

Research as a Cognitive Process:  
Some Thoughts on Data Analysis

LOTIE BAILYN

July, 1975

WP 801-75

Research as a Cognitive Process:

Some Thoughts on Data Analysis\*

My aim, in this paper, is to consider research as a cognitive process, in distinction to research as a validating process, and to draw some implications of this view for the analysis of data. When research is viewed as validation, data<sup>1</sup> is collected in order to test a preformulated conception. But data also serves a cognitive function: it enables the researcher to think about the problem under study as the study progresses, and thus becomes involved in the conceptualizing process itself. It is this function of data that I wish to examine.<sup>2</sup> In particular, I want to discuss the strategies of data analysis that enhance research as a cognitive process.

No researcher, of course, would deny this double role of data.<sup>3</sup> But because our "espoused theory" (Argyris and Schön, 1974) continues to view research primarily as validating, we know more about the principles of data collection and analysis that flow from the attempt to validate than we do about the principles that guide the optimal use of data in conceptualization. One reason, perhaps, why we do not know as much about these strategies is that researchers do not often talk about how they arrive at their conclusions.<sup>4</sup> It is a very personal process, and publication norms generally preclude the presentation of results in the manner in which they actually evolve. Only the researcher is aware of the vagaries of this evolution, and this personal history is often forgotten as soon as the results reach print. But if we do not know how

---

\*An earlier version of this paper was presented at a Colloquium of the Organization Studies Group, Sloan School of Management, M.I.T., on May 1, 1975. I am most grateful to Michael Ginzberg, Edgar H. Schein, and John Van Maanen for their careful reading of earlier drafts of this paper and their many incisive comments.

this process actually works, it is not possible to see what it implies for the analysis of data.

For this reason, I want to start this paper by presenting one such personal history. It will, I hope, illustrate not merely how initial beliefs and commitments guide the process of research, but also how empirical findings influence and shape the evolving conceptions of the researcher. Moreover, though it necessarily is a very specific story dealing with one researcher's attempt to understand one phenomenon, we will be able to use it later to exemplify some of the analytic strategies discussed in the main part of this paper.

#### An Example of Research as a Cognitive Process

My story begins in England, in 1969, when I had an opportunity to work with Robert and Rhona Rapoport in their survey of university graduates in Britain, a study that served as the empirical part of a book, Sex, Career, and Family (Fogarty, Rapoport, and Rapoport, 1971). Their data consisted of responses to a rather lengthy questionnaire on many aspects of people's careers and family situations. My own work in those materials centered on the questionnaires of a subset of some 200 women eight years after their graduation from university and another set of responses from their husbands. In this subsample, therefore, analysis by couples was possible. Since most previous research in the area had included only second-hand information about spouses, this aspect of the data was particularly appealing.

I had previously speculated about the problems of professional women (Bailyn, 1965), but had not had an opportunity to work with empirical data that dealt with the issues of work and family. That earlier

paper nevertheless expressed the assumptions with which I approached the English data: I was in favor of married women having careers if they chose to do so, and though I realized that there were impediments in their way, I felt that restricting their desire to work would be more harmful to them and to their families than following this desire would be. It was with some alarm, therefore, that I viewed an early finding showing that couples in which the wife was working were less happy and more dissatisfied with their lives than those in which the wife stayed home.

My first reaction was to try to disprove the finding. Hoping that I had stumbled on a chance association, I used different definitions of the wives' work or career orientations and tried various indicators of satisfaction. But every attempt brought further verification of the finding. I considered a number of factors that might account for the relationship. Income was an obvious candidate: if families with less income were likely to have working wives, and low income was accompanied by family strain--highly plausible and empirically corroborated assumptions --the association might disappear when income was controlled. It did not. I knew, of course, that cross-sectional data could not provide evidence of the direction of causality. It was certainly possible that wives in unhappy couples sought work as an escape from the unsatisfactory family situation, a process that was easier for me to accept than its opposite. And, indeed, I am sure that this happened in some cases: that there were working couples in which satisfaction would have been even less had the wife decided to stay home. But if this were the primary way in which the original association could come about, I would have

expected the relation between wife's work and unhappiness to be considerably less in those couples where the woman was working only for financial reasons--and it was not.

These early analytic steps resulted in a provisional, if reluctant acceptance of the finding and a search for conditions and contexts in which it does or does not apply. Such a turn, however, necessitated greater conceptual clarity. My interest centered on married women with careers--women who wanted to combine work with their family roles. I had learned enough in my initial forays into the data (made particularly easy by a highly flexible interactive computer program<sup>5</sup>) to realize that the fact of working during these early years of marriage did not validly identify such women, since some worked merely for financial reasons and others, though not working at that time, were patiently waiting for the chance to resume outside employment. It was necessary, therefore, to devise a better indicator for what I had in mind. In the end, women were classified into types of career orientation not on the basis of their actual work patterns at the time of the survey but by a combination of their attitudes toward married women's careers and the satisfactions they derived from work.

Thus I arrived at a variable that defined the career orientations of the wives in the sample in a way that seemed to reflect my concerns more accurately than had the simple distinction between working and non-working with which I began. But the step from which a new concept would eventually evolve had not yet been taken. This only occurred when the career orientations of the wives were cross-tabulated with career and family orientations of the husbands--when couple patterns were developed

involving various combinations of each partner's orientation to work and family.

These patterns were evaluated in terms of the marital satisfaction associated with them. Gauged by this measure, only one combination turned out to be "bad": that in which husbands with traditional career-orientations were paired with wives who wanted to integrate careers into their own lives. In other words, when husbands are willing to subordinate some portion of their career needs to those of family, happy marriages are as likely to occur when wives are career-oriented as when they are not. Hence it turned out that it was the husband's attitude--his accommodation of work to family, to use the term I later chose--that was crucial in understanding the original finding.

In my final analysis of the English data I concentrated on this interpretation (Bailyn, 1970). It seemed a worthwhile focus for a paper because the emphasis on the husband's orientation to his own work (not only to his family or to his wife's work) introduced a new element in the discussion of married women's careers. It arose from the "couples" aspect of the data at hand, and also fitted into the more symmetric view of the relation between work and family for both men and women that was then evolving. This paper also included a comparison of the traditional couples (wives oriented to family, husbands to career) with the integrated ones (husbands accommodative, wives integrating careers into their lives). I will not go into the details of these results, but only note that they demonstrated that this integrated pattern is not one of role reversal; on the contrary, the accommodative husbands continued to get a great deal of satisfaction from their work and the integrated wives were still very much involved with their families.

Thus an analysis that started with a concern with women's careers ended with a concept applicable to men's relation to their own work. This evolution made it possible to pursue the research with an entirely different data base, consisting only of male respondents: a survey of the career experiences of M.I.T. alumni from the 1950's. These data provided enough information about wives' work patterns to allow me to continue the investigation of the effect on their families of men's orientations to work, (now in a more typically validating phase of the research process). This analysis showed that it was the men whose own lives were integrated and balanced, not those oriented exclusively to their careers and also not those primarily involved with their families, whose wives were most likely to be involved in professional careers (Bailyn, 1973). And though the dependent family variable here (career involvement of professional wives) was different from that used in England (marital satisfaction), the general conclusion from the English data that men's orientations to their own work have important family consequences was corroborated.

The example is coming to an end, even though the story is not over. The M.I.T. data made it possible to extend the work in another direction: to raise the question of the implications of accommodation not only for men's families but also for their work. Preliminary analysis on this point indicated that in the technical occupations in which these men were involved, what is good for their families may be problematic for their work (Bailyn, 1974). New research, therefore, must concentrate on the way family needs and work requirements interact with each other in the personal and career development of men and women.<sup>6</sup> New data and another cognitive research phase are needed.

By presenting this example I hope I have been able to convey something of the flexibility in the relation between ideas and data that makes research a cognitive rather than simply a validating process. The more fluid the procedures, the more continuously responsive to surprises, nuances, and anomalies they are, the more likely will the result be a new formulation. I will come back to this example whenever possible in the subsequent pages, where I try to specify the analytic strategies that allow one to maximize the cognitive value of empirical data.

#### Implications for Data Analysis

When research is viewed as a cognitive process, analysis is continuous; it occurs in all phases, from the initial collecting to the final writing. And though, for the sake of clarity, I will subsequently break this process into three stages--collection, compilation, and presentation--the principles of analysis I want to examine apply to the process as a whole. In particular, I concentrate throughout on two general analytic principles. One concerns the cognitive complexity of the data and the other relates to the interplay between concepts and data.

For data to be maximally useful for research as a cognitive process, it must be at the proper level of complexity. If it is too simple and undifferentiated it will not provide the researcher with input capable of affecting already existent views about the phenomenon under study. On the other hand, it must not be so complex as to overwhelm the analyst. As will be obvious when we discuss the specific manifestations of this principle at the various stages of the research process, controlling the cognitive complexity of the data is a major analytic task.



Second, the process of analysis proceeds by means of a continuous interplay between concepts and data. If one envisions concepts as residing in a conceptual plane and data in an empirical one, I would venture the proposition that the more links, and the more varied the links, that can be established between the two planes, the better the research. Some links, of course, must exist or there would be no empirical research at all. "Theorizing" occurs all within the conceptual plane, and an analysis that proceeds only at that level is not an empirical investigation. At the other extreme, "empiricism" refers to manipulations that stay entirely within the empirical plane. Critics who accuse the empirical social sciences of doing nothing but proving the obvious have usually limited their vision to the empirical plane and neglected to consider the links between it and the conceptual one.

It is the continuous interplay between concepts and data that perhaps more than anything else distinguishes research as a cognitive process from research as a validating process. In the typical validating study there is only one foray from the conceptual to the empirical plane: to select the potentially validating data and assess its significance. This is obviously an important step, but it does not describe the total research effort, which needs its cognitive phases as well. And these require many more links between the two planes.

Specific strategies of data analysis may be developed from the application of these two principles. Such strategies vary according to the phases of the research, and thus it is necessary to distinguish the essential task of each of the three research stages.

The function of the collection phase is to generate the kind of data that will permit research to proceed in a cognitive fashion. During data compilation—which is perhaps the most creative phase—the collected data is transformed in such a way that it both guides and reflects the evolving conceptions of the analyst. The personal history given above was concentrated in this stage. It is here that conceptually key variables, such as the career orientations of the wives in the example given, are defined; it is here, also, that standard techniques of data reduction may find their place. The stage of data presentation is crucial because the validation of the whole process often resides less within one particular study than in its fit with the shared conceptual understanding of other researchers concerned with similar problems.<sup>7</sup> Research results must be presented in a way that allows them to be assessed in the context of an evolving consensus.

These steps are obviously not strictly sequential. First drafts, for instance, are often more valuable for compilation than for presentation, and even data collection can, in certain designs, be intertwined with data compilation. But their essential functions are distinct, and thus it is fruitful to examine them separately when considering the implications of the principles of research as a cognitive process.

#### Specific Implications at Various Stages of Analysis

Chart I summarizes, for each phase, some of the specific strategies of data analysis that flow from this view of the research process. The cells of the table can be identified by the principle of analysis they exemplify (the row they are in) and by the research stage (column) to which they mainly apply. Within each cell different strategies are identified by letter. Thus 1,2,a locates sequential development as a strategy for

CHART I\*

IMPLICATIONS FOR DATA ANALYSIS

A N A L Y T I C S T A G E S

ANALYTIC PRINCIPLES	1. Data Collection	2. Data Compilation	3. Data Presentation
1. Cognitive Complexity of Data	a. surplus data  (1,1)	a. sequential development b. non-restrictive assumptions  (1,2)	a. openness b. reduction  (1,3)
2. Interplay between Concepts & Data	a. structure of variables  (2,1)	a. theoretical generalization b. empirical test of interpretations c. quantitative and qualitative indicators  (2,2)	a. links to literature b. interweaving of results and interpretations  (2,3)

\*I am grateful to Lance K. Heiko for a comment which led to an improvement in the format of this chart.

controlling the cognitive complexity of the data (row 1) in the compilation phase (column 2). The discussion that follows will proceed through the chart cell by cell. Whenever possible I will give specific examples from the research described in the beginning of this paper, and draw on other studies only when a particular strategy was not used in that case.

1,1: Controlling Cognitive Complexity of Data in the Collection

Stage. The prescription to provide surplus data (1,1,a) is made in order to insure that the data collected will be sufficiently complex to be cognitively useful. In large scale surveys this usually comes about in any case because of the low marginal cost of including a few extra questions. And, indeed, secondary analysis, such as my use of the Rapoport data, is possible because most questionnaires contain many more items than are used in the basic analysis.

But the strategy is equally applicable to more controlled research designs. Let us take as an example a study by Fry (in preparation) which attempts to describe learning environments in terms of their "situational complexity." Fry's hypothesis is that complexity in classroom environments and individual styles of learning can be categorized on an equivalent set of dimensions, and that the optimum learning situation occurs when the individual learning style matches the character of the learning environment. To test this hypothesis it is sufficient (though by no means simple, as is obvious from Fry's work) to measure learning style and to characterize learning environments. But if the research is also to serve a cognitive function and to influence the researcher's understanding of the phenomenon he is studying, it will require other kinds of data as well. For if the hypothesis of fit is confirmed, the

researcher, though his existing conception of the learning process will be strengthened, will gain no new insights; and if not confirmed, the lack of surplus data to provide possible explanations for the negative result will, most likely, lead the researcher to conclude that measurement was at fault, and thus also not affect the original conception.

This example provides clues to two kinds of surplus data with potentially great cognitive yield: data that will help one understand the process by which an expected association comes about; and data that will allow one to test possible explanations for negative results. To continue with the Fry example: If the hypothesis is confirmed, one would then want to know why a learning environment that fits a person's learning style is more effective, for there are various possible explanations. Is it because one feels more at ease in such an environment, or feels freer to experiment? Or does it require less time and effort to adjust to and hence leaves more for actual learning? Conversely, what might account for disconfirmation of the hypothesis? Perhaps the environment is not accurately perceived; or maybe one had expected a non-congruent learning environment and had already made anticipatory adjustments which then precluded one's taking advantage of the congruence actually found. In both cases, people's reactions to their experiences, though not directly related to the main hypothesis, are obviously relevant. In particular, questions about the subjects' perceptions of the environment in which they find themselves and their evaluation of their experiences in it, might provide appropriate surplus data.

Other studies, of course, deal centrally with people's perceptions and experiences. In such cases the crucial surplus data might involve no more than standard demographic information to indicate membership in

certain social groups. Or, if the design allows for greater sophistication, the surplus data might deal directly with environmental characteristics. Every research problem, naturally, will provide its own specific guidelines. But the general principle is to try to collect data on classes of variables that are not central in the initial conceptualization of the problem, so that new interpretations (both of expected and of unexpected findings) can be formulated and subjected to empirical test.

1,2: Controlling Cognitive Complexity of Data in Compilation Stage.

Control of the cognitive complexity of data is perhaps most crucial and most difficult to achieve in the compilation phase, as anyone who has ever been faced with six inches of computer output relating everything to everything else can testify. An interactive computer program, such as I used in the analysis of the English data, makes it easier for the analyst's thinking to keep up with the flow of data. But even without an interactive program, certain strategies can help control the analysis. To be sure, they are not the same as those dictated by the economics of batch processing, but their potential cognitive yield should outweigh any extra cost or effort.

Sequential development (1,2,a) is, in my mind, the best way to ensure enough complexity in the data without producing a cessation of all thought because of overload. One must proceed by looking first at the distributions of only a few variables and relating them to each other before introducing new ones, and by testing indicators and selecting those that are conceptually the clearest while rejecting those whose usefulness has been superseded in the evolving analysis. In my work on the English data, for example, I started by looking at women's current work status. But this variable, though crucial in initiating the sequence of analysis, was

completely transformed as the analysis proceeded, and hardly entered at all into the final research report.

The problem is to know where to start and where to stop: what initiates an analytic sequence and when does the analyst view it as finished? In general, my tendency is to start by trying to define a key classificatory variable that comes as close as the data will allow to one's central concern. Once one such key variable is defined, the analysis proceeds by relating that variable to other available information. What other information is chosen at this point depends on one's original understanding of the phenomenon under study, and is guided by past research, theories, personal experiences, or the researcher's hunches. The sequence stops when a formulation has been reached that is pleasing<sup>8</sup> to the analyst and fits the empirical relations that were found.

Sooner or later, of course, all researchers are faced with the situation where there is no focal concern to initiate a sequence. This occurs most often when analytic results are required by contract (by the need to write a paper, for instance) rather than by the desire to understand a particular problem. In such cases it makes sense to start routinely --by looking at all the marginals, for instance, or relating everything to a few demographic characteristics such as sex or age--and move into a cognitive research sequence when something surprising catches one's eye.<sup>9</sup>

Obviously these are very general prescriptions; each particular case will necessarily be different, and inevitably there will be false starts. But the strategy of working slowly and sequentially--taking a few variables at a time, looking first at individual distributions, then pair-wise relations, then at more complicated interrelations--should

minimize the time and effort spent on false starts and maximize the cognitive yield of the data. To be sure, a very experienced, self-disciplined analyst could proceed sequentially through six inches of computer output. But it is not an easy task and such a procedure severely limits the available sequences.

But aside from sequentiality there are various ways of looking at data that can affect its cognitive potential. In particular, I find the assumptions of linearity and additivity, which underlie so many standard techniques of data reduction, to be very restrictive. When I speak of non-restrictive assumptions (1,2,b), therefore, as a strategy of data compilation, I primarily mean looking at data in such a way that interactions are clearly visible. The effects the social scientist studies depend on the context in which they occur (Cronbach, 1975), and our data analysis must be sensitive to this fact. The example presented above is anchored in such an interaction: wives' career patterns have a different effect when their husbands are accommodative than when they are traditionally career oriented; only one specific combination of husbands' and wives' patterns stands out as problematic. If data compilation always proceeds by averaging effects over conditions instead of looking at them separately for each condition, such results will not emerge.

1,3: Controlling Complexity of Data in Presentation Stage. When research results are written up, the appropriate complexity of the data is geared more to the reader than to the researcher--though the discipline imposed on the analyst at this stage often clarifies his or her own understanding of the phenomenon under study. The prime task of researchers in this phase is to present data in such a way that their procedures are



open to the scrutiny of others. Such scrutiny is crucial when research is pursued in the manner presented here, since alternative explanations are often more obvious to the interested reader than to the original analyst. Hence the prescription of openness (1,3,a).

Perhaps the most closed way to present one's data is merely to indicate which results are statistically significant and which are not. This mode of presentation could allow two relationships which actually are very similar to be viewed in opposing ways, simply because their P-values happened to fall on either side of a predetermined value. And the unwary reader would have no way of knowing how close to the line each one in fact was. Further, presenting only the significance of results often hides their strength. A correlation of .66, for instance, may be insignificant in a small sample. But the reader should be allowed to decide that the design was a fault rather than being forced into the conclusion that there actually is no association between the variables involved. And this is only possible when the correlation is given in the final report, even if it has to appear without a \* (e.g. Hall, 1963).

But correlations themselves, though more open than P-values, are still relatively closed. A paper, for instance, that reports that the correlation between maternal employment and independence in daughters is positive, while that between maternal employment and independence in sons is negative, though intriguing, leaves many questions unanswered (cf. Maccoby and Jacklin, 1974, pp.3-8). Suppose that independence is measured on a ten-point scale, then the following two hypothetical tables--both of which fit the reported correlations--would lead one to very different conclusions:

Mean Independence Scores:	Mother	
	Employed	Not Employed
Sons	7.0	9.0
A) Daughters	6.0	4.0
<hr/>		
Sons	4.0	6.0
B) Daughters	6.0	4.0

Table A indicates that though independence is more characteristic of boys than of girls under all conditions, maternal employment decreases this sex differential. The data in Table B, in contrast, show no such androgynous effect of maternal employment. They imply, rather, that a working mother creates more independence in her daughter than in her son, whereas a mother who is primarily at home fosters more independence in her son.

The most open kind of presentation would give the actual distributions of the variables under consideration. But such a format would be overwhelming to the reader and thus a focus must be introduced. It is here that the strategy of reduction (1,3,b) comes in. Its purpose is to reduce the complexity that would result from a full presentation by focusing on one particular aspect of a distribution. This might be on its central point, for example, or on the weighting at one or both of its extremes--depending, of course, on the particular conclusions the presented data support. Such reduction is always necessary to a certain extent, but is most crucial when more than one variable is involved.

Table 1 (adapted from Bailyn and Schein, in preparation) is a relatively open, yet reduced table. All three of the variables represented in it are given in reduced form. The stub variable present position

Table 1

Engineers' Alienation from Work by Present  
Position and People-Orientatation

<u>Present Position of Initial Staff Engineers:</u>	<u>Value Placed on Working with People</u>	
	HIGH	LOW
Staff Engineer	61% (N=51)	19% (N=58)
Engineering Manager	31% (N=45)	18% (N=39)
General Manager	17% (N=64)	7% (N=15)

Note. The figures in the table represent the percentage in each cell who are alienated from work: the % who fall at the Low end of an Index of Work Involvement.

(which defines the rows of the table) focuses on only those respondents whose first jobs were in engineering staff positions and further eliminates those of this group whose present jobs do not fall into the three listed categories. The reduction in the column variable, people-orientation, is not obvious from this table alone. But readers of the full report would know that the focus here is on the two extremes of the distribution, with a large group of respondents, whose orientation is neither primarily to people nor to things, not represented at all. Finally, the presentation of the cell variable of work involvement focuses entirely on the alienated end of the distribution.

These reductions make it possible to see the main point of the table clearly. It shows that present position and people-orientation interact with each other in their effect on work involvement: alienation from work is primarily the response of those engineers who very much value working with people but who have not moved into managerial positions.

But the table is also sufficiently open and complex to present a fair amount of additional information to an interested reader. It shows, for instance, that there are main effects as well as interaction effects: respondents who greatly value working with people are more alienated in technically based careers than are those whose orientation is more to things; so, too, are those who currently work in staff positions, particularly when compared to those whose work involves general managerial duties. Even further removed from the main point of the table, because it ignores the focal variable of work involvement, is the information implicit in the cell frequencies. These show that engineers who value working with people are much more likely to have moved into management than are those whose values are not centered on people.

To summarize, the presentation of data must be sufficiently focused so that the reader can easily assimilate the central research results. At the same time, it must be open enough to allow interested readers to make their own inferences and thus to evaluate the cognitive process the analyst went through to reach the presented conclusions.

Now, as we move to the strategies that flow from the second analytic principle--the interplay between concepts and data--we see a shift in emphasis. Where the first set was designed primarily to maximize the potential cognitive yield of data, this set serves also the function of controlling and validating the process of analysis.

2,1: Interplay between Concepts and Data in Collection Stage. In one sense, the interplay between concepts and data in the stage of data collection is obvious. It involves the problems of measurement and defining operations that are dealt with in every textbook. Though these aspects are clearly important, I will not deal with them here. Rather, I will concentrate on a strategy that is crucial to research viewed in the manner of this paper--defining the structure of the variables (2,1,a).

By structuring the variables I mean defining conceptually distinct categories of information and specifying the particular variables that fit into each. The tripartite partition of variables into those measuring outcomes (dependent variables), determinants (independent variables), and conditions (intervening variables), represents an often used scheme.<sup>10</sup> Allocating specific variables into such general categories is necessary to guide the research process in all phases: even as simple a task as deciding which way to run percentages depends on some notion of structure among the variables involved. Nonetheless, it is a strategy that is

peculiarly relevant to the collection phase because a clear structure to guide that stage greatly eases all analytic tasks. Indeed, one of the reasons that secondary analysis is difficult is that such a structure must be evolved at the same time that the data is being compiled.

Generally, the strategy of structuring the variables guides the collection phase by helping the analyst anticipate the kinds of data that must be collected and pointing out areas where potentially important surplus data may be found. More specifically, such a structure can help researchers in the actual development of their instruments. The following example on the individual level shows how even a very general initial structuring can help one design a questionnaire.

Suppose one wants to distinguish among behavior, attitudes, and personality dispositions, but is dependent for all one's measurement on a written questionnaire. If, as is most likely, one associates behavior with outcome and personality dispositions with determinants, one can be guided by this structure in wording one's questions. Thus one could introduce questions that deal with anticipated reactions to hypothetical situations to measure outcome (see, e.g., Evans, 1974); and one could word questions in a manner designed to negate, as much as possible, the effect of specific situations when trying to tap personality dispositions. In some sense, of course, all questionnaire responses are on the same level--representative, perhaps, only of respondents' attitudes or the attitudes they feel are called forth by the questions. But, by being aware of the structure of one's variables from the start, and letting it guide the development of one's instrument, the usefulness and probable validity of such an approach are maximized.

2,2: Interplay between Concepts and Data in Compilation Stage. In the compilation phase, the interplay between concepts and data is particularly important. Without repeated links between the conceptual and empirical planes at this stage there would be no way to control the validity of the results emerging from the techniques designed specifically to maximize the cognitive potential of the data. All the strategies in this cell fall under the general heading of multiple corroboration--an analytic form of triangulation (Webb et al., 1966) applicable even within a single, if sufficiently complex, measurement instrument.

Theoretical generalization (2,2,a, see Rosenberg, 1968) refers to the use of alternate indicators for a concept to test the generalizability of a particular result. In the example given above I used different indices to represent accommodation in the English and in the M.I.T. samples. Far from undermining the value of such a replication, I feel it strengthens the case. Within the English study itself, my initial attempts to "disprove" the negative relation between women's work and family satisfaction by switching indicators, inadvertently strengthened the original finding, by showing that the association was not dependent on chance covariation of the particular measures I had first used.

It is interesting to speculate on the relevance of this strategy for the repeated use of the same instrument. Would students of leadership, for example, have more confidence in Fiedler's theories (1967, 1971) if he did not always rely on the LPC (Least Preferred Coworker) to measure leadership style? But such speculation goes beyond the scope of what is meant by theoretical generalization in the present context, where it represents a way of linking concepts with alternative indicators in order to test the validity of a particular finding.

The empirical test of an interpretation (2,2,b) is the essential feature of what Rosenberg (1968) has called the "logic of survey analysis"-- though in my mind it is applicable to the analysis of any kind of data-- and is one of the reasons that it is important to have surplus data. It is a crucial analytic step because it provides a way to control and validate the process of "data dredging" (Selvin and Stuart, 1966). It does this by prescribing a procedure that guards against the extremes of staying too much in either the conceptual or the empirical plane. First, by making one think about one's findings and generate explanations for them, it forces researchers to link their empirical results to the conceptual plane. And then it requires analysts to corroborate these post hoc interpretations by deducing testable consequences from them and actually checking these with the data at hand.

The use of income to test the original finding in the example given above illustrates the strategy. The interpretation to be tested was that the association between working wives and unhappiness could be explained by differences in family income. Notice that showing that income relates to both women's work and family strain--which could have been established by argument or by others' research results--is not a sufficient test of the interpretation. The existence of such relations only makes it possible for income to account for the initial finding. Empirical corroboration, however, implies that it actually does so; and this cannot be known unless the association is looked at when income is controlled. In the analysis of the English data, it will be recalled, the necessary first order relations did hold but the original association did not disappear when income was held constant. Hence the hope that income might account for the initial finding had to be relinquished.



The final strategy in this cell, the use of quantitative and qualitative indicators (2,2,c), refers to a principle of corroboration based on a notion of opposites: if you have qualitative data, quantify wherever possible; if your data base is quantitative, don't ignore its qualitative potential. The former is relatively obvious. It deplores the use of unspecific quantifiers--e.g. "few"; "many"; "in almost every case"--in the analysis of qualitative data. It is based on the assumption that by forcing the analyst to seek indicators that can actually be counted and to present results in "quantitative" terms, the reliability of qualitative analysis is increased.

The opposite, the use of qualitative indicators with quantitative data, is perhaps more surprising. The principle is the same: switching the character of indicators is not possible without linking the empirical and conceptual planes, and such links serve as corroboration of findings. Though the application in this case is less obvious, once the principle of using qualitative indicators in the analysis of quantitative data is accepted, the possibilities are limited only by the creative imagination of the researcher.

Let us assume, for example, that our quantitative data base consists of responses to a completely structured questionnaire. There is no qualitative data, and yet certain qualitative strategies are available to the analyst which can increase the cognitive potential of the data as well as serve a corroborating function. Here only two will be mentioned. The first refers to an emphasis on content in addition to frequency. Thus one might look at individual items that comprise scales, and select those, for instance, that most differentiate two groups. Consideration of the actual wording of these key items might provide an

understanding that would not have emerged if one had limited oneself to total scale scores. In the context of a structured questionnaire, the content of such individual items represents "qualitative" data.

The second possible qualitative strategy refers to a holistic emphasis (cf. Weiss, 1966): analysis by people as well as by variables. With structured questionnaire data from large samples this could take a number of forms. For instance, one could explicate or validate an index by looking at the questionnaires of the people who fall at its extremes (see, e.g., Bailyn and Schein, in preparation, Chapter 5). Or, one could try to understand the meaning of an empirical association by analyzing the people whose responses do not show the expected correlation (cf. Kendall and Wolf, 1949). In either case, the whole questionnaire of a person represents a richer and more qualitative datum than do the response frequencies to a particular set of questions, and thus provides a potential for corroboration not often utilized in questionnaire studies.

It should be noted, parenthetically, that these strategies have tactical implications for the computer processing of questionnaire data. They require, for example, that identification numbers be entered as retrievable information. Also, they point to the importance of not deleting responses to individual items once scale scores have been derived.

2,3: Interplay between Concepts and Data in Presentation Stage. In the last cell of the chart, the interplay between concepts and data serves a communication function: it allows readers to follow the process by which conclusions have been derived from the data, and permits them to judge these conclusions in a larger context. Links to the literature (2,3,a) are often very useful at this stage, and have their place in

the "results" section of research reports as much as in the introduction.

The final strategy that flows from the view of research presented here is sometimes viewed with suspicion. To interweave results and interpretations (2,3,b) in a research report goes against most publication norms, where "results" and "conclusions" are neatly separated by section headings. It should be obvious by now that such a presentation follows the principles of validating research, where a particular formulation is validated by a single empirical datum. Such a datum is then, correctly, presented in a "results" section which resides between two versions of the conceptual formulation: the unvalidated one (presented as "hypotheses") and the validated one given in the "conclusions" section. But to force results of research undertaken in the cognitive vein into such a format does disservice in my opinion, to the process involved as well as to the reader. It is far better to admit what one is doing and lay the process open to inspection by others, thus maximizing the possibility of proceeding from a cognitive phase of research to a validating one. And this purpose is better served by documenting the way results were interpreted and how these interpretations were tested empirically, a documentation that cannot follow the traditional format.

These then are some of the strategies that flow from viewing research as a cognitive process. They would, I believe, if properly used, enrich the analysis of empirical data.

\* \* \* \* \*

I would like to conclude with an anecdote. One night I came into the kitchen to find my husband opening drawer after drawer and muttering to himself that he didn't know what he was looking for. Since he had an empty bowl in his hand, and a container of ice cream was on the table, it seemed to me obvious that he was looking for the ice cream scoop which, as a matter of fact, was lying on an adjacent counter. But having momentarily forgotten the ice cream, he continued to open drawers in a seemingly random fashion. This is the obvious part of the process: you can't find something unless you know what you are looking for. But there is another, equally important part, which relates to the ability to perceive the relevance of something you find for understanding the problem at hand. It occurred to me, as I watched my husband, that if by accident a bright red plastic spoon had caught his attention, it might have reminded him--perhaps by his reflecting on the functions of spoons--of what he was lacking. Unfortunately, there was nothing in those kitchen drawers that was arresting in this sense, and so the search continued.

It is the research strategies that aid in identifying and utilizing such arresting findings--findings that become fruitful, heuristic, generative as part of a continuous process of conceptualization--that have been the subject of this paper.

Notes

<sup>1</sup>I am using the word "data" in two senses: first, as a singular collective noun to refer to the totality of available empirical information; and second, in its more usual sense, to refer to specific results from a study, in which case it takes the plural form.

<sup>2</sup>Such a view is obviously applicable to all kinds of data. In the social sciences it has been most explicitly recognized by advocates of qualitative research (e.g. Glaser and Strauss, 1967; Becker, 1958). But Polya's (1954) distinction between demonstrative and plausible reasoning, which applies to mathematics, is similar; and just recently an appeal for the collection of data for more than validation--for "close observation of effects the hypothesis under test says nothing about" (p. 122)--has come from within "scientific psychology" itself (Cronbach, 1975).

<sup>3</sup>The fact that empirical research and social theory are mutually dependent on each other was already documented by Merton (1949) almost thirty years ago.

<sup>4</sup>A classic exception is Hammond (1964). More recently, contributors to the Festschrift to Herman Wold (Dalenius, Karlsson, and Malmquist, 1970), were asked to "report your personal experience of the research process and how new knowledge is created in the process" (p. 18). In the present context the chapter by Johan Galtung, "On the Structure of Creativity" (1970, pp, 43-61), is particularly relevant.

<sup>5</sup>The Conversational Mode Survey Analysis Program developed by L. Hawkins of Survey Analysis Ltd., London.

<sup>6</sup>Members of the Organization Studies Group at M.I.T., particularly Edgar H. Schein, John Van Maanen, and myself, are currently pursuing some of the theoretical implications of such an interactive view of careers.

<sup>7</sup>This view, though not that expressed in standard texts on research methods (see, e.g., Kerlinger, 1973), is gaining currency in discussions of the sociology of science (e.g. Mitroff and Mason, 1974; Kuhn, 1970).

<sup>8</sup>Sometimes what is pleasing will be empirical corroboration of an intuitively or theoretically "obvious" point; at other times one is most pleased by a conceptually clear interpretation of an unexpected or contradictory finding. The criteria of closure actually used in any given case will depend, of course, on the particular problem and researcher involved.

<sup>9</sup>The concept of "relative deprivation," for instance, seems to have evolved in this way (Stouffer et al., 1949).

<sup>10</sup>The technique of path analysis forces researchers to provide a very specific structure of their variables, one which also includes the causal links among the variables. Path analysis, therefore, would seem to be uniquely suited to effect this strategy. It does not, however, fit some of the other specifications described in this paper—particularly not those relating to cognitive complexity of data during compilation (1,2). It probably is more useful, therefore, in a validating than in a cognitive research phase. Further, it is not always possible to predetermine causal links, though if data collection is guided by a clear structure of the variables involved in the research, the likelihood of accurately specifying these links is greatly increased.

References

- Argyris, C., & Schön, D.A. Theory in practice: Increasing professional effectiveness. San Francisco: Jossey-Bass, 1974.
- Bailyn, L. Notes on the role of choice in the psychology of professional women. In R.J. Lifton (Ed.), The woman in America. Boston: Houghton Mifflin, 1965.
- Bailyn, L. Career and family orientations of husbands and wives in relation to marital happiness. Human Relations, 1970, 23, 97-113.
- Bailyn, L. Family constraints on women's work. Annals of the New York Academy of Sciences, 1973, 208, 82-90.
- Bailyn, L. Accommodation as a career strategy: Implications for the realm of work. (Working Paper No. 715-74). Cambridge, Mass.: M.I.T., Alfred P. Sloan School of Management, July 1974.
- Bailyn, L., & Schein, E.H. Work involvement in technically based careers: A study of M.I.T. alumni at mid-career. Book in preparation.
- Becker, H.S. Problems of inference and proof in participant observation. American Sociological Review, 1958, 23, 652-660.
- Cronbach, L.J. Beyond the two disciplines of scientific psychology. American Psychologist, 1975, 30, 116-127.
- Dalenius, T., Karlsson, G., & Malmquist, S. (Eds.). Scientists at work: Festschrift in honour of Herman Wold. Stockholm: Almqvist & Wiksell, 1970.
- Evans, P.A.L. The price of success: Accommodation to conflicting needs in managerial careers. Unpublished doctoral dissertation, M.I.T., 1974.
- Fiedler, F.E. A theory of leadership effectiveness. New York: McGraw-Hill, 1967.
- Fiedler, F.E. Leadership. Morristown, N.J.: General Learning Press, 1971.
- Fogarty, M.P., Rapoport, R., & Rapoport, R.N. Sex, career, and family. London: Allen & Unwin, 1971.
- Fry, R. Diagnosing learning environments: An observational scheme for matching learner styles with levels of situational complexity. Doctoral dissertation in preparation, M.I.T.
- Galtung, J. On the structure of creativity. In T. Dalenius, G. Karlsson, & S. Malmquist (Eds.), Scientists at work: Festschrift in honour of Herman Wold. Stockholm: Almqvist & Wiksell, 1970.
- Glaser, B.G., & Strauss, A. The discovery of grounded theory. Chicago: Aldine, 1967.

- Hall, R.H. The concept of bureaucracy: An empirical assessment. American Journal of Sociology, 1963, 69, 32-40.
- Hammond, P.E. (Ed.). Sociologists at work: Essays on the craft of social research. New York: Basic Books, 1964.
- Kendall, P.L., & Wolf, K.M. The analysis of deviant cases in communications research. In P.F. Lazarsfeld & F.N. Stanton (Eds.), Communications research 1948-1949. New York: Harper, 1949.
- Kerlinger, F.N. Foundations of behavioral research (2nd ed.). New York: Holt, Rhinehart, & Winston, 1973.
- Kuhn, T.S. Postscript-1969. In The structure of scientific revolutions (2nd, enlarged ed.). Chicago: University of Chicago Press, 1970.
- Maccoby, E.E., & Jacklin, C.N. The nature of the evidence. In The psychology of sex differences. Stanford: Stanford University Press, 1974.
- Merton, R.K. The bearing of empirical research on sociological theory. In Social theory and social structure: Toward the codification of theory and research. Glencoe, Ill.: Free Press, 1949.
- Mitroff, I.I., & Mason, R.O. On evaluating the scientific contribution of the Apollo moon mission via information theory: A study of the scientist-scientist relationship. Management Science, 1974, 20, 1501-1513.
- Polya, G. Induction and analogy in mathematics. Vol. 1 of Mathematics and plausible reasoning. Princeton: Princeton University Press, 1954.
- Rosenberg, M. The logic of survey analysis. New York: Basic Books, 1968.
- Selvin, H.C., & Stuart, A. Data-dredging procedures in survey analysis. American Statistician, 1966, 20 (3), 20-23.
- Stouffer, S.A., Suchman, E.A., DeVinney, L.C., Star, S.A., & Williams, R.M. The American soldier: Adjustment during army life. Vol. 1 of Studies in social psychology in World War II. Princeton: Princeton University Press, 1949.
- Webb, E.J., Campbell, D.T., Schwartz, R.D., & Sechrest, L. Unobtrusive measures: Nonreactive research in the social sciences. Chicago: Rand McNally, 1966.
- Weiss, R.S. Alternative approaches in the study of complex situations. Human Organization, 1966, 25, 198-206.