility", in *Psychological Bulletin,* Vol. 54, no. 2, March 1957.

Cluster-analysis as well as the metric analysis of conjoint ordinal data (see note 9) confirm the essential soundness of the modern factor-analytical approach to the measurement problem.

Already implied in the foregoing discussion is the problem of measurement-level. The measurement-theory has gained by S. S. Stevens' lucid formulation of postulates (those of order, interval or distance and of origin) of the real numbers series that can be projected into the data.

As far as the empirical data satisfy these demands for order, distance or known origin, we are allowed to speak of ordinal, interval or ratio scales. The degree of freedom in ascribing numbers to objects is in descending order: any increasing numbers can be ascribed in ordinal scales, linear transformation is allowed for interval scales (of the type $y = ax + b$) while, in the ratio-scales, only the measurement-unit can deliberately be chosen (the transform being of $y = ax$ type).[12]

Practically all scales of attitudes, values or norms, most psychological tests and inventories merely attain the ordinal level. Thus, if we assign to a person a higher rank on the authoritarianism scale or if we give him a higher estimate of his I. Q., we only know that he is likely to be more authoritarian or intelligent, without knowing exactly how much more. Computations of means and standard deviation is senseless since no safeguards are given that $8 - 7 = 6 - 5$ or $4 - 3$ or $20 - 19$ (intervals being unknown). This, we shall show instantly, has consequences for evaluation of research-results. This justifies the ever lasting efforts to obtain metric information out of ordinal data. The recent techniques of Kruskal, Guttman, Lingoe and Roskam justify a hope that computer programs will enable us to transform the by origin ordinal data into metric scales, thus solving the urgent problems of quantification in the social and behavioral sciences.

Evaluation of Data

One of the important consequences of Stevens' measurement theory was the general mistrust of parametric statistical tests that have been improved by R. A. Fisher and others to powerful scientific tools. Early in the fifties, countless non-parametrical tests (those not presupposing the knowledge of variances, means and other parameters), appear in the scientific journals, soon followed by attempts at their codification.[13]

In Siegel's book the impact of the theory of games (becoming popular at that time) upon statistical inference could be noted. The two kinds of error that are usually made by drawing statistical inferences from data can be visualized as a simple two persons game as follows:

[12] See S. S. Stevens, *Handbook of Experimntal Psychology,* John Wiley and Sons, 1951; I draw on the 1963 edition, chapter 1. Also see W. S. Torgerson, op. cit.

[13] W. J. Dixon, F. J. Massey, *Introduction to the Statistical Analysis,* New York 1951; S. Siegel, *Nonparametric Statistics for the Behavioral Sciences,* New York, 1956.

		In fact: H₀ is true	Alternative hypothesis is true
The decision is:	to accept the null-hypothesis (H_0)	right decision	Error II (β)
	to reject the null-hypothesis (H_0)	Error I (α)	right decision

The pay-off matrix is, of course, the matrix of probabilities; thus we may assume an inverse relationship between error I and error II: the stricter our demands on the right rejection of the alternative hypothesis (α), i. e. the higher significance level we deliberately choose, the higher the chance that we shall return with empty hands, that no causal association will be ascertained even if they exist in reality (β).

One should look for a *reasonable* but not an exaggerated level of certainty ($P_\alpha \leq .05$ is usually preferred in our own studies).

It can be shown that errors made are reversely related to the number of individual observations (N) and to the nature of the test itself. The latter can be proved if N and α are kept constant, and is expressed as the power of a test (P_w).

$$P_w = 1 - \beta.$$

Parametric statistics are of higher power than those of the nonparametric kind. Siegel expresses the power efficiency of the latter in terms of the former:

Power efficiency of test B $= (100) \frac{N_p}{N_b}$ percent. If for instance we need 50 persons to test a hypothesis (at a constant α) by a test B and only 40 persons in a parametric test, we conclude that the test has a power efficiency of 80 percent. Wherever possible, Siegel presents the nonparametric tests together with their estimated power efficiency (usually based on the t-test).

This was an important development. For the first time, new techniques appeared to be simpler and easier to grasp than the old ones: psychometrics was to be replaced by a scalogram board, the correlation and the analysis of variance by simple functions of the chi-square tests. Moreover, the complex experimental designs or work-extensive survey, were modified, too. F. Stuart Chapin and E. Greenwood deserve their merit for the propagation of a controlled study-design under the somewhat barbaric term of the export-factor experiment.[14] By selecting individuals either according to the assumed effect of an unknown variable that has operated on them (a retrospective design) or according to the known cause of yet to prove effects (a prospective design) and by matching the selected group (either individually or group-wise) with a control group a very efficient research-design is obtained, requiring a minimal effort

[14] E. Greenwood, *Experimental Sociology*, New York 1945; F. Stuart Chapin, *Experimental Design in Sociological Research*, New York 1947.

of data collection and data-processing. This is chiefly due to the reduction of the original sample by matching.

It is queer that this consistent structure (ordinal scales as combined with non-parametric tests and an efficient design of controlled study) did not become more popular among sociologists. Its real triumph was marked in the field of social medicine (or, if we forget that sociologists were not involved: medical sociology) in explaining the deaths of long-cancer by excessive smoking habits.

We must, of course, acknowledge that any ex-post-factor experiment is but a substitute for an experiment, as based on (a) a genuine manipulation with the independent variable; and (b) the comparison with a control group, equal in all aspects (the allotment of individuals to either group being guided by the rules of randomization). Unfortunately, real experiments remain limited (at least in the Western countries) to the fields of communication and small group research (there being even few experiments in industrial setting, owing, among others, to institutionalized bargaining about working conditions with the trade unions; it is impossible, for instance, to "operate" with renumeration, replacing the piece-rate by group-tariff).

It is this lack of manipulation and of randomization that met with criticisms of Stauffer and Selvin (see note 6). As mentioned above, McGinnis countered this criticism by pointing to two different conceptions of causality and to the necessity to multiply the controls in testing the "finitely conditional hypotheses".

The control by matching or multiphase sampling , may turn out a less efficient device than the multivariate analysis (or its nonmetric counterpart), in this respect. Large matrices of order 60 (variables) $\times$ 2200 (individuals) or more correlations can be computed in a few minutes, as experience taught us.[15]

Not only that; multiple or partial correlations of high order can be computed by allowing the computer to search for the variables explaining the largest part of variance. Another possibility is to apply various programs of factor-analysis (principle components, canonical factor-analysis) together with a standardized rotation program (Kaiser's Varimax or Promax) often with estimated (by iterations) communalities along the main diagonal.

The advantage of analyzing the synchronic influence of a large number of variables that these parametric techniques offer over the control of three to seven variables by sub-sampling or matching is so evident that we are willing to pay for it by violating some assumptions and established rules.

There is also the necessity to apply one kind of test to a heterogeneous set of variables. These variables may not be normally distributed, which can be easily ascertained by inspection. As the number of individuals is large this needs not be a serious handicap. What seems even worse is the different

[15] See my *Hazardous Habits and Human Well-Being*, Groningen 1963 (in Dutch and English), for correlations and factor-analysis of 34 variables over a representation sample (1297 persons) of the Dutch population; or my *Absences and Well-Being of Workers*, Assen 1965, an ex-post-facto study of 107 working groups from large industrial plants in the Netherlands (N = 2,227 workers); recently, a correlation-matrix of 105 subjective items was computed and a principle component analysis (as combined with the Varimax-rotations) applied. Also see R. Wippler: *Social Determinants of Leisure Behavior*, Assen 1968 (in Dutch, provided with an extensive summary in English), for an analysis of 65 variables (N = 883 individuals, selected from a larger representative sample including housewives and aged persons).

measurement-level of the variables concerned: they may imply dichotomies (man-woman differential), nominal classifications (political preference, religious denomination, region of origin), ordinal attitude scales or some genuine interval and ratio-scales (incomes, age, number of years living in the residence, etc.).

How to standardize the tests? It might seem imperative to use the lowest measurement level to choose a test. The trouble is that we lack suitable coefficients that would express both the significance and the intensity of the relationship measured. The solution we decided to adopt was to use the parametric tests (Pearson's product moment r), but only as a means for detection of relationship (thus abstaining from drawing the conclusion as to the nature of the relationship). Our confidence, based so far on empirical experience (we applied to hundreds of variables both parametric and non-parametric tests and compared the results) has recently been confirmed by the Dutch statistician H. de Jonge. By using the computer to stimulate data (a Monte Carlo technique for the exact permutations test) he was able to show that for monotone data Student's test gives results not deviating from those as achieved by some nonparametric tests (that of Yates and Cochran or Wilcoxon), which he also proves mathematically.

Though encouraging, this solution is not an ideal one, nor the only possible one. One alternative would be to use special computer programs of Kruskal—Roskam type to raise the ordinal scale to metric level and only then draw a correlation matrix of thus quantified material (some doubts about the feasibility of this still plague us). Another alternative would be to use one or another kind of factor analysis as based on weak assumptions that begin to appear in the statistical literature. Finally, finite mathematics also makes progress and offers new solutions how to disentangle the intricate matrix of social relationships.

Impact of Technology on Model Construction

The limited space and the actual scope of this short survey do not allow to enlarge on the problem of synthesis of research results into a unitary frame of reference. Only a few trends will be mentioned here, mainly in so far as these concern the research-process itself.

There is, to begin with, the possibility to simulate social processes by means of computers, to let the computer generate (hypothetical) data. The use of Monte Carlo techniques (in computation of random numbers tables to facilitate the random sampling, in computation of permutations and stochastic distributions) is sufficiently known, yet not fully exploited.

Simulation also finds application in the game-theory that still waits for practical use[16] (some attempts have been made to apply it to international affairs and to build a theory of rational conduct of states).

Markov-chains also proved useful to analysis of social mobility, of some "social contagion" (communication) phenomena or the cases of social change over time.[17]

---

[16] We draw on R. Duncan Luce and Howard Raiffa: *Games and Decisions. Introduction and Critical Survey*, John Wiley & Sons, 1957 (Second printing, 1967). For simulation we consult T. H. Naylor, J. L. Balintfy, D. S. Burdick, Kong Chu: *Computer Simulation Techniques*, John Wiley & Sons, 1966.

[17] An insightful book is James Coleman: *Introduction to the Mathematical Sociology*. The Free Press, Glencoe, 1964. Also see J. L. Kemeny, J. L. Snell, *Mathematical Models in the Social Sciences*, Boston, Ginn & Co., 1962.

Personally, I cherish a high expectation of the theory of finite graphs which I used to apply to the analysis of causal matrices ever since the early fifties and that has received special attention in the recent times.[18]

This theory is not only suitable for an exact and efficient analysis of a sociomatrix or of various kinds of communication networks by means of matrix-multiplication which a computer easily performs (but which can be done mechanically, as well). It also can be applied to matrices of causal relationships to uncover clusters of interrelated variables. These can be analyzed more in detail by partial correlation, by path-analysis[19] or their nonparametric substitutes, or by computing the second or higher powers of the matrices in order to estimate the secondary or tertiary effects of given causes.

Conclusion

These are by no means all techniques resorted to by social scientists in the last two decades. Limitation of space and the personal predilection made a selection necessary. Yet we hope to have signalled some more relevant trends that are representative of the present development. Though only a sample, they may embarras a neophyte; he may judge the way of social research too hard to follow. This would be a misperception. A combination of techniques that we might call "a simple testing design" (ex-post-facto experiment as based on good ordinal scales and on simple nonparametric tests) is not yet obsolete and offers opportunities for any social research-worker of modest means. Moreover, even the richest, the most sophisticated, techniques do not replace substantial thinking, sharp theoretical analysis and inventiveness. Yet we dare say in the light of the present survey that if a researcher comes home empty-handed from his survey-campaign or experimental work, this is probably not due to the lack of suitable techniques to tackle the intricate social relationships.

NOVĚJŠÍ SMĚRY V TECHNOLOGII SOCIÁLNÍHO VÝZKUMU

Autor konstatuje, že nikde není vývoj sociologie tak výrazný v posledních dvaceti letech jako v metodologii a technice výzkumu. Ve svém příspěvku analyzuje a komentuje novější techniky a problémy při sběru dat, jejich vyhodnocení a modelování.

---

[18] See my *A Dutch Community*, Leiden, 1956; for a systematic treatment: F. Harrary, R. Z. Norman, D. Cartwright: *Structural Models: An Introduction to the Theory of Directed Graphs*, John Wiley & Sons, 1965.

[19] For the analysis of spurious correlations by means of the "vanishing partials" see Herbert A. Simon: *Models of Man. Social and Rational*, John Wiley & Sons, 1957. Path-analysis has been described by Sewal Wright, "The method of path coefficients" in *Annals of Mathematical Statistics, V*, 1934, pp. 161—215. Sociologists "rediscovered" it only recently: Otis D. Duncan, "Path analysis: sociological examples", *Amer. Journal of Sociology, 72*, no. 1, 1966, pp. 1—16.