1 Introduction

PART 1

Some assumptions

A set of assumptions lies behind this approach to qualitative analysis, which first will be listed and then briefly discussed.

1. Very diverse materials (interviews, transcripts of meetings, court proceedings; field observations; other documents, like diaries and letters; questionnaire answers; census statistics; etc.) provide indispensable data for social research.
2. As compared with both the quantitative analysis of data and the actual collection of data by qualitative analysts, the methods for qualitatively *analyzing* materials are rudimentary. They need to be developed and transmitted widely and explicitly throughout the social science community.
3. There is need for effective theory – at various levels of generality – based on the qualitative analysis of data.
4. Without grounding in data, that theory will be speculative, hence ineffective.
5. Social phenomena are complex: Thus, they require complex grounded theory. This means conceptually dense theory that accounts for a great deal of variation in the phenomena studied.
6. While there can be no hard and fast rules governing qualitative analysis – given the diversity of social settings, research projects, individual research styles, and unexpected contingencies that affect the research – it is possible to lay out general guidelines and rules of thumb to effective analysis.
7. Such guidelines can be useful to researchers across a broad spectrum of disciplines (sociology, anthropology, political science, psychology, public health, nursing, and education) and, regardless of "tradition" or "theoretical approach," just as long as they believe their work can be furthered by the qualitative examination of materials. Also, such analytic methods can be useful whether researchers are wedded to the idea of social science per se or to more humanistic versions of social research ("understanding," "enlightenment").
8. Finally, research is basically work – sets of tasks, both physical and conceptual – carried out by researchers. Development, use, and teaching of qualitative analysis can be enhanced by thinking specifically of analysis in terms of the

organization and conduct of that work. Thus, what we know about work (from research on that phenomenon) can be applied to the improvement of research methods.

Materials as data

Among social scientists a distinction is commonly drawn between quantitative and qualitative research. The distinction in part has its origins in the history of some disciplines, especially perhaps sociology and social anthropology – in sociology, because so many disciplinary trends since World War II have fostered questionnaires and other survey methods of collecting data and their statistical treatment; and in anthropology, because qualitative analysis of field data is the primary mode, although quantitative methods have lately been more employed, to the distress of many who steadfastly rely on qualitative methods. "Qualitative methods" has generally been used, also, to refer to the work of researchers who work as differently as ethnographers, clinical and organizational psychologists, grounded-theorist sociologists, or macrohistorians/sociologists. *Qualitative* researchers tend to lay considerable emphasis on situational and often structural contexts, in contrast to many *quantitative* researchers, whose work is multivariate but often weak on context. Qualitative researchers tend, however, to be weak on cross-comparisons because they often study only single situations, organizations, and institutions. (See, however, recent discussions and methods pertinent to cross-site qualitative analysis: Miles and Huberman 1983, pp. 151–209; Miles, p. 1284; and see others who are inventing and testing procedures for merging quantitative and qualitative analysis: Louis 1982; Smith and Robbins 1982; Jick 1983; Sieber 1983; McClintock et al. 1983.)

Quite aside from historical considerations, it is our contention that the genuinely useful distinction (which we will touch on further) is in how data are treated *analytically*. (There is neither logical nor any sensible reason for opposing these two general modes of analysis. I do not discuss in this book their use in conjunction with each other because I have had no recent research or teaching experience in combining the two.) In quantitative research, statistics or some other form of mathematical operations are utilized in analyzing data. In qualitative research, mathematical techniques are eschewed or are of minimal use, although assuredly rudimentary or implicit counting and measuring are usually

involved (How many? How often? To what degree?). Qualitative analysis may utilize a variety of specialized nonmathematical techniques, as noted below, or as commonly practiced may use procedures not appreciably different from the pragmatic analytic operations used by everybody in thinking about everyday problems. (Leonard Schatzman terms these *natural analysis*. See Schatzman, forthcoming.) Qualitative researchers, however, when addressing scientific rather than practical or personal problems, are more self-conscious and more "scientifically rigorous" in their use of these common modes of thinking.

In any event, moving to the research materials themselves: They occur in a variety of forms, all of which have been utilized by social scientists – as well as by investigators in fields like history, psychology, education, and law – although different disciplines and their specialties have favored one type of material rather than another. For instance, among those primarily utilizing qualitative methods, ethnographers have relied mainly for data on field observations converted into field notes and on interviews. Historians may interview if their work is on contemporary or relatively recent events, but principally they utilize many different kinds of documents, depending on their specific research aims and on the availability and accessibility of materials: records of various types, memoirs, official and personal letters, diaries, newspapers, maps, photographs, and paintings. Researchers in clinical psychology base conclusions primarily on their clinical observations of patients' nonverbal as well as verbal behavior, and on therapeutic interviews. Many sociologists prefer to analyze written texts rather than engage in field research or interviewing; others generate materials through tape recordings of conversations, transcripts of court trials, and the like. While some materials (data) may be generated by the researcher – as through interviews, field observations, or videotapes – a great deal of it already exists, either in the public domain or in private hands, and can be used by an informed researcher provided that he or she can locate and gain access to the material – or is lucky enough to stumble on it.

These materials, then, are useful for qualitative analysts in all of the social sciences. In some disciplines or their specialties, materials are converted into quantitative data through counting and measuring operations. In others, counting and quantitative measurement are minimal and these operations may even be rejected on reasonable, well-thought-out grounds. Whether qualitative or quantitative analysis predominates is sometimes a matter of ideology (which can be frozen into

tradition), but more often is a matter of rational choice. At any rate, qualitative analyses are more than merely useful: They are often indispensable.

Of course in daily life everyone engages in some form of qualitative analysis – much as Moliere's citizen used prose – without thinking twice about the matter since no judgments, no decisions, no actions can be taken in their absence. So, in a genuine sense, both common sense and "researcher" conclusions are based on "qualitative data." Without denigrating the care, self-awareness, and systematic character of a large proportion of everyday, pragmatic analyses (indeed, researchers themselves would be irate if accused of lacking those virtues in their daily thinking), it is clear enough that researchers are expected by their colleagues to adhere to disciplinary practices associated with the "good researcher," and will criticize or ignore as incompetently done any research products judged deficient in careful, scrupulous, systematic treatment of reliable data.

More important for our purposes here is that improved qualitative analysis requires more explicitly formulated, reliable, and valid methods than currently exist. *Analysis* is synonymous with *interpretation* of data. It refers to research activity which, as will be detailed later, involves several different but related elements (or operations). (See Miles and Huberman 1983, p. 214, for slightly different emphases.) Qualitative analysis occurs at various levels of explicitness, abstraction, and systematization. At the beginning of a research project, when the researcher reads a sentence or sees an action, the analysis may be quite implicit; but analysis it surely is insofar as perception is selective, mediated by language and experience. Later in the investigation or even during the first days when an observed scene, interview, or perused document challenges the researcher's analytic sense, the conclusions will be drawn more explicitly and probably more systematically. Depending on the purposes of the investigator, the final conclusions drawn in the course of the research can vary greatly by level of abstraction. At the lowest levels they can be "descriptive," and at the highest levels, the researcher may aim for the most general of theory. But description itself can be "low level" – perhaps only reproducing the informants' own words or recording their actions – or can be reported at a much more complex, systematic, and interpretative level. If social theory is aimed for, it can be formulated with more or less systematic treatment and with varying degrees of abstraction. In addition, the theory at any level can be broader or narrower in scope; and it may be linked with other theory which is more or less developed.

Methods for qualitative analysis of data

Social scientists who engage entirely or primarily in qualitative analysis generally would agree that quantitative methodology is much more explicitly presented in standard manuals and during training. As we noted some years ago in *The Discovery of Grounded Theory* (1967), quantitative analysts since the 1920s have developed relatively rigorous methods for collecting and treating their data, and have written extensively about those methods. By contrast, much of the attention of qualitative researchers is still focused on improving and making explicit their techniques for the collection of data – analytic considerations being at best quite secondary and, such as they are, transmitted on an apprenticeship basis in tacit rather than explicit fashion. However, a number of researchers have developed effective methods for the qualitative analysis of different types of materials. The character of some of these methods is suggested by their respective names: conversational analysis, (qualitative) network analysis, biographical analysis, sociolinguistic analysis, dramaturgical or social drama analysis, textual analysis. These methods, or sets of techniques, have evolved in conjunction with particular lines of research and theoretical interests or commitments.

Grounded theory

The methodological thrust of the grounded theory approach to qualitative data is toward the development of theory, without any particular commitment to specific kinds of data, lines of research, or theoretical interests. So, it is not really a specific method or technique. Rather, it is a style of doing qualitative analysis that includes a number of distinct features, such as theoretical sampling, and certain methodological guidelines, such as the making of constant comparisons and the use of a coding paradigm, to ensure conceptual development and density.

This approach to qualitative analysis was developed by Glaser and Strauss in the early 1960s during a field observational study of hospital staffs' handling of dying patients (1965, 1968). Contributing to its development were two streams of work and thought: first, the general thrust of American Pragmatism (especially the writings of John Dewey, but also those of George H. Mead and Charles Peirce) and including its emphases on action and the problematic situation, and the necessity for conceiving of method in the context of problem solving; second,

the tradition in Chicago Sociology at the University of Chicago from the 1920s through the mid-1950s, which extensively utilized field observations and intensive interviews as data-collecting techniques, and furthered much research on the sociology of work. Both the philosophical and the sociological traditions assumed that change is a constant feature of social life but that its specific directions need to be accounted for; they also placed social interaction and social processes at the center of their attention. In addition, Chicago Sociology almost from its inception emphasized the necessity for grasping the actors' viewpoints for understanding interaction, process, and social change. The study of dying by Glaser and Strauss, with its initial use of the grounded theory style of analysis, drew from both of those philosophical and sociological traditions. (For a fuller historical understanding of the background of grounded theory, it would be useful to read John Dewey's *Logic: The Theory of Inquiry*, 1937, and Everett C. Hughes's papers on occupations and work and on fieldwork in *The Sociological Eye*, 1971.[1])

Of course, theory is generated and tested even by researchers whose analytic methods remain relatively implicit, but the grounded theory style of analysis is based on the premise that theory at various levels of generality is indispensable for deeper knowledge of social phenomena (Glaser and Strauss 1967; Glaser 1978). We also argued that such theory ought to be developed in intimate relationship with data, with researchers fully aware of themselves as instruments for developing that grounded theory. This is true whether they generate the data themselves or ground their theoretical work in data collected by others. When we advocated that position in 1967 there was perhaps more need to remind social scientists of that necessity for grounding their theory than now.

Complex theory

One of our deepest convictions is that social phenomena are complex phenomena. Much social research seems to be based on quite the opposite assumption; either that, or researchers working in various research traditions describe or analyze the phenomena they study in relatively uncomplex terms, having given up on the possibility of ordering the "buzzing, blooming confusion" of experience except by

[1] Barney Glaser had studied with Paul Lazarsfeld at Columbia University, and so brought to the development of the grounded theory approach some of Lazarsfeld's emphasis on multivariate analysis. The Chicago tradition similarly emphasizes variation.

ignoring "for a time" its complexity. Their assumption apparently is that later generations will build on current endeavors – a kind of accumulation premise that seems reasonable, since one cannot study everything at the same time. Nevertheless much more complexity can be handled than is often done by quite competent or even gifted researchers. This is why grounded theory methodology emphasizes the need for developing many concepts and their linkages in order to capture a great deal of the variation that characterizes the central phenomena studied during any particular research project. We shall have much to say about this issue of complexity throughout this book.

Guidelines and rules of thumb, not rules

Affected by a mistaken imagery (based on speculative philosophy) of effective scientific research – exact, precise, explicit about its technology – students of social life often assume that is should be possible to lay down rules (later if not right now) for carrying out social investigations. We do not believe this is an accurate characterization of how any kind of work is carried out; and it is not likely ever to be true for researchers who aspire to developing new theory or to extending extant theory. Even in the more precise scientific investigations of physicists or chemists, contingency is inevitable; thus, discretion is advisable and often essential. Moreover, the best opinion among philosophers these days holds that such codification of investigation is impossible anyhow.

We shall not argue the point further except to repeat that several structural conditions mitigate against a neat codification of methodological rules for social research. These include the diversity of social settings and their attendant contingencies which affect not merely the collection of data but how they are to be, and can be, analyzed – quite aside from researchers' often different aims in doing their analyses. Researchers also have quite different investigatory styles, let alone different talents and gifts, so that a standardization of methods (swallowed whole, taken seriously) would only constrain and even stifle social researchers' best efforts.

Hence we take the stand about our own suggested methods that they are by no means to be regarded as hard and fixed rules for converting data into effective theory. They constitute guidelines that should help most researchers in their enterprises. For that – as we shall attempt to show – researchers need to be alive not only to the constraints and challenges of research settings and research aims, but to the nature of

...eir data. They must also be alert to the tempora. aspects or phasing of their researches, the open-ended character of the "best research" in any discipline, the immense significance of their own experiences as researchers, and the local contexts in which the researches are conducted.

Our guidelines for developing theory are not merely a kind of laundry list of suggestions, however: they are stronger than that, for they emphasize that certain operations must be carried out. Coding must be done, and generally done early and continually. Analytic memos must be done early and continually in conjunction with the coding. And a few concepts, loosely strung together, cannot satisfy the requirements for formulating social theory. Yet, we emphasize also that personal pacing and experiences can be ignored only to the detriment of effective and analytic work. We do not believe that strict instructions can be given for how to proceed in detail with all kinds of materials, by everyone, holding for all kinds of research, at all phases of the research project. Methods, too, are developed and change in response to changing work contexts. However, we have throughout this book included lists of rules of thumb. These are to be thought of as operational aids, of proven usefulness in our research. *Study* them, *use* them, but *modify* them in accordance with the requirements of your own research. Methods, after all, are developed and changed in response to changing work contexts.

Our guidelines and rules of thumb, then, will be useful to any researcher who shares our concern for achieving better comprehension of social phenomena – through the development of some level of theory – regardless of the substantive character of the materials or of the particular discipline in which he or she has been trained. We believe that the same assertion holds for researchers who are committed to different traditions or theoretical approaches, even within the same discipline; this, provided these traditions and approaches cash in on their strengths – raising important problems or looking at relevant or neglected areas of social life – rather than box their adherents into dogmatic positions which foreclose on the possibility of actually challenging some of what their own traditions currently stand for.

Underlying some contemporary positions are the contrasting assumptions that either a social science is possible or that it is to be eschewed in favor of more humanistic versions of knowledge about human activity. Our own position is somewhere between these extremes, though some practitioners of grounded theory methodology might lean in either direction on that continuum of belief. Nevertheless, we believe that the methodological guidelines and general procedures can be of service to researchers regardless of where they stand on this particularly divisive and long-standing dispute among social scientists.

Research investigation as work

The last assumption that underlies the grounded theory approach is that research should be understood and analyzed as work. Essentially we are advocating a highly self-conscious approach to the work of research: to how it is and can be actually carried out under a variety of circumstances, during its various phases, by researchers who stand in different relationships to the work of getting and examining and interpreting the information that becomes their data. Consequently, this book is not only based on an explicit sociology-of-work perspective, but is designed to help readers think in those terms about their own research endeavors. We should note also that research work consists of more than sets of tasks or a clear formulation of the goals of those tasks. It involves the organization of work – the articulation of tasks (itself a type of work) including the management of physical, social, and personal resources necessary for getting the research work done, whether working alone, with someone else, or in a team.

Perhaps it is also necessary to add that a sociology-of-work perspective emphasizes temporal features, both of the investigatory process itself and of the phenomena being studied. This constitutes our own bias toward reality, of course. For all that, we believe a sociology-of-work perspective on research activity can be useful even if a reader chooses to ignore for the moment or to downplay or deny temporal considerations when doing his or her research work. Admittedly, however, our approach to analysis, which emphasizes complexity of phenomena and the unexpected contingencies affecting both the phenomena under study and the course of the research itself, tends to bring temporality into focus for the analyst.

We should add that while much research involves routine operations and can at times be boring, assuredly also at its most creative it is exciting, fun, challenging, although sometimes extremely disturbing and painful. This means that researchers, as workers, can and should *care* very deeply about their work – not being simply possessive about its products or jealous of their research reputations, but find deep and satisfying meaning in their work. They and it are immensely interactive in exactly the sense used by John Dewey when writing about artists (he

did not regard artistic and scientific activity as basically different): An "expression of the self in and through a medium, constituting the work of art, is *itself* a prolonged interaction issuing from the self with objective conditions, a process in which *both of them* [our italics here] acquire a form and order they did not first possess" (Dewey 1934, p. 65). In short, the researcher, if more than merely competent, will be "in the work" – emotionally as well as intellectually – and often will be profoundly affected by experiences engendered by the research process itself.

Qualitative analysis of data: an introduction

Besides those general assumptions that lie behind our approach to the qualitative analysis of materials, some additional remarks will be useful before the more technical details of grounded theory analysis are discussed.

Complexity

The basic question facing us is how to capture the complexity of reality (phenomena) we study, and how to make convincing sense of it. Part of the capturing of course is through extensive data collection. But making sense of complex data means three things. First, it means that both the complex interpretations and the data collection are guided by successively evolving interpretations made during the course of the study. (The final products are analyses done at a relatively high level of abstraction: that is, *theories*.) The second point is that a theory, to avoid simplistic rendering of the phenomena under study, must be conceptually dense – there are many concepts, and many linkages among them. (Even the best monographs often are rather thin in their conceptual treatment, as betrayed by the monograph's index, which lists few if any new concepts.) The third point: It is necessary to do detailed, intensive, microscopic examination of the data in order to bring out the amazing complexity of what lies in, behind, and beyond those data. (Later, we shall say much more about complexity and capturing it through analysis.)

Experiential data

To that analysis, as will be seen, analysts bring experiences of various kinds. If not new to the research game, then they bring research skills

and savvy to their analyses. What is in their heads also in the way of social science literature also affects their analyses. This is true, whether in the form of specific hypotheses and concepts or, more diffusely, an informed theoretical sensitivity (ways of thinking about data in theoretical terms) – to nuances in their data that less well-read researchers may lack in some degree. Equally important is the utilization of experiential data, which consists not only of analysts' technical knowledge and experience derived from research, but also their personal experiences (see also the next section, Induction, Deduction, and Verification). These experiential data should not be ignored because of the usual canons governing research (which regard personal experience and data as likely to bias the research), for those canons lead to the squashing of valuable experiential data. We say, rather, "Mine your experience, there is potential gold there!"

Experiential data are essential data, as we shall see, because they not only give added theoretical sensitivity but provide a wealth of provisional suggestions for making comparisons, finding variations, and sampling widely on theoretical grounds (Schatzman, forthcoming). All of that helps the researcher eventually to formulate a conceptually dense and carefully ordered theory. The researcher's will not be the only possible interpretation of the data (only God's interpretations can make the claim of "full completeness"), but it will be plausible, useful, and allow its own further elaboration and verification.

We should add that the mandate to use experiential data gives the researcher a satisfying sense of freedom, linked with the understanding that this is not license to run wild but is held within bounds by controls exerted through a carefully managed triad of data collection/coding and memoing (to be discussed shortly). This triad serves as a genuinely explicit control over the researcher's biases.

Induction, deduction, and verification

The grounded theory of analysis involves – as does all scientific theory which is not purely speculative – a grounding in data. Scientific theories require first of all that they be conceived, then elaborated, then checked out. Everyone agrees on that. What they do not always agree on are the exact terms with which to refer to those three aspects of inquiry. The terms which we prefer are induction, deduction, and verification. Induction refers to the actions that lead to discovery of an hypothesis – that is, having a hunch or an idea, then converting it into an hypothesis

and assessing whether it might provisionally work as at least a partial condition for a type of event, act, relationship, strategy, etc. Hypotheses are both provisional and conditional. Deduction consists of the drawing of implications from hypotheses or larger systems of them for purposes of verification. The latter term refers to the procedures of verifying, whether that turns out to be total or a partial qualification or negation. All three processes go on throughout the life of the research project. Probably few working scientists would make the mistake of believing these stood in simple sequential relationship.

Because of our earlier writing in *Discovery* (1967) where we attacked speculative theory – quite ungrounded in bodies of data – many people mistakenly refer to grounded theory as "inductive theory" in order to contrast it with, say, the theories of Parsons or Blau. But as we have indicated, all three aspects of inquiry (induction, deduction, and verification) are absolutely essential. Of course, deduction without verification or qualification or even negation of an hypothesis or set of hypotheses is truncated inquiry. Obviously, too, verification cannot occur without deduction: Hypotheses for data collection without reference to implications of theoretical hypotheses are useless. And how can there be hypotheses without either thinking through the implications of data or through "data in the head" (whether experiential or from previous studies) that eventuates in so-called hunches, insights, and very provisional formulations of hypotheses?

In fact, it is important to understand that various kinds of experience are central to all these modes of activity – induction, deduction, and verification – that enter into inquiry. Consider induction first: Where do the insights, hunches, generative questions which constitute it come from? Answer: They come from experience with this kind of phenomenon before – whether the experience is personal, or derives more "professionally" from actual exploratory research into the phenomenon or from a previous research program, or from theoretical sensitivity because of the researcher's knowledge of technical literature.[2] As for deduction: Success at it rests not merely on the ability to think logically but with experience in thinking about the particular kind of data under scrutiny. The researcher is able to think effectively – and propositionally – because he or she has experiences to draw upon in thinking about those data, including the making of comparisons that help measurably in furthering the lines of deduction. Further, a special kind of prepa-

[2] See the writings of Charles Peirce, the American Pragmatist, whose concept of *abduction* strongly emphasized the crucial role of experience in this first phase of research operations (Fann 1980; Hartshorne et al. 1958).

ration underlies this deductive ability: experience not only with deductive procedures but with those used specifically in research endeavors. And verification: Quite clearly, this is not primarily a matter of activity or ability. It involves knowledge about sites, events, actions, actors, also procedures and techniques (and learned skills in thinking about them). Again that knowledge is based on personal and professional experience.

The crucial role of experience has been underplayed by philosophers of science, probably because they do not actually have a working knowledge of research, and by positively minded if methodologically reflective social scientists, who wish to rule out of court anything that smacks of "subjectivity" and who wish to minimize soft data in favor of hard (or "real") data.

If, then, experience and associated learned skills at verification, deduction, and induction are central to successful inquiry, do not talent–gifts–genius contribute to that success? Obviously the answer is yes; but not so obviously it is a qualified and complex yes. Why? Because different abilities are relevant for each of these central investigative modes. Some people are better at generative questions, intuitive flashes, hunches, etc. Some are better "theorists" – better at developing hypotheses and drawing out implications. And some are best at doing the verifying work: the laboratory whizzes, the gifted interviewers, the sensitive field observers, the highly skilled questionnaire designers. Some people can do two of these central modes of inquiry well, and some all three. Moreover, doing each well or not so well implies a continuum for each mode (verification, deduction, induction). The "real geniuses" do them all, and brilliantly. Yet as should be evident (we shall say more about this later in the book), analytic capacities can be developed, and competent if not brilliant accomplishments at one or more modes of inquiry achieved. Good research analysis can be taught and learned: It is not at all merely an innate skill.

We should add that in the event an extant grounded theory is used at the beginning or early in the research project, then deductions are made from it in the form of theoretical questions, hypotheses, suggested theoretical sampling, possible categories, and so on. They lead directly into the initial phase of collecting and analyzing data. Thus the role of deduction is the same as if the researcher began without using such a grounded theory. (See the Appendix, Discovering New Theory from Previous Theory.) This is in marked contrast to a very frequent mode of using previous theory – usually drawn from a well-known theorist, like Goffman, whose theory may be well grounded – but this theory is misused because it is not really checked out in the further inquiry. It

is only applied like a label to one's data. This practice almost totally relieves the researcher of three very important responsibilities: of (1) genuinely checking or qualifying the original data; (2) interacting deeply with his or her own data; and (3) developing new theory on the basis of a true transaction between the previous and newly evolving theory. While this practice and its citations may flatter the theorist, and may give the illusion of adding to "knowledge," it really does not advance the collective scientific enterprise. In this regard, effective social science research must follow the example of physical science research in its intertwining of the formulating of provisional hypotheses, making deductions, and checking them out – all with the use of data.

An example

Here is an example of the beginnings of a complex analysis, based on field observational data but certainly supplemented by experiential data. It will serve as a brief introduction to the grounded theory style of analysis and introduce a couple of important terms for analytic processes. This example is taken from an actual study, and the field workers did make the observations and go through some of the analytic operations described. (See Strauss et al. 1985. Other materials from this project are given in chapters on coding and memoing.)

Imagine that in a study of whether and how the use of machines in hospitals affects the interaction between staff and patients, we observe that many machines are connected to the sick persons. We can formulate a category – machine–body connections – to refer to this phenomenon. Our observations also lead us to make a provisional distinction (which may or may not turn out to be significant after further research) between those machines where the connection is external to the skin of the patient, and those where the connection is internal (through various orifices: nose, mouth, anus, vagina). This distinction involves two dimensions of the machine–body category: internal and external connections. The basic operation of making those distinctions is *dimensionalizing*.[3] But since further distinctions can be made – either by thinking about previous observations or making new ones – the process of dimensionalizing will continue. That is termed subdimensionalizing. Subdimensions may also be generated analytically by questions that

[3] This discussion of dimensionalizing was much furthered by a working session with Leonard Schatzman, a colleague, who has been thinking through the details and implications of dimensions and dimensionalizing (see Schatzman, forthcoming).

sooner or later will occur to us about some of those distinctions. Thus, about the internal connections: Don't they – or at least some of them – hurt? Are they safe? Are they uncomfortable? Are they frightening? We can think of these subdimensions (hurt, safety, discomfort, fear) dichotomously – as yes or no – or as continua running from very much to not at all. Or we can slice up a continuum roughly into "more or less" subcategories, as for instance, terribly uncomfortable, very uncomfortable, somewhat uncomfortable, a bit uncomfortable, not at all uncomfortable. (In quantitative analysis, continua can be given "values," running from 0 to 100.) All of these subdimensions, subcategories, and questions come not only from inspection of field/interview data but, understandably from our experiential data (those orifices are sensitive, so that connection probably hurts; or, that tube looks horrible coming out of his belly, so is it really safe?).

Those last questions refer to consequences: "If it looks like that, then it may have the consequences of endangering life." This may be amended by specific conditions through adding: "It may endanger his life, especially if he moves too quickly or turns over in his sleep or it falls out and then he gets an infection." Or there may be questions raised which involve the staff's strategy: "Why did they connect it up that way rather than another?" Or the patient's strategy: "Did he try to bargain to get it done another way?" Questions about interactions will also arise: "What went on between the personnel and the patient when he was being hooked up? Did they tell him beforehand and warn him? Did they just do it and so he got frightened?" (That last is a question also about consequences of interaction.)

Those questions are given provisional answers – that is, they have the status of hypotheses. Some may be checked out by further observations or interviews. But now the researcher can be more directed than previously in making observations and doing interviews. He or she is likely to realize (recognizing when observing) that a nasal connection is likely to be uncomfortable but perfectly safe, and so will interview around that hypothesis. Or thinking about unsafe conditions, the researcher may either ask staff for examples of when those connections proved unsafe for the patient – thus eliciting relevant data – or be on the lookout for unsafe nasal connections in terms of further conditions, like: because the connection got disconnected, or because of the way the connection was made.

This line of reasoning can lead to further subdimensionalizing and further questions and provisional hypotheses. Thus, for connections that become disconnected more or less easily: How do they become

disconnected? By accident, carelessness, purposefulness (as on the part of the annoyed or uncomfortable or fearful patient)? What tactics or techniques are used by the personnel to minimize or prevent disconnection: Special care? Warning the patient about moving? Emphasizing that one's safety depends on staying immobile or in not loosening the connection no matter how it hurts? Or by eliciting "cooperation," promising that the connection will remain only for several hours or be removed periodically to give relief? These questions and hypotheses and distinctions may not turn out to be "realistic"; but if they are, then further *directed inquiry* will tell the researcher: yes–no–maybe; as well as, *why*. Understandably, the researcher is likely to raise questions eventually (or observed events will occur that raise and partly answer the questions) about many more conditions; also about consequences not only for the patient but for kin, staff, different types of personnel, for the ward's functioning, and probably also for the redesign of particular models of machinery.

The directed inquiry will also very naturally and easily lead the researcher to ask: Where can I find instances of "x" or "y"? The technical term for this is *theoretical sampling* – for the researcher, after previous analysis, is seeking samples of population, events, activities guided by his or her emerging (if still primitive) theory. This sampling is harnessed at least implicitly (explicitly by the experienced researchers) to *making comparisons* according to various subdimensions. Thus, the researchers may compare, either "in imagination" or through their own experiential data, certain machine connections that are uncomfortable with those that are not. The researchers have already thought about the discomfort or the anxiety engendered by various connections. But they can go further afield and, say, make (or discover) comparisons between what happens when a dangerous disconnection occurs versus a nondangerous one: For example, once when there was actually a power blackout in the hospital, the researchers rushed around observing what was happening because various pieces of equipment had got disconnected electrically. They discovered much variation, one of the most interesting being the manual emergency motoring, done for about two hours, of dialysis machines in an associated clinical building that had no backup for the dangerously malfunctioning electrical system.

Directed by his or her theorizing, the researcher can sample even more widely by thinking about safety or discomfort with respect to other machines – whether body-connected or not – like x-ray equipment, airplanes, toasters, lawnmowers, or the body-shaking power tools manipulated by men who are employed to break up cement on street

surfaces. The purpose of thinking about those comparisons is not to pursue a more encompassing theory about machines in general or safe/dangerous machines in general, but to stimulate theoretical sensitivity in the service of generating theory about medical machinery in hospital work. Out-sampling then links with in-sampling.

Understandably, too, some ideas and thinking about those comparisons can come from personal experiences with the machines, from watching others use the machines, from reading novels or autobiographies or nonfictional literature about them.

Several points about work processes

Next, several things are especially worth noting about the basic research work processes – thinking, going to the field, observing, interviewing, note taking, analyzing. *First*, the raising of *generative questions* is essential to making distinctions and comparisons; thinking about possible hypotheses, concepts, and their relationships; sampling, and the like. These come from examination and thinking about the data, often in conjunction with experiential data. The original generative question may come from insight, which actually sparks interest in an aspect of some phenomenon and thus challenges the researcher to study "it." But these insights occur along the course of a study (although perhaps especially in the earlier phases), and open up questions about other phenomena or other aspects of the same phenomena.

Second, the researcher will be making a number of interesting, if at first quite provisional, linkages among the "discovered" (created) concepts. The coding is beginning to yield *conceptually dense theory* which will of course become much more dense as additional linkages are suggested and formulated.

Third, the theory is not just discovered but *verified*, because the provisional character of the linkages – of answers and hypotheses concerning them – gets checked out during the succeeding phases of inquiry, with new data and new coding.

Fourth, the *relevance of the coding* to the real world of data is a central issue. Of course, "there is no end to the logical elaboration of dimensions, the drawing of distinctions, the making of linkages, but to run riot with logical elaboration is dangerous – if fun. This thought process *must* be linked with, tied in tightly with, the examination and collection of new data" in order to be of service to the research itself. (We shall discuss this point later, under the heading of Deduction and Induction.)

Fifth, there is the issue of *integration*: Which dimensions, distinctions, categories, linkages are "most important," most salient – which, in short, are the *core* of the evolving theory? This issue becomes solved during the course of the inquiry. Conveying how integration happens is not easy, and we shall discuss and illustrate this work later. Suffice it here to say that integration actually begins primitively and provisionally with the first linking up of dimensions, categories, etc. Integration becomes increasingly more certain and "tighter" as the research continues. The *core category or categories* that will best hold together (link up with) all the other categories – as they related to it and to each other – will take hard work and perhaps special techniques to put together in a convincing fashion: convincing both to the researcher and to those who will read his or her resultant publications.

Sixth, theoretical ideas are kept track of, and continuously linked and built up by means of *theoretical memos*. From time to time they are taken out of the file and examined and sorted, which results in new ideas, thus new memos. As research proceeds to later phases, memo writing becomes more intense, more focused, and memos are even more frequently sparked by previous memos or sum up and add to previous ones. *Sorting* of memos (and codes) may occur at any phase of the research. Both examination and sorting produces memos of greater scope and conceptual density. The systematic operation of sorting is especially important in later phases, as the analyst moves into planning the writing up of materials for publication.

Seventh, it is vitally important to recognize the *temporal* as well as *relational* aspect of the triad of analytic operation: data collecting, coding, memoing. Grounded theory practitioners need to understand how very different their perspective on that triad is from that of most other styles of analysis. Figure 1, a simplified diagram of a coding paradigm will illustrate some of the main features of this triad. Note that data collecting leads quickly to coding, which in turn may lead equally quickly – or at least soon – to memoing. Either will then guide the searches for new data. *Or* – and this is important to understand – they may lead directly to additional coding or memoing. Or – please note! – they may lead to inspecting and coding of *already* gathered (and perhaps already analyzed) data. The *return to the old data* can occur at any phase of the research, right down to writing the last page of the final report of the theory. Furthermore, as the diagram indicates, at any phase of the research coding can lead to more coding; or memoing, directly to further and more integrated memos, helped out of course by the sorting of codes and memos.

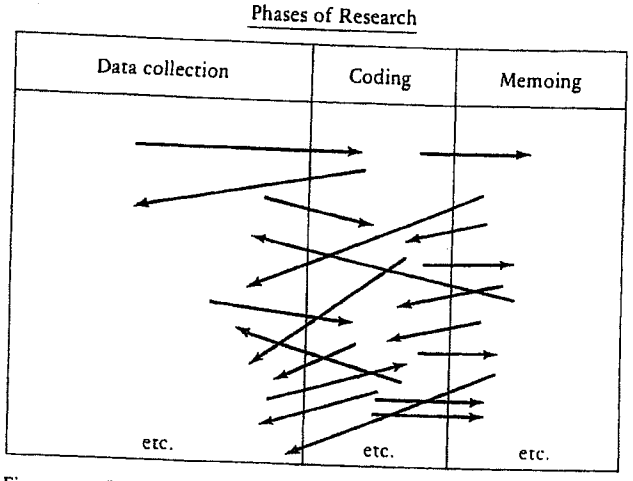

Figure 1. Coding paradigm.

This reexamination of all data throughout the life of the research project is a procedure probably engaged in by most qualitative researchers. But they do not usually double back-and-forth between collecting data, coding them, memoing in terms of data collection, coding, and memoing. The more positivistic research traditions proscribe the use of old data for verifying hypotheses, and so drive the researcher forward in a more linear direction, thereby cutting out the potential dividends of this recommended doubling back-and-forth procedure.

Eighth, during the *writing*, need for additional integration will often be recognized, the researcher sometimes then going back to the data, collecting some new data, or thinking through the sorted memos and codes, to "fill in," thus achieving the necessary integration. However, there is much variation concerning how much those operations will be relied upon during the writing period. How much depends on the degree of thoroughness with which the coding and memoing has been carried out; also on what the researcher realizes ought to be emphasized for particular audiences for whom he or she is writing; also on the writer's previous research/writing experience. Also, in team research it happens that so much data will accumulate, so fast, that although much coding is done and many theoretical memos are written, when the researchers sit down to write their various papers and monographs, they discover substantial holes in the previous analyses. This is especially so when some decisions about what to write, and for whom, evolve fairly late in the study.

The writing then does not just reproduce what is in the memos, although memos can often be rephrased or parts of them can be used pretty much as written in the final publication. The writing is, then, both analytic and creative. It can result in various types of publications (papers, monographs) and speeches, depending both on the substance of the research and the researcher's perceptions of the audiences. But the main point is that all the technical operations touched on in this section go on *continually*, from the outset of the research project until its close.

In the reception to published theory of this kind there is, we have found, a double-edged irony corresponding to two contrasting audiences. When lay people, or professional people of the population who have been studied – such as nurses or physicians – read the paper or monograph, they do not read it as theory, but either as a more or less accurate description of what's been happening to themselves and others of their acquaintance, or as "a new way of seeing what we all know that's very useful" – even an eye-opener. Then there is the audience of social scientists, who may read the publication, recognizing its "solid sociology," to quote an admirer of one of our publications, but without recognizing that the bright and even "brilliant ideas" in the publication arose not from personal gifts but from the hard work of research. The first irony should very much please the grounded theorist. The second will on occasion drive him or her wild with annoyance; but so be it: More-informed social science colleagues will know better.

A glossary of major terms

A number of important terms pertaining to qualitative analysis have appeared in the preceding section. They will be further discussed in the next chapter and then used throughout the book. We shall give capsule definitions of them now, since it is essential to have a firm grasp of them.

Data collection. the finding and gathering – or generating – of materials that the researcher will then analyze

Experiential data. data "in the head," drawn from the researcher's personal, research, and literature-reading experiences

Coding. the general term for conceptualizing data; thus, coding includes raising questions and giving provisional answers (hypotheses) about

categories and about their relations. A code is the term for any product of this analysis (whether category or a relation among two or more categories).

Dimensionalizing. a basic operation of making distinctions, whose products are *dimensions and subdimensions.*

Category. since any distinction comes from dimensionalizing, those distinctions will lead to categories. (Thus, *Machine–body connection* is a category.)

Property. the most concrete feature of something (idea, thing, person; event, activity, relation) that can be conceptualized, which will allow the order of specificity required by the analyst for purposes of his or her research

Hypotheses (used exactly as in the usual scientific lexicon). a provisional answer to a question about conceptual relationships

Core category. a category that is central to the integration of the theory

Theoretical sampling. sampling directed by the evolving theory; it is a sampling of incidents, events, activities, populations, etc. It is harnessed to the making of *comparisons* between and among those samples of activities, populations, etc.

Theoretical saturation. when additional analysis no longer contributes to discovering anything new about a category

Conceptual density. the multiplicity of categories and properties *and* their relationships

Integration. the ever-increasing organization (or articulation) of the components of the theory

Variation. product of comparisons; grounded theory analysis rests on a multitude of comparisons – directed by theoretical sampling – and so grounded theory is multivariate. Making comparisons among categories and properties involves connecting (*crosscutting*) them.

Theoretical sensitivity. sensitive to thinking about data in theoretical terms

Theoretical memos. writing in which the researcher puts down theoretical questions, hypotheses, summary of codes, etc. – a method of keeping track of coding results and stimulating further coding, and also a major means for integrating the theory

Theoretical sorting. sorting of the theoretical memos in the service of integration: Codes are also sorted, toward the same end.

Integrative diagrams. a visual device which also furthers cumulative integration along the full course of the research

Generative questions. questions that stimulate the line of investigation in profitable directions; they lead to hypotheses, useful comparisons, the collection of certain classes of data, even to general lines of attack on potentially important problems.

PART 2

Grounded theory analysis: main elements

In this portion of the introductory chapter, a number of essential research operations are presented. Some of the discussion cannot be completely understood, at least in detail, until the illustrative materials in later pages help to provide visualization for the points made here. So, you might wish to read this chapter quickly to get an overview, then return to it, or parts of it, for reading or study later.

Our approach to the qualitative analysis of data is termed grounded theory "because of its emphasis on the generation of *theory* and the *data* in which that theory is *grounded*."[4]

Grounded theory "is a *detailed* grounding by systematically" and intensively "analyzing data, often sentence by sentence, or phrase by phrase of the field note, interview, or other document; by 'constant comparison,' data are extensively collected and coded," using the operations touched on in the previous section, thus producing a well-constructed theory. The focus of analysis is *not* merely on collecting or

[4] As noted in the preface, this part of Chapter 1 is reproduced almost wholly from Barney Glaser's *Theoretical Sensitivity*, 1978, with some editing and supplementation. The quoted sentences and paragraphs are identifiable by the relevant quotation marks. For more detailed statement of these technical aspects of the grounded theory mode of analysis, readers are advised to consult *Theoretical Sensitivity*.

ordering "a mass of data, but on *organizing many ideas* which have emerged from analysis of the data."

We have already seen the basic ingredients in producing complex, conceptually woven, integrated theory; theory which is discovered and formulated developmentally in close conjunction with intensive analysis of data. These procedures vary during the course of a research project. So, that issue will be discussed first, then we shall turn to a more detailed discussion of elements of the main procedures touched on previously. They are:

1. the concept-indicator model which directs the coding
2. data collection
3. coding
4. core categories
5. theoretical sampling
6. comparisons
7. theoretical saturation
8. integration of the theory
9. theoretical memos
10. theoretical sorting

Research phases and the operations

We shall now discuss the essential procedures for discovering, verifying, and formulating a grounded theory. These are in operation all through the research project and, as the case illustrations later will show, go on in close relationship to each other, in quick sequence and often simultaneously. But what about their relations to different phases of the entire research project? More will be said in answer to that question, but a few words should help in reading the concrete materials to be presented throughout this book.

As we shall see, the earliest phases of the research are more "open" than later ones are. There is no attempt to foreclose quickly on one or more categories. Many months may pass before the researcher is more or less certain of them and very many more before those core categories are saturated, and linked in a multiplicity of ways with other categories. In the earliest phases, a number of categories probably will be generated which later will be dropped as not very useful, or as unrelated to the core categories. Likewise a number of hypotheses will fall by the wayside, but are freely if provisionally generated by the enthusiastic researcher. Yet, from the earliest days, theoretical sampling directs the data collection and comparative analysis is done from the word go. The first

memos are far less integrative than later they will be, and they too may poke up blind alleys or be focused very closely on the early microscopic analysis of data.

Once the core category or categories have been committed to, then the researcher will be seeking to relate other categories to them, thereby gradually densifying the theory. Also, more confidence will be placed in any new categories that "emerge" from further coding. Further highly directed theoretical sampling will function to generate additional relevant categories and properties. There is likely to be some sorting too, both of codes and memos, during this later phase (presumably by the middle of the project). Memos are likely to become increasingly elaborate, summarizing the previous ones; or focused closely on closing gaps in the theory. Earlier integrative diagrams will be made more elaborate, covering both more concepts and more connections among them. All of that continues until the last phases of the project.

Near the end, achieving integration will be a major focus. Also, considerable thought will be directed at which audiences to write for or speak to, and about what topics; also, what published papers to begin aiming for. Finally, there is the task of pulling the entire theory together for its presentation in a monograph. If a team is involved in this research, then there will be conferencing over who will write which papers, give which talks, write which chapters of the monograph. Or if they decide to publish more than one monograph, there is the question of: Who will write which monographs or portions of them?

Having said all that, we should emphasize that no sequential mini-steps can firmly be laid out in advance of the evolving phases of a given research project. Each enterprise will have its own detailed sequences, depending on: the circumstances of what kind of data are available, accessible and required; the nature of the data and the interpretations that the researcher will make of them; the experience of the researcher or researchers; the many contingencies that affect both the researcher personally (and interactionally, if a team also); the character of the audiences for whom they decide to write their publications; and the scope and generality of the theory for which the researchers aim. Only the general lineaments of the unfolding project can be anticipated in advance. The major differences between the grounded style of qualitative analysis and other qualitative analysis modes, however, is not in the relative unpredictability of project phases, but the differences per stage in the combinations and permutations of the operations (theoretical sampling, comparative analysis, theoretical saturation, memo

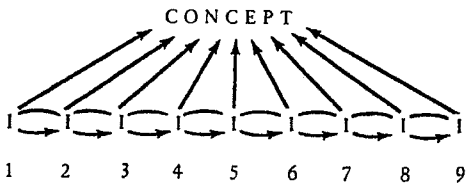

Figure 2. Indicator-concept model.

sorting, and so forth). These operations are essential to the development of densely woven and tightly integrated theory.

Basic operations

Concepts and indicators

"Grounded theory is based on a *concept-indicator* model, which directs the *conceptual* coding of a set of *empirical indicators*. The latter are actual data, such as behavioral actions and events, observed or described in documents and in the words of interviewees and informants. These data are indicators of a concept the analyst derives from them, at first provisionally but later with more certainty." (See the chapters on codes and memos and the chapter illustrating the research seminar analyses, where many illustrations of this indicator-concept model are given and sometimes pointed out explicitly for the reader.)

The concept-indicator model in Figure 2 is based first of all on the constant comparison of indicator to indicator. That is: Many indicators (behavioral actions/events) are examined comparatively by the analyst who then "codes" them, naming them as indicators of a class of events/behavioral actions. He or she may give this class a name, thinking of it then as a coded category. By making "comparisons of indicator to indicator the analyst is forced into confronting similarities, differences, and degrees of consistency of meaning among indicators. This generates an underlying uniformity, which in turn results in a coded" *category*. A second procedural step is that after "a conceptual code is generated, then indicators are compared to the emergent concept . . . . From the comparisons of additional indicators to the conceptual codes, the codes are sharpened to achieve their best fits to data." Meanwhile "further properties of categories are generated, until the codes are verified and saturated," yielding nothing much new.

In this model of concept indicators, "concepts and their dimensions have *earned* their way into the theory by systematic generation from data . . . *Conceptual specification* is at the focus of grounded theory . . . because the operational meaning of the concept derives from the use of its earned distinctions in the grounded theory."

"Changing indicators, thereby generating new properties of a code, will proceed only so far before the analyst discovers saturation of ideas through the *interchangeability of indicators*." That is, the events/behavioral actions which are converted analytically into indicators may vary in detail or in fact just be repetitious – but anyhow the indicators seem to "add up to the same thing" analytically. So the more the researcher "finds indicators that work similarly regarding their meaning for the concept, the more the analyst *saturates* the properties of the concept for the emerging theory. Nothing new happens as he or she reviews the data. The category and its properties exhaust the data. Meanwhile the analyst continues to saturate other categories by use of the constant comparative method."

Data collection

There is some ambiguity associated with the term *data collection*. Many social scientists do generate their data, through field observation, interviewing, producing videotapes, taping proceedings of meetings, and so on. But, as noted earlier, there are other sources of data: published documents of all kinds and private documents like letters and diaries. Use of those latter sources involves work too – searching for the data, getting access to them, taking notes on them, and nowadays xeroxing those data. In some kinds of library research, the researcher will even use the library much like an ethnographer, deciding upon which shelves to find the data sources (books, periodicals), and like the ethnographer happily coming upon fortuitously useful data, too (see Glaser and Strauss 1967).

The initial data collected may seem confusing, the researcher flooded by their richness and their often puzzling and challenging nature. It should not remain *that* confusing (only challenging) for very long because the analysis of these data begins (in our style of research) with the very first, second, or third interview or after the first day or two of fieldwork if at all feasible. It follows also that the next interviews and observations become informed by analytic questions and hypotheses

about categories and their relationships. This guidance becomes increasingly explicit as the analysis of new data continues.

Data collection never entirely ceases because coding and memoing continue to raise fresh questions that can only be addressed by the gathering of new data or the examining of previous data. Theory-guided data collection often leads to the search for – or quick recognition of – valuable additional sources of data: for example a series of directed interviews to supplement the more casual interviews done during the daily fieldwork; or the use of published biographies to supplement a series of interviews. We call these "slices of data," for different kinds of data give different views or vantage points, allowing for further coding, including the discovery of relationships among the various categories that are entering into the emergent theory.

Coding

Coding, as noted in a previous section, is an essential procedure. Any researcher who wishes to become proficient at doing qualitative analysis must learn to code well and easily. The excellence of the research rests in large part on the excellence of the coding. (See Chapter 3 for illustrations and further discussion.)

Coding paradigm. One important point about coding that is sometimes misunderstood is this: While coding involves the discovery and naming of categories, it must *also* tell the researcher much more than that. It is not enough, for instance, to code an event qua indicator as an instance of a category – say, as "machine breakdown" – by writing the name of the category in the margins of the page next to the indicating lines of print. Also, the researcher needs to code the associated subcategories which are reflected either in the same lines or which will be reflected in other lines within the same or different interview, fieldnote, or document. (See especially Chapters 3–5.)

So we suggest the following *coding paradigm*. It is central to the coding procedures. Although especially helpful to beginning analysts, in a short time this paradigm quite literally becomes part and parcel of the analyst's thought processes. Whether explicit or implicit, it functions as a reminder to code data for relevance to whatever phenomena are referenced by a given category, for the following:

conditions
interaction among the actors

strategies and tactics
consequences

Because beginning researchers sometimes seem to experience difficulty in discovering "conditions" when inspecting their data, we shall note the following. Conditions are often easy to discover – indeed sometimes the interviewees or actors will point to them specifically – but if not, then look for cues like the use of words such as "because," "since," "as," or phrases like "on account of." Likewise, consequences of actions can be pointed to by phrases like "as a result," "because of that," "the result was," "the consequence was," and "in consequence." Strategies and the more specific tactics associated with strategies seem to present no difficulties for inexperienced analysts. *Interactions* are also easy to discover: They are those interactions occurring between and among actors, other than their straightforward use of tactics and strategies. Exemplifications of how the coding paradigm works will be found throughout this book. Remember that without inclusion of the paradigm items, coding is not coding.

Open coding. The *initial* type of coding done during a research project is termed open coding. This is unrestricted coding of the data. This open coding is done (as some of the case illustrations will show) by scrutinizing the fieldnote, interview, or other document very closely: line by line, or even word by word. The aim is to produce concepts that seem to fit the data. These concepts and their dimensions are as yet entirely provisional; but thinking about these results in a host of questions and equally provisional answers, which immediately leads to further issues pertaining to conditions, strategies, interactions, and consequences. As the analyst moves to the next words, next lines, the process snowballs, with the quick surfacing of information bearing on the questions and hypotheses, and sometimes even possible crosscutting of dimensions. A single session with a single document can often astonish even the experienced researcher, especially when the document at first glance seemed not to promise much in the way of leads. The point is really that the potential is not so much in the document as in the relationship between it and the inquiring mind and training of a researcher who vigorously and imaginatively engages in the open coding.

Novices at this type of coding characteristically get hung up, will argue intensely, about the "true" meaning of a line – or about the "real" motives of the interviewee lying behind the scrutinized line. In terms of open coding, their concern is entirely irrelevant. Why? Because

the aim of the coding is to *open up* the inquiry. Every interpretation at this point is tentative. In a genuine sense, the analyst is not primarily concerned with this particular document, but for what it can do to further the next steps of the inquiry. Whatever is wrong in interpreting those lines and words will eventually be cancelled out through later steps of the inquiry. Concepts will then work or not work, distinctions will be useful or not useful – or modified, and so forth. So the experienced analyst learns to play the game of believing everything and believing nothing – at this point – leaving himself or herself as open as the coding itself. For all that, the coding is grounded in data on the page as well as on the conjunctive experiential data, including the knowledge of technical literature which the analyst brings into the inquiry.

This grounding in both sources of data gets researchers away from too literal an immersion in the materials (documents, fieldnotes, interviews, etc.) and quickly gets them to thinking in terms of explicit concepts and their relationships. This stepping away into conceptualization is especially difficult for even experienced researchers who may, in a particular study, either have gone a bit native through personally participating in the field of study, or who know too much experientially and descriptively about the phenomena they are studying and so are literally flooded with their materials. Yet the conceptual stepping back must occur if one is to develop theoretical understanding and theories about the phenomena reflected in the materials. Open coding quickly forces the analyst to fracture, break the data apart analytically, and leads directly to excitement and the inevitable payoff of grounded conceptualization. In research seminars, open coding is additionally valuable since students often find it much easier to code someone else's data, being more emotionally distant from them, and so learn through the open-coding procedures how more quickly to fracture their own data.

A word should be said here, however, about the difficulties novices often have in generating genuine categories. The common tendency is simply to take a bit of the data (a phrase or sentence or paragraph) and translate that into a precis of it. For instance: The interviewee is expressing grief or joy or aggression since he or she has declared "I was full of grief" or "I was mad as hops and so slugged him." The novitiate analyst is merely writing shorthand translation notions on the side of the interview page rather than generating theoretical categories. (In effect they are, as are many researchers who use other methods of analysis, remaining totally or mostly on a descriptive level, not much

different from that of the actors themselves.) However, when a nurse tells the researcher that "I tried to keep my composure when the patient was yelling, by leaving the room" then that phrase can be converted analytically into "professional composure," plus notations about the structural condition threatening her composure and the tactic she uses for maintaining her composure. This can lead the researcher to write a memo in which questions are raised immediately about other pertinent conditions and tactics, as well as about situations where the nurse's tactic failed, or she had no chance to use one, and so lost her composure.

In our teaching experience, the most difficult step (other than integrating the total analysis) for beginners at this style of analysis is actually to get off the ground with genuine coding. Until they have learned this, they are frustrated. Yet it is essential that they learn this skill, since everything that follows rests on it. Other than the general guidelines given directly below (and in Chapter 3, on coding), we find in teaching students that the following *rules of thumb* are useful:

1. Look for in-vivo codes, terms used by the people who are being studied. The nurse's "tried to keep my composure" is an instance.
2. Give a provisional name to each code, in-vivo or constructed. Do not be concerned initially about the aptness of the term – just be sure to name the code.
3. Ask a whole battery of specific questions about words, phrases, sentences, actions in your line-by-line analysis.
4. Move quickly to dimensions that seem relevant to given words, phrases, etc.
5. These dimensions should quickly call up comparative cases, if not then concentrate on finding them.
6. Pay attention to the items in the coding paradigm, as previously listed.

There are several additional guidelines for open coding that tend to ensure its proper use and success. "The *first* is to ask of the data a set of questions. These must be kept in mind from the very beginning. The most general question is, *What study are these data pertinent to?* This question keeps reminding the researcher than an original idea of what the study was may not turn out to be that at all – in our experience often it is not. [The case illustrations drawn from the research seminars will show how that can happen.] Another question to ask continually when studying the data is, *What category does this incident indicate?* This is the short form. The long form is, What category or property of a category, or what part of the emerging theory, does this incident indicate? As the theory becomes increasingly well formulated this question becomes easier to answer. The continual asking of this question helps to keep the analyst from getting lost in the rich data her/himself,

by forcing the generation of codes that relate to other codes. Lastly, the analyst continually asks: *What is actually happening in the data?* What is the basic problem(s) faced by the participants? What accounts for their basic problem or problems? [Another way to phrase all of this is, What's the main story here, and why?] All of these questions tend to force the generation of a core category or categories which will be at the center of the theory and its eventual write-up."

The *second* guideline for open coding – remember, this is primarily an initial coding procedure – is to *analyze the data minutely.* As noted several times earlier, this means frequently coding minutely. This effort is entirely necessary "for achieving an extensive theoretical coverage which is also thoroughly grounded." A contrasting "approach to open coding (the *overview approach*) is to read the data over rather quickly, which yields then an impressionistic cluster of categories. We do not recommend this approach by itself because it yields only a few ideas and does not force the evolution of conceptual density. It does not, either, give any idea of what has been missed. To continue in that vein gives conceptually thin and often poorly integrated theory."

The more-microscopic approach to open coding "minimizes the overlooking of important categories, leads to a conceptually dense theory, gives the feeling – to the reader as well as to the analyst – that probably nothing of great importance has been left out" of the theory, and forces both verification and qualification of the theory. We should note, however, that when a code seems relatively saturated – "nothing new is happening" – then the analyst will find himself or herself moving quickly through the data, finding repetitions in the line-by-line examination, and so will scan pages until something new catches the eye. Then the minute examination begins again. Indeed, additional data gathering, especially when guided by careful and imaginative theoretical sampling, is very likely to call again for microscopic analysis. (The seminar cases in this book will illustrate very clearly this intense scrutiny, as the students linger for many minutes over particular words, phrases, and sentences, doing their line-by-line analyses.)

So this kind of intensive analysis may be done from time to time. The rule of thumb here is to do this if you sense that some portions of the total analysis are not satisfying or important relationships among categories might be nailed down by additional open coding. Of course, given the usual masses of data, you cannot continue to do open coding more than occasionally – but then there would be no point in doing that anyhow. However, once you sense the usefulness of again engaging in open coding, do not delay the work. The sooner, the better, since

that may lead quickly to useful theoretical sampling and slightly redirect your new data collecting.

A *third* important guideline for open coding is: "frequently, to interrupt the coding in order to write a theoretical memo. This leads quickly to accumulated memos as well as moves the analyst further from the data and into a more analytic realm." A *fourth* guideline is: "The analyst should not assume the analytic relevance of any 'face sheet' or traditional variable such as age, sex, social class, race, until it emerges as relevant. Those, too, must earn their way into the grounded theory."

It is important to understand that "open coding both verifies and saturates individual codes." Initially they are likely to be crude, so they will need much modification. Anyhow they are provisional so will end up considerably modified, elaborated, and so on. Hence, the analyst must not become too committed to the first codes, must not become "selective too quickly, tempting as that is, since initial codes can seem highly relevant when they are actually not. Open coding proliferates codes quickly, but the process later begins to slow down through the continual verifying that each code really does fit . . . . Eventually the code gets saturated and is placed in relationship to other codes, including its relation to the core category or categories – if, indeed, they or it are not actually the core."

Axial coding. Axial coding is an essential aspect of the open coding. It consists of intense analysis done around one category at a time, in terms of the paradigm items (conditions, consequences, and so forth). This results in cumulative knowledge about relationships between that category and other categories and subcategories. A convenient term for this is *axial coding*, because the analyzing revolves around the "axis" of one category at a time. It is unlikely to take place during the early days or even weeks when the initial data are collected and analyzed. However, axial coding becomes increasingly prominent during the normally lengthy period of open coding, before the analyst becomes committed to a core category or categories and so moves determinedly into selective coding (to be discussed next). During the open-coding period, however, the very directed axial coding alternates with looser kinds of open coding, especially as the analyst examines new aspects of the phenomena under study. It also runs parallel to the increasing number of relationships becoming specified among the many categories, whether this part of the coding is done as intensively as the axial coding or not. Of course, within this increasingly dense texture of conceptualization,

linkages are also being made with the category, or categories, that eventually will be chosen as "core."

Selective coding. Selective coding pertains to coding *systematically* and concertedly for the core category. "The other codes become subservient to the key code under focus. To code selectively, then, means that the analyst delimits coding to only those codes that relate to the core codes in sufficiently significant ways as to be used in a parsimonious theory." The core code becomes a guide to further theoretical sampling and data collection. The analyst looks for the conditions, consequences, and so forth, that relate to the core category, coding for them. Selective coding then, is different from open coding but occurs within the context developed while doing open coding. During selective coding, understandably, the analytic memos become more focused and aid in achieving the theory's integration. Selective coding can begin relatively early, but becomes increasingly dominant, since it is more self-consciously systematic than is open coding.

Sociologically constructed codes and in vivo codes

"The categories are of two types" – sociological constructs and in vivo codes. The latter "are taken from or derived directly from the language of the substantive field: essentially the terms used by actors in that field themselves." Often while doing open coding, the researcher will hear the actors using these terms, and will incorporate them into his or her analysis. "In vivo codes tend to be the behaviors or processes which will explain to the analyst how the basic problem of the actors is resolved or processed. These codes fracture the data directly because they represent analytic categories, as used by the researcher." They can also lead to associated theoretical codes: "for example, 'monitoring' a patient's clinical conditions implies – and the actors often say this explicitly – various conditions under which the monitoring is done, the consequences of the monitoring, and so on."

In vivo codes "have two characteristics: analytic usefulness and imagery. Their analytic usefulness relates the given category to others, with specified meaning, and carries it forward easily in formulation of the theory. Imagery is useful insofar as the analyst does not have to keep illustrating the code in order to give it meaning. Its imagery implies data that have sufficient meaning so that the analyst does not clutter his or her writing with too many illustrations. In vivo terms have

a very vivid imagery, inclusive of much local interpretative meaning: they have 'grab' for the participants. And they are seldom forgotten by readers because their terms are colorful. They also have much analytic force since the actors do use them with ease and with sufficiently precise meaning."

"Sociological constructs, on the other hand, are codes formulated by the sociologist ('awareness context,' 'illness trajectory')." (The constructs of course need not be sociological but psychological or anthropological, and so forth, depending on the disciplinary theory that is being formulated.) These constructs "are based on a combination of the researcher's scholarly knowledge and knowledge of the substantive field under study. As a result, they can add more sociological (social science) meaning to the analysis than in vivo codes. They add scope by going beyond local meanings to broader social science concerns. They have much analytic utility because they are constructed clearly and systematically. They may have little imagery (some analysts think that the flatter they are, the more scientific and less impressionistic they are; but others prefer them to resonate with more imagery)."

In the illustrations given later, readers will see the analysts generating many in vivo and sociological codes. As mentioned earlier, this generation is a provisional matter and so is the *labeling* of codes, which is easily changed if better terms are invented later. It is important that researchers should feel free to invent and change those terms. "There is little point in struggling to find exactly the right term, especially when one first notices the phenomenon which leads to the labeling – the important activity is first to notice and then invent or apply a term resonant enough so that the category can be referenced, focused on, and remembered." Analysts can learn to coin these terms with some facility after some experience in doing this style of qualitative analysis. Of course, that facility is not just a linguistic matter but a matter of improving one's theoretical sensitivity *and* associated analytic ability.

Core categories

"The goal of grounded theory is to generate a theory that accounts for a pattern of behavior which is relevant and problematic for those involved. The generation of theory occurs around a *core* category (and sometimes more)." "Since a core category accounts for most of the variation in a pattern of behavior," its different kinds of appearances

under different conditions, "the core category has several important functions for generating theory. It is relevant and works. Most other categories and their properties are related to it, which makes it subject to much qualification and modification. In addition, through these relations among categories and their properties, it has the prime function of *integrating* the theory and rendering it *dense* and *saturated* as the relationships are discovered. These functions then lead to theoretical *completeness* – accounting for as much variation in a pattern of behavior with as few concepts as possible, thereby maximizing parsimony and scope."

"The analyst should consciously look for a core variable when coding data. While constantly comparing incidents and concepts, he or she will generate many codes, being alert to the one or two that might be the core. The analyst constantly looks for the 'main theme,' for what appears to be the main concern of or problem for the people in the setting, for what sums up in a pattern of behavior the substance of what is going on in the data, for what is the essence of relevance reflected in the data." (As noted earlier, What's the main story here? is a kind of motto–question that the analyst asks repeatedly, to remind himself or herself to keep trying to answer the above questions.)

"As the analyst asks those questions, while analyzing, he or she becomes sensitized to their potential answers." "Possible core categories should be given a 'best fit' label as soon as possible, so that there is a handle for thinking about them. The researcher may have a feel for what is the core, but be unable to formulate it to his or her satisfaction, so must use a provisional label until a better one can be formulated."

"After several workable coded categories develop, the analyst attempts to *theoretically saturate* as much as possible those which seem to have explanatory power." Thus, relations among categories and their properties become apparent and conceptually dense. Theoretical sampling is done to further the saturation of categories because they are related to many others and recur often in the data. With qualitative analysis, "these relationships must be kept track of in memos, which get spread out or filed until sorted," and get built into integrative memos. "The core category must be proven over and over again by its prevalent relationship to other categories."

"The more data, the more certain one can become of the eventually chosen core category. Time and data can be expensive; in smaller studies the researcher often has to take a chance: and certainly deciding on a core category can test skill and ability. If the analyst decides too

rapidly, using a relatively small amount of data, there is a risk that he or she might end up with an undeveloped theory which has little integration and little explanatory power."

There are several criteria for judging which category should serve as the core category.

1. "It must be *central*, that is, related to as many other categories and their properties as is possible, and more than other candidates for the position of core category. This criterion of centrality is a necessary condition for putting a category at the heart of the analysis: It indicates that the category accounts for a large portion of the variation in a pattern of behavior."
2. "The core category must appear *frequently* in the data. (More precisely: The indicators pointing to the phenomena represented by the core category must appear frequently.) By frequent recurrence it comes to be seen as a stable pattern, and consequently becomes increasingly related by the analyst to other categories. If it does not appear frequently, that does not mean that it is uninteresting, only that it is not the core category."
3. "The core category *relates easily* to other categories. These connections need not be forced; rather they come quickly and abundantly. But because the core category is related to many other categories and recurs frequently, it takes more time to saturate the core categories than the others."
4. "A core category in a substantive study has *clear implications for a more general theory*. (See Chapter 11, on generating a formal theory.) Thus, an analyst looking at hospital shifts sooner or later may realize the implications of shifts as a basic structural condition for any twenty-four hours a day work operation, and begin to conceive of generating a theory about work shifts in organizations. The various analytic operations which follow, however, have to utilize data bearing on work shifts from many different substantive areas. Intensive scrutiny of these data is necessary, of course, before the core category or categories for this general theory can be determined."
5. "As the details of a core category are worked out analytically, the theory moves forward appreciably."
6. "The core category allows for building in the *maximum variation* to the analysis, since the researcher is coding in terms of its dimensions, properties, conditions, consequences, strategies and so on." All of these are related to different subpatterns of the phenomenon referenced by the core category. Such variation (also called variance) is, as a colleague once expressed to us, emphasized more usually in quantitative analysis than in discussions of qualitative analysis. He spoke accurately, since many qualitative analysts do not seek for variance, but for very general patterns. It is one of the hallmarks of the grounded theory mode, however, to seek variation. (See additional remarks on this topic, a little further on.)

Who should code?

When it is a matter of an individual researcher embarked on his or her project, the answer to that question is obvious. But what if a team is working together on a project? Should all its members code, or only the most experienced, the most efficient, the most brilliant coders; or the professor rather than student assistants; or, on a large project, the top echelon and not the mere data collectors? Some years ago, a qualitative researcher, Julius Roth (1963) severely criticized the principle investigators of survey researches for their exploitation of the "hired hands," who did nothing but the dirty work of data collecting – contrasting this situation with the deep commitment and involvement of the typical fieldworker, who of course did all the research work, including the brainy-work of coding. Those are the two extreme answers to the issue of who should do the coding.

However, the reasonable answer to this issue takes its cues from structural and organizational conditions bearing on the project, on its aims and its audiences. For instance, a large cross-site qualitative project with, say, two professors back home directing it, and concerned with producing "good results," and fast results (for career reasons), might handle the who-should-code issue quite differently than might – and probably do – smaller and more collaborative teams consisting of peers or virtual peers (cf. Miles 1983, especially pp. 131–2). In these terms, then, think of organizational conditions like amount of funding, numbers of data sites, amount of data to be collected, number of team participants, the degree of homogeneity of team composition. The team structures can correspondingly look different: some are hierarchical, some quite collaborative, and so on. But also, the aims of the project might include – in their various combinations of salience – reaching fast results, or the "best" results, or the most effective results for a given expected audience. Or they might include furthering the creativity of each team member, or of the total team which is expected to do further research together. And the product of all this productive research activity can take various forms during a given project: a collectively written monograph, or two or more monographs written by different members or combinations of members, individual or joint or collective papers – or all of these.

So the answer to the coding issue is going to be inevitably and profoundly affected by such considerations. Each person on a team may code his or her materials, because of greater familiarity with the data – and because there is *so* much of it cumulating for the total project. But, at a team meeting, they may together begin coding someone's presented material, or throwing in individually collected data during the analytic discussion. (See Chapter 6 for an instance of this.) Or one may code some of another's data after reading a memo by the

other. Or two teammates may meet in a session to do (or end up doing) joint coding. And they may do that after a team meeting, too, or reading a memo, etc.

My own research projects over the years have involved small teams, composed of more or less experienced people, all trained initially by me, and ended up doing highly collaborative work. And work designed to produce both "findings" and theory, but also designed to increase the creative potentials of each member. So every team participant engages in all the research procedures outlined in this chapter. Sometimes each does that separately, sometimes in twos, or as an entire team, depending on circumstance or ad hoc design. In large part, they tend to code their own data: That is understandable, but perhaps we have failed a bit in not forcing more intercoding of each other's materials, leaving that mainly to joint and team sessions.

Anyhow, to summarize with these *guidelines* pertaining to non-solo projects devoted to doing really creative research – I believe:

1. Each data collector should code much of his or her own data, but
2. code some of the others' data, separately as well as jointly and as a total team,
3. and this should be done from the onset of the initial data collecting to the very end of the project;
4. meanwhile, all should be engaged in theoretical sampling, making comparative analyses, conceptually densifying, integrating, etc.

I should add that there sometimes is one especially difficult problem encountered by students taught in our research seminars. When they attempt to code their own materials alone, without the support of the seminar's analytic discussion, then they may find this not nearly as easy or "deep" and may not have sufficient self-confidence in their coding. For this reason they are urged to meet occasionally without the instructor, as well as to work jointly with another student, between the only occasional opportunities to present their materials in class or to confer individually with the instructor.

Theoretical sampling

Theoretical sampling is a means "whereby the analyst decides *on analytic grounds* what data to collect next and where to find them." "The basic question in theoretical sampling is: *What* groups or subgroups of populations, events, activities (to find varying dimensions, strategies, etc.)" does one turn to *next* in data collection. And for *what* theoretical

purpose? "So, this process of data collection is *controlled* by the emerging theory." It involves, of course, much calculation and imagination on the part of the analyst. When done well, this analytic operation pays very high dividends because it moves the theory along quickly and efficiently. This type of sampling, so essential to the grounded theory mode of analysis, is of course neither the same as is utilized in quantitative research nor subject to the same canons (see Glaser and Strauss 1967).

Neither is it what Leonard Schatzman has aptly termed *selective sampling* (Schatzman and Strauss 1973), a frequently used sampling method in qualitative analysis. "Selective sampling refers to the calculated decision to sample a specific locale or type of interviewee according to a preconceived but reasonable initial set of dimensions (such as time, space, identity) which are worked out in advance for a study."