CHAPTER 3

THEORETICAL SAMPLING

"Theoretical sampling is the process of data collection for generating theory whereby the analyst jointly collects, codes, and analyzes his data and decides what data to collect next and where to find them, in order to develop his theory as it emerges. This process of data collection is *controlled* by the emerging theory, whether substantive or formal." This general statement was delineated in some detail in chapter 3 of DISCOVERY. In this chapter we add to this previous discussion both general ideas and specific procedures that have emerged as we and others have theoretically sampled during literally dozens of studies.

The general procedure of theoretical sampling, as we now shall describe it, is to elicit codes from raw data from the start of data collection through constant comparative analysis as the data pour in. Then to use the codes to direct further data collection, from which the codes are further theoretically developed with respect to their various properties and their connections with other codes until saturated. Theoretical sampling on any code ceases when it is saturated, elaborated and integrated into the emerging theory. This process produces cumulatively *intense theoretical sensitivity* into one's data as the integrative matrix of the emerging theory grows denser. The analyst progressively sees more in his data of relevance and fit as he explains how the basic social process resolves the basic social-psychological problems of the participants in the substantive area under study.

The explicit, consistent and persistent use of theoretical sampling is considerably different than those methods of generating theory employed by our colleagues whose work is most similar to our method of generating grounded theory. The initial decisions in theoretical sampling are based only on a general sociological perspective about a substantive area within a population, not on a preconceived problem or hypothsis. The researcher within the first days in the field already begins to generate codes, to emerge hypotheses and to integrate them. Thus much of his complex analysis is done *while* collecting data, not only after the data is finally collected. Theoretical

sampling is not simply a single line, unidirectional method of moving from less to more directed observation and from specific data to conceptual rendition. It also requires that the analyst engage in many other operations while accomplishing the latter along multiple lines and directions, and while going back and forth between data and concept as one generates theory. This book details at length those operations, such as theoretical coding, memoing, generating a process and problem, integrating and densifying.

Theoretical sampling is not what Leonard Schatzman has aptly termed "selective sampling," which is a frequently used sampling method in qualitative analysis.[1] Selective sampling refers to the calculated decision to sample a specific locale according to a preconceived but "reasonable" initial set of dimensions, (such as time, space, identity or power) which are worked out in advance for a study. The analyst who uses theoretical sampling cannot know in advance precisely what to sample for and where it will lead him. Only as he discovers codes and tries to saturate them by looking for comparison groups, does both what codes and their properties and where to collect data on them emerge. It is never clear cut for what and to where discovery will lead. It is ongoing.

We turn now to two sections. One is on explications of the logic of theoretical sampling with regard to inductive-deductive phasing of theoretical sampling; logical vs. conceptual elaborations of codes and the logic of comparability. In the second section we detail advances in procedures of collecting data with regard to the initial stages, to the question of where to go next and to the recording of data. We also discuss different kinds of comparisons, besides between collected data, such as experiential and literature comparisons to the emergent theory and we discuss the use of secondary data.

LOGIC OF THEORETICAL SAMPLING

Inductive-Deductive: Sociological and social psychological research typically involves alternating between inductive and deductive logic as the research proceeds. Maintaining the balance between the two logics and following their interplay is dictated by whether the research is initiated primarily as deductive (derived or preconceived) or as inductive. Grounded theory is, of course, inductive; a theory is induced or emerged after data collection starts. *Deductive work in grounded theory is used to derive from induced codes conceptual guides* as to where to go next for which com-

1. Leonard Schatzman and Anselm L. Strauss, *Field Research, Strategies For A Natural Sociology,* (Englewood Cliffs, NJ.: Prentice-Hall Inc., 1973).

parative group or subgroup, in order to sample for more data to generate the theory. We discuss this conceptual deductive logic below. Suffice as to say, deduction is in the service of further induction and the *source* of derivations are the codes generated from comparing data, *not* deductions from pre-existing theories in the extant literature. To repeat, the focus of deduction is on more comparisons for discovery, not on deriving an hypothesis for verification (though this may occur as a byproduct). "Strange" data are never a source of embarrassment in disproving a deduced hypothesis, but an excellent occasion for what they may contribute upon comparison to theoretical expansion, refinement and enrichment of the emerging theory.

Accordingly, the theory is rooted in data not an existing body of theory. Later as the generating continues, comparisons with extant theory may link it to a number of diverse theories which touch upon various aspects and levels of the emerging theory. This linkage, at minimum, can place the generated theory within a body of existing theories. More often, as we have said, it transcends part of it while integrating several extant theories. It may shed new perspectives and understandings on other theories and highlight their process. Other theories are neither proved or disproved, they are placed, extended and broadened. Comparisons with extant theories require generating analyses just as comparisons new data and emergent fit concepts.

This approach differs markedly from more conventional ways of handling the deductive, inductive relationship between theory and research. Perhaps more commonly, investigators logically *deduce* research hypotheses from a pre-existing abstract framework. Thus, with a logical approach, research hypotheses are created *before* the researcher has initiated investigation. The investigator may then feel compelled to find the information which is presupposed by the hypotheses that were logically derived. In this event, the researcher may become side tracked from relevancies by remaining committed to hypothesis in his conceived theoretical framework although the specific information for hypothesis testing may remain rather inaccessible or be relatively irrelevant. Commitment to pre-conceived hypotheses may limit the kinds of observations, information and insights that the researcher makes and actually may have access to. He finds himself subject to a nonstrategic, fixed and immovable research design. His theoretical sensitivity is thwarted.

When the strategies of theoretical sampling are employed, the researcher can make shifts of plan and emphasis early in the research process so that the data gathered reflects what is occurring in the field rather than speculation about what cannot or should have been observed. He can follow his emerging theoretical sensitivity.

The formulation of hypotheses in the discovery approach differs considerably from other types of social scientific research and theory construction. Research hypotheses in the generating approach are drawn from

emergent connections between the emerging coded and questions (deduced from the codes) which the researcher has about what is happening in the field of inquiry. When the grounded theory approach is used the researcher constructs his theoretical framework out of the data. Through comparing the data as it is collected, the researcher *creates* more abstract levels of theoretical connections. In short, theory is gradually built up inductively from the *progressive stages* of analysis of the data.

Questions may be raised at this point concerning how during sampling the emerging theoretical sensitivity can be reconciled with the theoretical biases, proclivities and premises of the individual investigator. This issue brings up some significant points concerning the methodological orientation of the researcher. Most generally, the background experiences of one's education and training is used to *sensitize* the researcher to address certain kinds of broad questions. Thus, one's methodological orientation is broadly, rather than narrowly, based in one's scholarly discipline. Its sensitizing nature lacks specification of attributes, but forms guidelines and reference points which the researcher uses to deductively formulate questions which may then elicit data that leads to inductive concepts being formulated later.

But, most importantly, the sensitizing concept is not simply verified through the research process. Instead, it is used to uncover *data* that otherwise might be overlooked. The use of sensitizing concepts is at the beginning of the research, rather than becoming the whole of the analysis. Then, formulation of definitive concepts occurs *after* the researcher has taken the data through analytic phases of conceptualization.

When the grounded theory approach is used, relationships between theory and methods become tightly integrated. Empirical relationships are created between the theory that is being constructed, the data it explains, and the inductive process by which the former was generated from the latter.

Theoretical sampling is, then, used as a way of checking on the emerging conceptual framework rather than being used for the verification of preconceived hypotheses. Comparisons are made continually between kinds of information to generate qualifying conditions, not disprove hypotheses. While in the field, the researcher continually asks questions as to fit, relevance and workability about the emerging categories and relationships between them. By raising questions at this point in time the researcher checks those issues while he still has access to the data. As a result, he continually fits his analysis to the data by checking as he proceeds.

Importantly then, the researcher continually *analyzes* while still in the midst of the social world or circumstances he is studying. Thus, unlike other field workers, his descriptive notes and questions may bring out important categories early in the research. Rather than keeping a somewhat behavioristic account to be analyzed after the field experience is closed the analyst attempts to reflect upon what he is experiencing and begins to code and

analyze from the start of the research. This early effort literally forces theoretical sampling deductions, which sustains the development of analysis and facilitates field inquiry around *theoretical questions.* In this way, the amount of time for *data-collection* can perhaps be shortened without sacrificing the richness of the data, which sacrifice, paradoxically, can occur through dilution when the researcher gathers endless voluminous notes unfrettered by theoretical sampling. This approach differs considerably from those qualitative methodologists who stress keeping the field notes behavioristic devoid of coding and analyzing so as to avoid contamination of the data.[2] But, the field work itself becomes stronger and more complex when analysis coincides with the written observations, that is, under the condition that the analysis corresponds to the empirical reality at hand.

Conceptual vs. Logical Elaboration: Deductive elaborating is vital to the theoretical sampling phase of grounded theory. This we call conceptual elaboration as opposed to the logical elaboration founded in deductive, hypothesis testing research. Conceptual elaboration during theoretical sampling is the systematic deduction from the emerging theory of the theoretical possibilities and probabilities for elaborating the theory as to explanations and interpretations. These become hypotheses which guide the researcher back to locations and comparative groups in the field to discover more ideas and connections from data. These directing hypotheses are proven only in the sense that they are useful for relevant theoretical elaboration of the emerging theory. The data constantly check deductions that lead nowhere, as the analyst takes his directions from emerging relevancies. These deducted hypotheses are not forced on the data when they fit poorly, they are discarded and others emerge in their place by constant comparative analysis. This method underlines and assumes the fact that the interpretations of patterns must be researched and grounded just as much as the patterns themselves. Interpretations cannot be justifiably borrowed from extant theory.

In contrast, logical deductions are a re-entry of the primarily deductive approach after a bit of grounded theory making. The subtle switch back to the deductive approach occurs because the analyst flashes on an extant theory that seems to explain or interpret what is going on. Almost automatically he starts deriving hypotheses from extant theory that makes sense in explaining the inductive theory. For example, one can borrow interpretatations for a delinquent pattern from deviant subculture theory.

Analysts working under this tacit assumption of borrowing interpretations or logical elaborations (as most of us do) are tempted to use logical

2. John Lofland, *Analyzing Social Settings,* (Belmont, California—Wadsworth Publishing Co., Inc. 1971). Chapters 5 and 6.

possibilities that exceed the bounds of the data in an attempt to fulfill the exigencies of an explanatory scheme. And unfortunately, this forced logic can go unrecognized because it also derives its legitimization from a grounded base (i.e. the grounded data at hand). While conceptual elaboration is necessary and leads to density within an analysis, it can also become countervailing when it shades into logical elaboration by leading the researcher away from the grounded nature of his data into the realm of verifying what is "supposed" to be and not theoretically sampling for what "could be or is."

In order to circumvent such a predicament the analyst must impose analytical limitations or focus within his memo development. This amounts to disavowing allegiance with logic-deductive analysis (as much as possible) and only allowing himself to extend his analysis within the conceptual bounds of the data (i.e. to the conceptual possibilities called forth within the framework of the data).

To repeat, elaborations of the emerging theory must be researched. The best that can occur is *emergent fits* to the logical derivations. But more often several dangers to avoid, when using and grounding deductions from extant theory, become reality. The analyst feels so good about them that he is saturated and researches no more. It is a standard, easy and legitimate way out of the hard task of generating. Then we never know the true grounding or verification of these "neat" interpretations.

When research occurs that follows exclusively the dictates of these deductions from other theory, relevance is easily lost. These dictates usually derail the analyst from studying the emerging basic social psychological problem and basic social process as he pursues "pet" ideas of others. This loss occurs in trying to force the data in a direction that is not appropriate and yields either no good or unmeaningful data or a useless verification of an aside theory. The continual reappearance of the core problem and variables is stifled, the central analytic framework splattered, not integrated. In logical elaboration, the researcher goes to another world (often by arm chair) to explain what is happening in the world wherein lies the data. Grounded theory requires going back to the field for data *for and by* conceptual elaboration.

To summarily discard deductive elaborating as a tool for discovery is clearly unwise. It is vital in the constant comparative analysis of data for generating theory. To write about the subtle interplay of deductive and inductive generating is not easy. The analyst need only remember that the deductive is in the service of an inductive method; it is subserviant to it, and ideas arrived at deductively must be discarded unless grounded. And they usually are discarded with ease, since ideas arrived at inductively flood in and push the deductive out with their ability to fit and work and be relevant. It is only the doctrinaire researcher who may not allow the inductive to correct deductions and emerge as primary, and who will explain "poor fit" with yet another pre-existing theory.

Logic of Comparability of Groups: "The basic question in theoretical sampling is: what groups or subgroups does one turn to next in data collection? Our criteria are those of theoretical purpose and relevance not of structural circumstance or of preconceived preconception."[3] Groups are chosen as they are needed rather than before the research begins. Theoretical purpose, as a criterion, is embodied in generated *ideas* or ideas deduced from them. These ideas are properties of groups *not* the group itself or its most apparent description. Thus apples can be compared to oranges if the comparison is the kinds of vitamins found beneath and in the skin. The premature baby ward can be compared with a cancer ward on the criterion of awareness.

Apparent non-comparability is often seen by people, unaccustomed to grounded theory, as a problem. "We all agree that you must not compare incomparables," a pretentious colleague wrote us. *Actually apparent non-comparability is irrelevant, if the variable to be compared has a value in each group.* Comparing on the basis of properties of groups has the purpose of generating theory. In the bargain if important differences (or indeed, similarities too) also exist between the groups, they become part of the analysis as qualifying conditions of the general hypotheses or idea that emerged from the initial comparison. These differences do not disqualify a comparison, they enrich it. Comparing the apparently non-comparable increases the broad range of groups and ideas available to the emerging theory. This especially is important in sampling for formal theory generation (see chapter nine). One can compare a political speech, to a doctor-patient exchange and to a milkman's approach to a housewife all on the idea of cultivating others for an instrumental purpose.

The researcher must always bear in mind that groups are, from a theoretical viewpoint, clusters of variables, which are not all readily apparent, and many of which are to be discovered by comparison. Thus to compare on one or two initial variables easily brings into relief other properties of the group that find their way into the emerging theory. Comparisons then are ruled out based only on a dearth of generation after trying, *not* based on an apparent "non-comparable" difference.

Further, these ideational property comparisons may even seem more non-comparable when they are based on an interchangeability of indicators.[4] Any concept is indicated by what may be called a reasonable set of indicators which therefore may be seen as interchangeable. The social value of a patient may have indicators, some readily seen, others learned after time. Thus a researcher may compare the treatment of patients according to their

3. *Discovery,* Ibid, chapter 3.
4. Paul F. Lazarsfeld and Wagner Thielens, Jr., *The Academic Mind,* (Glencoe, Ill: The Free Press, 1958) pages 403-6.

social value by looking at the treatment of poor blacks in an emergency room, of a young mother in an intensive care unit, of a derelict alcoholic from skid row who later proves to be a dentist on a cancer ward, of a distinguished politician in a private room, in order to generate hypotheses about the calculation of social value as a source of the attention dimension of hospital, medical and patient care. Thus while indicators are interchangeable in indicating a concept, they also bring out different values of the ideational variable, greatly enriching the emerging theory.

The interchangeability of indicators allows for different researchers to collect different indicators on the same concept and agree that each others' indicators are "all right" for the concept. They may be part of the same study or of different studies. All that is required is a broad concensus of what may be included in the reasonable set of indicators for the concept. And then further they need merely collect ones like them. Thus ethnic indicators help establish social values, and there are many. If they disagree on the indicators, they usually can comparatively analyze them to generate the different dimensions of the concept. Thus diarrhea and perfume both indicate body pollution. The former in unfavorable and the latter favorable. The former is avoided and the latter is sought. Interchangeability brings out enriching differences on the same idea. Apparently too diverse a pair of indicators should not occasion one to be discarded, until thoroughly checked by comparative analysis. It is important to remember that the analyst is collecting facts as indicators to be compared and coded into ideas, he is not collecting facts to be rendered empirically in descriptions.

By the same token, the interchangeability applies to the same indicator indicating more than one concept. So different analysts may see a different concept in the same datum. Thus a nurse saying that a patient may expect to die in two days, can indicate to one analyst a level of awareness and to another a stage in the temporal passage (tragectory) toward death. Each indicator therefore can have more than one meaning. The point is, always, to achieve the cogency of indicator-concept meaning so others can see (and judge) it for themselves, that is how well the concept fits and work using the indicator.

We shall gain more insight into the logic of comparability in the next chapter where we discuss at length the coding relationship between indicators and concepts, which concepts we refer to as categories and their properties.

Suffice as to say that for grounded theory, comparability is ideational and analytic and that seeming non-comparability is not to the point. Similarities and differences (on various dimensions) are to be comparatively analyzed into categories and their properties and ordered by theoretical codes (see Chapter five). We can compare anything by comparing diverse indicators of various properties of the same category. Many people are captured by the logic of the comparative schema found in comparing experimental groups

by controlling, holding constant, or muting "significant conditions" while a prime variable is absent in one group and present in another. Or they are captured by the comparative schema used in surveys of comparing groups on the same dimension for their differences, for which selective sampling is used.

The logic of comparison is more general than these specific models. Comparison is basic to science. No one model is the only one. Our approach is to develop another model of comparison. One that has complete variable flexibility as detailed above, using the interchangeability of indicators as they generate concepts upon comparison. Our approach seems to work well in dealing with an unanticipatable set of natural groups that cannot be controlled.

Further, the groups need not be clearly defined or membered. We are not comparing populations, we are comparing ideational characteristics of groups that in turn delineate behavorial and attitudinal patterns. Indeed, one category of a group may well be its vague boundaries and membership. Amorphous groups are just as useful.

COLLECTING DATA

Initial Decisions: As we said in *Discovery,* initial decisions for theoretical sampling are based only on a sociological perspective and on a general problem area, such as what happens to students in a medical school that turns them into doctors, or how do milkmen keep housewives from cancelling their service and using super markets or how do lawyers develop a solo practice. These decisions are not based on a preconceived framework of concepts and hypotheses. We also said that the researcher should also be sufficiently theoretically sensitive—by training—so he has the tools within him to self consciously conceptualize and formulate a theory as it emerges from the data.

Since that time we have discovered this position can go both ways in degree of openness. The analyst can in most cases enter the field with complete openness. He can go anywhere and talk and listen to anyone and read anything with virtually no problem in mind and little training in a perspective, *provided he is capable of conceptualization.* He can do this because relevant problems and processes quickly emerge—almost too fast—sufficiently enough to start theoretically sampling for the emerging theory. False starts, and starts which are close but not quite central, soon become corrected by the constant comparisons in theoretically sampling. Indeed, the analyst with this complete openness is often more receptive to the emergent than others with a few pre-ideas and perspectives. He has less ideational baggage to give up or correct. The emergence of concepts never fails, as it cannot, since

social organization of life is always in process of resolving relevant problems for the participants in an action scene. And this social organization always provides a partial framework of "local" concepts to be used by participants in designating the principal structural, processual and interactional features on the action under study.

For example one simply knows before beginning the study of a ward in a hospital that there will be doctors, nurses, ancillary personnel, shifts, admittance procedures and so forth. Also yet unknown "local" concepts—and there are always at least a few or more—emerge quite quickly in the first day or two of listening, reading and watching. These beginning codes, no matter how conceptually primitive, quickly start theoretical sampling and constant comparisons of incidents. How relevant these concepts are to the basic problem and basic social process becomes a question of futher analysis. The *researcher in this mode does not have to know beforehand,* he has to believe his data and theoretically sample for it.

Other researchers, usually those with training of some duration, find it more comfortable to enter the field with some combination of a clear question or problem area in mind, a general perspective, and a supply of beginning concepts and field research strategies. This is less than being completely open, but still quite receptive to the emergent. We have mentioned some general questions above. The perspectives that many analysts prefer to use are limiting but still quite general. One may wish to look at the activity of nurses as a problem in work, or focus on temporal dimensions of patient care, or view interactions related to levels of awareness, or structural change and so forth. Insofar as a chosen general perspective relates to the basic problem and social process, they are fine. Seldom do they constrain or derail the emerging analysis, but if they do, they must be discarded. For example most action or interaction has relevant temporal aspects related to the core codes.

Several analysts start their field work with initial strategies. They will attend scheduled regular or special events hoping to observe also surprise events. Based on "local" concepts, they will go to the groups which they believe will *maximize the possibilities* of obtaining data and leads for more data on their question. Thus it was a natural to start the dying study on an intensive care unit and a cancer ward. They will also begin by talking with the most knowledgeable people to get a line on relevancies and leads to track down more data and where and how to locate oneself for a rich supply of data. Another way to begin is to ask a few knowledgeable participants to write a diary or a case history for an event, such as an AA meeting. Then the researcher can begin by analyzing this material.

We now also suggest that if an existing theory seems quite grounded in data, one can possibly begin with it. But in using this theory, as many are now using grounded theories, one must be cautious. The analyst must

theoretically sample to establish its emergent fit and to prevent its derailing the analyst from achieving maximum relevance. Just because concepts of the theory fit, they do not necessarily capture relevance, nor should they be allowed to steer the study away from relevance in favor of the relevance in the others studies from which the concept is taken. For example there are many studies on careers, each with different relevancies, but using many of the same concepts.[5] This use of existing theory concepts should not also preempt the emergence of concepts, in general and specifically if one works better than career in studying temporal aspects of "dying."

At first the analyst may feel that his non-preconceived field work yields only scattered observation. But as soon as he starts to comparatively analyze the data—preferably the first day—codes emerge yielding theoretical leads and theoretical sampling is off to a start. Researchers often fear it will not happen, before entering the field. But within a few days in the field they quite often are barraged with "core" concepts and feel "I can write the report already." Most frequently the problem is to slow down the conceptualization while theoretical sampling corrects for fit, work and relevance. So much that is apparently crucial in the beginning gets left behind later in the study. Thus the issue emerges after these initial decisions: where to go next for comparisons for theoretical elaboration.

Where Next: The answer is that the directions to sampling are arrived at deductively, as mentioned above, from what is induced by comparisons. The problems on "where next" are compounded by the *multiple options* for the comparative groups that emerge as the analyst is barraged with data and ideas for codes to be saturated and checked. For example, as the analyst picks up on the awareness of a dying patient, there are many directions in which he can start sampling: levels of awareness, ingredients of awareness, interactions over differentials in awareness and so on. He can only systematically start sampling in groups that give him data on each possible direction, until sufficient elaboration and saturation are achieved with relevance and fit.

In the beginning he starts with *opening coding* which leads him to sample in all directions which seem relevant and work. Later on when the researcher discovers his core variables—the basic social problem and process—his sampling becomes *selective* along the lines of his focus on the central issues of his emerging theory.

Systematically starting to theoretically sample in conceptually elaborated directions, written in memos, in order to saturate codes, exerts control over the tremendous momentum for data collection that soon starts after initial forays into the data. Since theoretical sampling requires joint collection,

5. Barney G. Glaser, *Organization Careers: a Sourcebook For Theory,* (Chicago: Aldine Publishing Co., 1968).

coding and analyzing, the analyst is forced to curb his zest for data collection by trying from the start to code and analyze. And it is a struggle to keep up this simultaneous activity because the strong tendency is to forget coding and analyzing and pile up interesting data, no matter what direction it takes.

It is vital to avoid burning up one's energy in data collection, leaving none for coding and analyzing. Thus our fellow analysts talk always of leaving (or tearing themselves away from) the field after 3 to 4 hours and going home to record field notes, code them and start analyzing them to get direction for the next day. If they stay too long in the field then they may have to devote the whole next day to coding and analyzing, or data collection will run away with them, and they will not be theoretically clear where.

Each analyst of course, develops his own pacing recipe according to his current temperament and the contingencies of the study. But in all cases this extends the time in the field—while cutting the field work—compared to the more traditional approach of collecting all data first and then analyzing it. The grounded theorist is likely to be in the field longer but collects less data since collection is controlled and directed to relevance and workability by theoretically sampling for the emerging theory. He manages to advantage himself the most of the momentum before it is lost; and along the way, he has minimized the loss of emergent ideas by memoing on the unfolding analytic elaborations.

Besides his theoretical pacing the analyst soon develops ongoing strategies to handle two other problems spawned by the momentum of theoretically sampling: 1) Staying open and 2) keeping in mind the current categories. Staying open refers to both open coding until he is sure he has reached relevancy and saturation. And even when more focus is called for by still staying open to continual input of new data that can modify current "grabbing" ideas and force analysis of exceptions.

One strategy is to constantly change his interviewing style, place and his interviewees in order to keep following up new ideas. Another is to continually jot down new ideas in the field separate from and even before data recording. Another is to note relevance by its constant or patterned reoccurence in informant's discussions and stories, and to note saturation by the end of discovering new properties on a category. Relevance can also be checked and elaborated by asking top informants to appraise and give more data on categories proving to become core to the analysis, which provides ongoing sampling, verification and stabilization of focus. For example after much field work and analysis on Intensive Care Units we went over our work with two Head Nurses who had spent considerable years in ICU work.

Grounded theory's ready modifiability allows openness to correction and change in emerging theory, and theoretically sampling for such, since there are no "pet" hypotheses. There is only trying to discover what categories and their inter-relations fit and work best. Whatever brings the analyst

closer to this *is the goal,* not proving anything or beknighting the person who deducted a precious hypothesis.

In responding openingly to the data the analyst allows people to go on and on—being open to all—especially at the beginning stages. Patterns occur in both what is and is not said. However, to expect something that is not said or pointedly left unsaid is to accept it as *probably* not important. To study risk taking among steeple jacks and never hear them refer to it, forces the theoretically sampling analyst to accept that it is not relevant enough to their work. They are focused on other social psychological issues.

The momentum of sampling itself breeds a forgetfulness about what one is sampling for theoretically, as new data barrages the analyst and inundates his thought processes with yet new leads and directions. This is not a lasting problem. Few of us completely remember our original effort at category development when our openness to respond to data, derails us. It is hard to remember every category especially when one is a participant observer absorbed with the care, concern and fatefullness about what is going on. But it is no lasting problem because the analyst starts hearing and seeing in terms of emerging categories; and later on, when recording, coding and analyzing, he discovers his data are imbued through association recall as the forgotten becomes relevant with what has previously emerged.

The analyst's natural inclination is to sample closely for what he does not know, based on what he knows. Theory generation starts developing a perspective by which the analyst more and more sees his data and the concepts that emerge. He becomes trapped by his own emerging theoretical rhetoric. Thus an analyst is usually only good for one study in a given substantive field with a given focus. The analyst's general perspective in the beginning becomes both altered and actualized in a substantive perspective that emerges. This guides the researcher quite well, when general theoretical directions lapse.

Other ongoing strategies that come with experience are to keep grounding deductive ideas by questions as they emerge in conversations and interviews, to keep the direction of the ideas and questions going toward *researched* interpretations, and explanations and to keep trying to get behind what emerges with underlying scope and parsimony of conceptualization (a major prime goal of theory making). The probing questions of the theoretical researcher achieves this. This continuous probing behind emergents and grounding them achieves saturation and densification of properties of categories yielding theoretical coverage (two other major goals of theory making). Constantly noting properties of contexts and conditions as one observes and interviews within these parameters also comes with experience.

Norman Denzin has discussed at length the strategy of Triangulation in theoretical sampling.[6] He refers to how to bracket in the properties of an idea by getting different vantage points on it from diverse people. Pat Mullen in her work with students has developed group theoretical sampling.

> "One strategy for theoretical sampling is to have a number of people report in a group session the results of interviews or observations over a variety of situations. For example, in a study of invisible physical impairments, interviewers were brought together after each has seen someone with an impairment such as diabetes, ileostomy, colostomy, mastectomy, back problem, hearing impairment, heart problem, special diet need, etc. Instant comparisons were made as the interviewers "coded" their interviews aloud, leading to the generation of significant and dense memo's on, for instance, the dimensions of impairments (e.g., social intrusiveness, clash with role expectations, stability, stigma, credibility), dimensions of disclosing (simple telling, educating, reminding, convincing), strategies for disclosing, and other important aspects of the study. New directions were set immediately, and the individuals could proceed with coding and memo writing with a fuller picture than they would have had working alone, because of the conceptual sensitivity which was excited by the group exchange.
>
> Subsequent meetings in which open and theoretical coding and memo's were shared speeded and enriched the analytical process considerably.
>
> This group focus upon a single study is also a valuable teaching tool which makes possible a great deal of progress in a 3 month seminar with a minimum of time devoted to data collection (only 2 interviews are needed from each person) and maximal time for analysis."[7]

Types of Comparisons: Lastly in this chapter I shall discuss how the question "where to go next" poses additional questions of what to compare and what types of comparison and groups to sample for.

In DISCOVERY, the chapter on the constant comparative method detailed the process of comparisons involved in generating theory and in sampling to further emerge the theory. *First, the analyst compares incident to incident* with the purpose of establishing the underlining uniformity and its varying conditions. Both the uniformity and the conditions become the generated concepts and hypotheses. Thus when the analyst observes several nurses and doctors hovering over a VIP on an intensive care unit and he observes the ignoring of a lower class black on emergency ward with a gaping knife wound, he may generate the concepts social value of a patient and medical attention and the hypotheses that the higher the social value the

---

6. Norman K. Denzin, *Sociological Methods: A Sourcesbook,* (Chicago: Aldine Publishing Co., 1970).
7. Correspondence from Patricia D. Mullen, Professor of Public Health, University of Washington, Seattle.

more and quicker the medical attention under the condition of immediate need of care.

Then as the analyst continues to code and compare he compares *secondly, the concept to more incidents* generating new theoretical properties of the concept and more hypotheses. Thus when the medical team discover that the dentist in emergency is actually a derelict alcoholic who has not practiced for years there attention and effort may shift to others. Social value is both apparent and re-calculated as further aspects are learned about the the patient. This comparison of concept to further incidents has the purpose of theoretical elaboration, saturation, and verification of the concepts, densification of the concepts by developing their properties, and the generation of further concepts.

Lastly, while the first and second type of comparisons continued throughout the study, the analyst also compares *thirdly concept to concept* with the purpose of establishing the best fit of many choice of concepts to a set of indicators, the conceptual levels between concepts that refer to the same set of indicators and the integration into hypotheses between the concepts, which becomes the theory. Thus the analyst may have generated the concept of "social value" of patients when looking at a full spectrum of patients who are recovering and dying and the concept of "social loss" when just looking at dying. Social value is on a higher level of generality than social loss, since it implies the latter as its scope encompasses all patients, including dying patients. For dying patients either concept may work well but social loss fits better into the imagery of the theory.

The reader can see by these cursory examples how quickly theoretical sensitivity develops to one's data and how deductively it leads to theoretical sampling for new aspects and groups to further generate the theory. The analyst begins to think of many other kinds of incidents and groups to sample for within his substantive area and often outside his area. This in and out of substantive area is a fundamental distinction to keep in mind when theoretically sampling.

An important rule is, when the analyst is still "young" at generating skills, as a sociologist and/or as a scholar within the area, to sample exclusively within the substantive area until focus on a basic social psychological problem and the process by which it is resolved *both* have been discovered and stabilized in an emerging theoretical framework. Once this has occurred then it is safe from undermining of relevance of core process to sample outside the substantive area of study to further elaborate the emerging theory. Thus the analyst an look at the variation of how attention is distributed according to social value in other areas besides patient care, to help develop the theory for patients.

If the analyst goes outside the bounds of his substantive area too soon, several dangers emerge. It can easily kill or dilute the emergence of the basic

social process in the area under study, by comparing it to areas where the process is less relevant. Both milkmen and doctors cultivate their clients but in the former it is more basic. This derailment may forever lose the core variable, since the sensitivity to it is diluted. It leaves the less than polished skilled analyst somewhat "splattered" and confused. This applies especially to students and new Ph.D's. We hear statements to the effect that "I do not know which way to go;" "Nothing seems most relevant;" "I'm confused," "I have too much, and it goes nowhere." Whereas sticking within the area until focus is stabilized, presents the analyst with a challenge he is forced to face and resolve: he must make something theoretically of his substantive data.

Once it is safe to go outside the data for comparisons with other data he may even take on kinds of incidents that are not arrived at through systematic research, but can be helpful. This class of comparisons is called *experiential incidents.* This refers to anecdotes and stories, which are given by the analyst himself or by others, that seem to compare to the data. Mention a hospital and everyone has an experience to tell and often cannot be stopped until they do. In these tales overstatements and omission abound and cannot be checked out but they are fruitful for developing sensitivity to what to sample for.

The prominent class of outside comparisons is the literature. It depends on the literature; some comparisons are theoretical concept to concept and others are concept to data. This, of course, integrates both the substantive ethnographic data and theoretical literature to the research study by locating both. It allows for new emergent categories. It extends the theory referred to. At this point the power of most theoretical literature, if not all, to derail the analyst toward the less than relevant and poor fits is lessened as he reads a "neat" work on a problem. This is why reading the theoretical literature should be avoided when possible until after the discovered framework is stabilized.

How necessary it is for an analyst to go outside the substantive area is debatible. It is usually stimulating for theoretical sampling, but still dangerous for its undermining effects on relevance. Going outside an area for generating formal theory is of course a must, it is the way it is done: comparing a wide range of diverse substantive areas to increase the level of generality (See Chapter 9).

This general warning against sampling outside the substantive area before an emergent theoretical framework is stabilized cannot be heeded too closely and carefully. Besides undermining the relevance of the substantive area, the literature's focus frequently becomes a "pet" interest of the analyst because of its respected author. So what occurs is a substantive theory about an area with dubious relevance and the implicit but not fulfilled effort to do a formal theory on a concept. But the formal theory itself has not

occurred for three reasons; it is written as a theory of a substantive area, it is written up as if general but it is not; and the wide range of systematic comparisons required to generate a formal theory have not been made. It is just interesting but nowhere.

There are various cycles of going outside the substantive area. To this point we have discussed substantive theory and formal theory. The former is about a specific area e.g. heart disease or route milkmen, the later about a concept in its full generality: e.g. cultivating. There is a type of theory in between which we call *general substantive theory,* it is more general than a substantive theory but not completely general as a formal theory. This is a theory about a general substantive area such as chronic illness, or organizational careers.[8] Each case is not a specific one and each case is not as general as the chronic conditions of everyday life or careers of any nature.

To sample for a general substantive theory does not require a framework based on a specific substantive area. It can be generated from comparisons from different cases within the general area. Thus to write a general substantive theory on chronic illness one can theoretically sample as many chronic illnesses as possible to generate a theory of say: redesigning lifestyles. The analyst is going outside his area only insofar as he had set himself to study one chronic illness. To do a general study of the area is to expand his focus on diverse substantive units, not to stimulate a formal focus on the concept with outside data.

Some Practical Issues of Collecting Data

In closing this chapter we wish to suggest some strategies for handling three practical problems of theoretical sampling; recording data, theoretical saturation and secondary analysis.

It is hard enough to keep the theoretical ideas in one's head much less the data, when it is not possible to record in the field, as one is actually interviewing and/or observing. The general approach, of course, is to dictate or write the details of the interview as soon as possible after leaving the field. We have found, that a good strategy to recapture most of the details later, is to jot down after leaving the field, the ideas (and perhaps an illustration) the analyst was collecting data on, even if he was only implicitly sampling on the idea. So much theoretical sampling is on a recall basis when under the strain of the field. Grounded theory has this implicit guiding function both for collecting and later for recalling all the details seemingly forgotten.

---

8. Glaser, *Organizational Careers,* ibid, and Anselm L. Strauss and Barney G. Glaser, *Chronic Illness and the Quality of Life,* (St. Louis, Mo: CV Mosby Co, 1975).

A few days later the analyst can refer to his "jots" which lead to a recall of the interview. His recall gets better and better as the generated theory develops and he gets more "into" the analyst. By this strategy the analyst records details from which he will generate properties of his categories and they inevitably spill over into details that will yield related categories. In sum, both in collecting and recording data, the analyst need not be conscious of all ideas and operations as this chapter might imply, as they imbue his work irregardless—once the emerging theory takes hold. This strategy is especially good when at the end of the day the analyst cannot or is too tired to record notes.

One problem, that emerges in theoretical sampling, is the feeling of theoretical saturation when in the field (for more see Chapter 7). Analysts soon discover that this is often a false sense when analyzing their current data in the privacy of their studies. It is false because it is premature in terms of the scope of the study. It usually is the saturation of the current interview situation, plus fatigue. As more data comes in from the myriad of other situations, new incidents emerge which lead to other properties of the category. Theoretical saturation of a category occurs when in coding and analyzing both no new properties emerge and the same properties continually emerge as one goes through the full extent of the data. Thus when one is in the field and feels he has saturated a category in one situation, he probably has, and is not slighting it to go on to sample for incidents on other categories. He will naturally come back to theoretically unsaturated categories in new field situations.

A major option for grounded theorists lies in secondary analysis.[9] Every analyst we have known has considerably more time, temperament and ability to do more grounded analyses of data than he could ever collect for in the field. This applies especially to teachers. This can be coupled with the fact that there are more and more large studies being done at great expense which focus on collecting masses of data that go far beyond the preconceived minute analytic purposes of these projects. These studies are the products of empirically oriented funding agencies who hire "fact producers," who have little or no trained ability to analyze data to any theoretical extent.

The previous definition of secondary analysis was the analysis of previously collected data for a purpose other than that for which it was collected. Now, I see the definition being altered to the theoretical analysis of previously collected data for any purpose, depending on what emerges. Secondary data can be analyzed for the original purpose as well as other purposes since we find in these studies, which produce huge amounts of data, that even the original purpose is under or slightly analyzed.

9. Barney G. Glaser, "The Use of Secondary Analysis by the Independent Researcher", *The American Behavioral Scientist*, (1963), pp. 11-14.

Indeed, we and our students are constantly asked to consult for the purpose of analyzing the data sufficiently to get the inhouse researchers going on an analytic tack. This has occurred a number of times, for in the end, the project directors "want something done" with the data. They like the grounded approach, not simply because we and our students are trained to analyze, but because of our response to the data *as it is* and as the theory emerges.[10] These large projects usually have the funds to hire analysts as consultants. They are quite often delighted to give or lend their data to someone who will do something with it.

The grounded theorist simply theoretically samples the data that has been obtained, by "appreciating what he has, not what the project did not collect." Infrequently he can ask project personnel to go back into the field. It is important to be sure that enough data has been collected with which to theoretically sample, without running into its limits too quickly. The analyst, however, can always suggest what data could be collected in future studies when his theoretical sampling breaches these limits for certain categories.

The quality of the data collected by others, so often a concern in this type work, is for the grounded theorist just more variables for his analysis. He must keep in mind that the accuracy of the facts is not so "crucial" (as we said over and over again in DISCOVERY) since they are used to generate concepts for a theory. And the resulting categories, hypotheses and processes can in turn be tested through further comparative analysis of the data, and through future studies, which point out modifications of the theory. The theoretical distillate of secondary analysis is the important thing, not the inevitable distortions coming from analyzing data one did not collect.

Now let us turn to our advances on the nature of coding, which of course, is intimately linked with the theoretical sampling process.

---

10. Jack Mezirow et al, *The Last Gamble on Education,* (Washington D.C.; Adult Education Association of the U.S.A. 1975).