CHAPTER 4

THEORETICAL CODING

The essential relationship between data and theory is a conceptual code. The code conceptualizes the underlying pattern of a set of empirical indicators within the data. Thus, in generating a theory by developing the hypothetical relationships between conceptual codes (categories and their propertes) which have been generated from the data as indicators, we "discover" a grounded theory.

Coding gets the analyst off the empirical level by fracturing the data, then conceptually grouping it into codes that then become the theory which explains what is happening in the data. Coding for conceptual ideas is a sure way to free analysts from the empirical bond of the data. It allows the researcher to transcend the empirical nature of the data—which is so easy to get lost in—while at the same time conceptually accounting for the processes within the data in a theoretically sensitive way. The code gives the researcher a condensed, abstract view with scope of the data that includes otherwise seemingly disparate phenomenon. This conceptual scope transcends the empirical arguments often surrounding disparate data on the same level.

Since the code is of central importance in the generating of theory, we will discuss now at some length the nature and generation of ideational codes—the building block of theory. There are basically two types of codes to generate: substantive and theoretical. Substantive codes conceptualize the empirical substance of the area of research. Theoretical codes conceptualize how the substantive codes may relate to each other as hypotheses to be integrated into the theory. For example, two substantive codes generated for an intensive care unit in a Hospital are social loss and attention. They may be theoretically coded into an hypothesis as a cause based on degree. Thus, the higher the social loss, the more the attention received by nurses.[1]

In the first part of this chapter we shall deal with substantive coding by beginning with opening coding, then selective coding and then detailing

1. Glaser and Strauss, "The Social Loss of Dying Patients," American Journal of Nursing, 64 (June, 1964) pp 119-21

some of the process and mechanics of generating codes. In the second part of this chapter we then discuss the nature of theoretical coding and give several lists of theoretical coding families. The two types of coding most often go on simultaneously, and this should be brought out in memos. But the analyst will focus relatively more on substantive coding when discovering codes within the data, and more on theoretical coding when theoretically sorting and integrating his memos.

SUBSTANTIVE CODING

Open Coding: The goal of the analyst is to generate an emergent set of categories and their properties which fit, work and are relevant for integrating into a theory. To achieve this goal the analyst begins with open coding, which is coding the data in everyway possible. Another way to phrase it is "running the data open."

The analyst codes for as many categories that might fit; he codes different incidences into as many categories as possible. New categories emerge and new incidences fit existing categories. He may even code for what is not obviously stated. This maximizes allowing the best fits, the most workable ones and the core relevancies to emerge on their own. It minimizes whatever "trip", whether from preconceived thought or from latent training, may be laid on the data that can force or distort its meaning.

Open coding is diametrically contrasted with a preconceived code for which the data may be coded, irrespective of degree of relevance, which itself is hard to determine. The preconceived code may be either substantive (e.g. legitimizing) or theoretical (looking for consequences), which thereby preconceives the relevance of core processes and action of the data, and in large measure buries the chance of its emergence.

Open coding allows the analyst to see the direction in which to take his study by theoretical sampling, before he becomes selective and focused on a particular problem. Thus, when he does focus, he is sure of relevance. It begins with the fracturing of data into analytic pieces which can then be raised to conceptual level. It begins to teach the analyst the kind of categories which can handle the data theoretically, so that he knows how to code *all* data. If all data cannot be coded, the emerging theory does not fully fit or work for the data and must be modified. Full theoretical coverage includes all relevant data as indicators in its framework.

Open coding with many theoretical memos and much saturation of categories gets one "out of" his data. This is vital if the analyst is intimate with the data from collecting it—evan after a few days of fieldwork. Open coding forces him to think and transcend his involved empirical view of his field notes. Thus, it is always much easier to code someone else's data, because of the de facto distance from it. Open coding allows the analyst the full range

of his theoretical sensitivity as it allows him to take chances on trying to generate codes that might fit and work. It is in the beginning with open coding—and a minimum of preconception—that the analyst is most tested as to his trust in himself, in the grounded method and in his skill to use the method and as to his ability to generate codes and find relevance. Many of our students suffer the initial anxiety that, in fact, nothing will emerge because they doubt their skill. They soon discover the opposite. Constant comparisons literally force generation of codes.[2]

Several rules govern open coding, which tends to insure its proper use and success. The first rule is to ask a set of questions of the data which must be kept in mind from the start. The most general question is *"What is this data a study of?"* While allowing complete emergence, this question continually reminds the researcher that his original intents on what he thought he was going to study just might not be; and in our experience it usually is not.

The next vital question to continually ask when studying field notes is, *"What category does this incident indicate?"* This is the short form. The long form is, "What category or property of a category, of what part of the emerging theory, does this incident indicate?" As the theory becomes more and more formulated this question becomes easier to answer. The continual asking of this question keeps the analyst from getting lost in the re-experiencing of his data by forcing him to try and generate codes that *relate* to other codes. It forces coding that earns its way into the theory by its grounding in the data.

Lastly the analyst asks continually: *"What is actually happening in the data?"* What is the basic social psychological problem(s) faced by the participants in the action scene? What is the basic social psychological process or social structural process *that processes the problem* to make life viable in the action scene? What accounts for the basic problem and process?" These questions (See Chapter 6) and more that arise as the theoretical codes emerge, keep the substantive directions in tractable focus, as they force the generation of a *core* category.

These three types of questions keep the analyst theoretically sensitive and transcending when analyzing, collecting and coding his data. They force him to focus on patterns among incidents which yield codes, and to rise conceptually above fascinating experiences. As the analyst's skill develops, these questions become upper most in his mind and part of his continual analysis and theoretical sampling.

The second rule is to *analyze the data line by line,* constantly coding each sentence. This may seem somewhat painstaking, but as codes emerge and saturate, it becomes easier and faster. It is necessary for achieving a full theoretical coverage which is thoroughly grounded. Another approach, (the

2. *Discovery,* Chapter V

"overview approach") is to *read over-all the data somewhat quickly,* which yields an impressionistic cluster of categories.[3] We do not recommend this approach by itself. The analyst cannot be sure of these categories—since this type of coding does not facilitate verification, saturation and relevance of codes nor any idea of what has been missed, because of a glossing over of details.

The "overview approach" by itself tends to yield thin theory with rich but dubious relevance and the feeling that much has been left out. Memos became too speculative and uncontained. It also, unfortunately, facilitates unchecked the analyst unwittingly forcing his "pet" theme on the data, instead of letting the data emerge in its own right by constant, close, careful generating of codes from each line of data. The line by line approach forces the analyst to verify and saturate categories, minimizes missing an important category, produces a dense rich theory and gives a feeling that nothing has been left out. It surely corrects for the forcing of "pet" themes and ideas, unless they have emergent fit.

The line by line approach raises another problem the resolution for which we have a third, offsetting rule. It is painstaking and timetaking to code carefully, but *the analyst must do his own coding.* Being very busy and having other things to think about, tempts analysts to hire a coder. This works when data is coded for pre-set coded type studies which need quantitative totals for description. IT DOES NOT WORK WITH GROUNDED THEORY STUDIES for several reasons.

A coder has to have a list of codes. We have none in the beginning and a list emerges only as categories emerge and saturate. Obviously the coder cannot code for categories until they have been emerged by an analyst, and even then he should not use the codes until they are sufficiently verified by the analyst. Also the coder wastes his time coding for categories saturated by others, when he does not know this. This knowledge is hard to come by quickly enough. A weekly meeting is too late; instant phone calls are too cumbersome. The quest in coding is theoretical meaning thus saturation *not* distributing a category over a population thus counting how often it appears, which warrants coding beyond saturation of meaning. Indeed, we do not even have units of distribution.

Coding is constantly stimulating of ideas. Another important, vital fourth rule is to *always interrupt coding to memo the idea,* in order to reap the subtle rewards of the constant input from reading the data carefully and from asking of it the above questions. The coder, hired for a specific task cannot stop and memo, and he is usually incapable by training to go beyond the data and theoretically memo. (We discuss memoing in detail in chapter

3. Robert K. Merton, *Social Theory and Social Structure,* (Glencoe, Ill: Free Press, 1957) pp 12-16

5.) Thus to syphon off this vital input from the analyst to a coder is a great loss to generating of theory. The latter cannot do it and the former misses the chance to do it. Further the latter as a coder has no stake in the analysis anyway, hence little motivation to push for those theoretical notions necessary to theoretical completeness.

In short, always code your own data, and if someone else is to code too, he should be hired as an analyst and given a stake in the resulting theory for his own career and treated as collaborator with full analytic rights. Then when coding varies over the same data, it can be analyzed to resolve the difference and establish the best fit code.

Collaboration suits the "one-uping" rule as an analytic mechanism. When the analyst asks the theoretical questions of his data detailed above, he essentially is forcing himself to start conceptually coding and thereby raising the empirical level of the data to a conceptual level suitable for theory generation. An analyst can do this for himself, but it takes practice in the beginning to keep one-uping himself: forcing his reperception of data into a concept. Thus many people have trouble conceptually coding—constant comparisons helps immensely as do the questions. But with a collaborator, an analyst can move considerably faster, even when he is highly skilled as a solo analyst. While one analyst is getting the empirical straight and saying it aloud, the other can focus solely on its generalizing and conceptualizing. To repeat, though not impossible by any means, it is hard to operate on both levels, so analysts must pace themselves for these levels by, say, inputting one day and conceptualizing the next (See Chapter 2).

In engaging in this collaborative mode, both analysts must be careful not to hurt the other's feelings. It is unnerving and galling to hear an empirical statement repeated back as a concept. *"One-uping" is not a put down;* it is a collaborative theoretical effort to raise the level of the data as quickly as possible while carefully fracturing it. This type of collaboration should be very focused on the line by line approach and careful notes and memos (or a tape) kept. The empirical stating analyst can in turn one-up the conceptualizing analyst with additional ideas and properties of the code and relations between codes while both are memoing.

Collaborative forms may be with two or more analysts on the same study or two or more doing it for each other each on their own studies, to keep each moving along faster than a solo trip. But once again, be careful of the psychological brutality of the one-uping process—the feeling is initially always one of being upbraided as if one missed the "real" conceptual point. After the experience, and the reward of its output, most analysts will ask others for this kind of boost—it works.

Another rule is directly linked with theoretical sampling (see Chapter 3). In the initial coding, as well as sampling and through the next stage of establishing of the core variable and some of its properties the analyst

should *stay within the confines of his substantive area and the field study.* As we said in Chapter 3, if the analyst goes outside too soon, he can easily get derailed from relevance, fit and workability, since what is at core in other substantive areas will differ from his area. After he is sure of the relevance, fit and workability of this initial conceptual framework, then he can start comparisons with other kinds of data in other substantive areas to help enrich the theoretical content of his substantive theory—as opposed to using outside data to develop a formal theory. For example, in generating a theory of cultivating clients applicable to milkmen, the social ranking between milkmen and housewives was not apparent as a relevant variable.[4] After the theory was generated, it was briefly compared to professionals cultivating—such as with doctors and clients and it became apparent that many of the strategies used by milkmen were because they were cultivating up the situational social scale, not down as lawyers or doctors frequently do. Also, milkmen have lesser rank and professionals higher with respect to knowledge.

Lastly, a firm rule is that the *analyst should not assume the analytic relevance of any face sheet variable* such as age, sex, social class, race, skin color etc., *until it emerges* as relevant. They are never necessarily a property of the process under study until discovered so. Of course, in most preconceived studies of description and verification, they are categorically assumed to be differentiating and therefore of interest irrespective of the data. *We have found them often of minor or no relevance in studies of process*—e.g. becoming an alcoholic.[5] Therefore they too must earn their way into the theory. Some analysts do provide such descriptions of persons and places not as part of the analysis, but merely to orient the reader regarding where and to whom the process may occur. The caution is that the process can occur elsewhere too for different persons, unless one attribute can be shown as a part of the process.

Open coding carries with it verification, correction and saturation phenomenon. As the analyst gets deep into the data, he discovers that all data can be subsumed as an indicator of some category in the analysis. Later in the study nothing occurs as a surprise, after constantly comparing, analyzing and generating, sufficient codes to handle differential emergents. In short a total saturation occurs: all data fit.

Open coding both verifies and saturates individual codes. Initially codes come very fast, and it is important to realize that these codes need correc-

4. Odis E. Bigus, "The Milkman and His Customer: A Cultivated Relationship," *Urban Life and Culture*, July, 1972, pp. 131-165

5. Odis E. Bigus, *Becoming "Alcoholic": A Study of Social Transformation*, (San Francisco, University of California PhD Thesis, 1975), and Mary Catherine Taylor, *Alcoholics Anonymous: Recovery Processes in a Self Help Group*, (San Francisco, University of California PhD Thesis, 1977).

tion by trimming and fitting. They do not capture all the data in a mature way. The analyst should not become selective too quickly, tempting as that is, since each code in a vacuum can seem highly relevant. Open coding proliferates codes fast, which then begins to slow down the coding by continually verifying that each code fits, eventually saturating the code and placing the code in its true relevance among other codes. This means finding its true relation to the *core variable* if it is not core itself. Verification, correction and saturation process is a part of the delayed action nature of grounded theory so the analyst should not be misled by initial quick results.

Memos force one to realize this. They make the analyst realize that the initial codes, may be only one way to code a datum, and may not be related to many other codes. And by going over and over the data, he realizes they may not be the best fit, hence unverified. Since the theory must be grounded, verifying its fit and relevance requires patience in going over and over the data to be sure it works with ease, before a secure investment is taken in selective coding for a focus on a core variable. Premature selectivity in a core variable or a basic social process can leave the analyst high and dry when nothing seems to work. Even two analysts in the same substantive area may think that they are going to study the same process, only to discover later that their different field situation has made them note very different processes.

Selective Coding: Now the question arises, how does the analyst know when it is safe to selectively code for a core variable and to cease open coding. While an analyst can see the prospects for a theory that does cope with the data entoto, it often seems wise, given human finiteness, to delimit the theory to one core variable (often a basic social process or condition among the few that may be going on. See Chapter 5). The other variables are not lost, but to focus on the analysis of one core variable merely demotes possible other core variables to a role subserviant to the variable under focus. Thus, to analyze the basic condition of awareness context of dying patients we used the categories of social loss and dying trajectory, although in other papers these were the core variables under focus.[6]

To selectively code for a core variable, then, means that the analyst delimits his coding to only those variables that relate to the core variable in sufficiently significant ways to be used in a parsimonious theory. The core variable becomes a guide to further data collection and theoretical sampling. The analyst looks for the conditions and consequences and so forth that relate to the core process. His analysis is guided by the core variable. Selective coding significantly delimits his work from open coding, while he sees his focus within the total context he developed during open coding. The analyst's

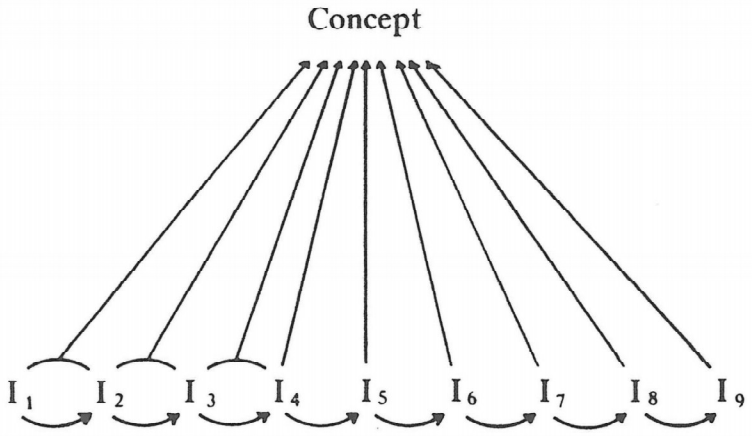
6. Glaser and Strauss, *Awareness of Dying,* (Chicago: Aldine Publishing Co., 1965)

memos clearly become more focused and start integrating while showing gaps for theoretical sampling.

How to find a core variable and its essential characteristics are developed at length in Chapter 6.

Concept and Indicators: It is important to look in some detail at just how a code is generated. Grounded theory is based on a *concept-indicator model*, which directs the conceptual coding of a set of empirical indicators.[7] This model provides the essential link between data and concept, which results in a theory generated from data. In Figure 1, we see a diagram of the model.

Figure 1



Our concept indicator model is based on constant comparing of (1) indicator to indicator, and then when a conceptual code is generated (2) also comparing indicators to the emerging concept. From the comparisons of indicator to indicator the analyst is forced into confronting similarities, differences and degrees of consistency of meaning between indicators which generates an underlying uniformity which in turn results in a coded category and the beginning of properties of it. From the comparisons of further indicators to the conceptual codes, the code is sharpened to achieve its best fit while further properties are generated until the code is verified and saturated.

This model must be distinguished from two other concept-indicator models. One is to a construct conceptual index by summing the indicators

7. Paul F. Lazarsfeld, "Problems in Methodology," in R.K. Merton et al (EDS) *Sociology Today* (New York: Basic Books, 1959) pp 47-67; "Evidence and Inference in Social Research" *Daedalus* LXXXVII (1958) pp 100-109; and with Wagner Thielens, *The Academic Mind* (Glencoe, Ill: Free Press, 1958) pp 402-407

which have been given differential values. Expositions of conceptual index construction abound. They do refer to the meaning of the indicators, but not analysis of the meaning. They primarily focus on distributing people among them to achieve a quantitative value. Our model focuses solely on the indicator's meaning. See Figure 2.

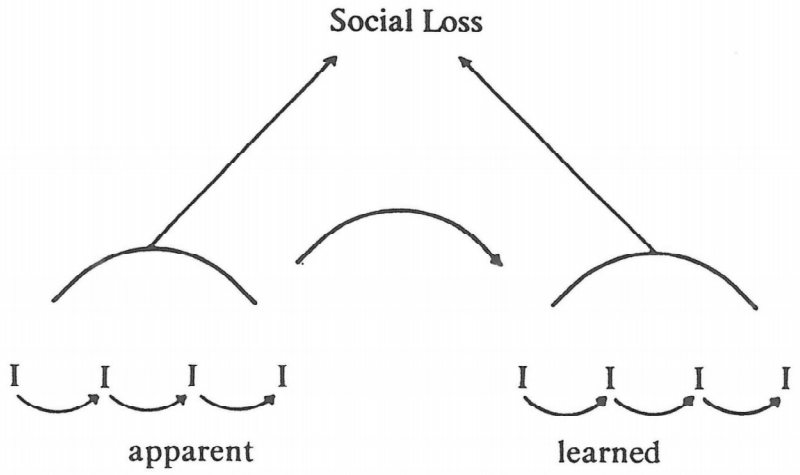
Figure 2

I + I + I + I + I + I + I = concept

The other concept-indicator model is to cluster the indicators into dimensions by a quantitative method such as factor analysis and then derive a concept for each cluster or take several clusters as dimensions (properties) of a higher level concept. This is a more sophisticated summing procedure where the meaning of the indicators is established by their total interrelations and thereby clustering into meanings.[8]

Neither of these two models employs our constant comparative approach which by comparing the meaning of indicators to each other slowly builds up to a concept and its properties. Figure three gives an example of the generation of the category of Social Loss with its two dimensions "apparent" and "learned".

Figure 3



Our model of concept-indicator generation has several vital facets of importance for grounded theory. Concepts and their dimensions (whether

8. See a methodology book on Factor Analysis and Lazarsfeld, *Ibid.* on Latent Structure analysis.

emerged or emergently fit) have *earned* their way into the theory by systematic generation from data. This may be seen as opposed to simply using the standard distinctions of received sociology *as if* they must be relevant. These distinctions, such as inner-outer, male-female, success-failure abound in texts, in logical derivations, in the writings of great men and in the training, verbal traditions. They are important to learn as a professional sociologist, but they cannot be used apriori in grounded theory unless they earn their way into it as workable and relevant.

Data determined distinctions prevent the fracturing of a concept into too many dimensions or developing too many properties of it. The over elaboration of a concept becomes quite easy when using received distinctions with logical elaboration. That style of analysis inhibits, by burying in its excess, the true use and relevance of earned distinctions that have derived their meaning *only* from the data and then from their grounded, systematic connection with other discovered categories. Pairing down this excess from received distinctions, and using only those earned helps achieve the goal for theory of parsimony of concepts, while at the same time richly densifying the theory.

In discovering a concept and its dimensions (such as apparent and learned social loss), the analyst soon realizes that the standard sociological way of defining a concept is too restrictive when compared to the specifying of its operational distinctions relevant to the emerging theory. *Conceptual specification* is the focus of grounded theory, *not conceptual definition.* This is because the operational meaning of the concept derives from the use of its earned distinctions in the theory. Thus, we have more of a notion of what "social loss" of a patient means by looking at a nurse's response to apparencies and learned calculations, than by defining it with the assumption that participants will respond accordingly. In this way the meaning of a concept can be modified or added to—as indicators change—thereby changing the applicable distinctions thus the concept's meaning. It is hard to keep changing a conceptual definition. In contrast, generation can easily keep changing distinctions for operational reasons as the theory emerges.

Changing indicators and thereby generating new properties of a code can go only so far before the analyst discovers saturation of ideas through the interchangeability of indicators (See Chapter 3). The more the analyst finds indicators that work the same regarding their meaning for the concept, the more the analyst saturates the distinctions and properties of the concept for the emerging theory. Nothing new happens as he reviews his data. The category and its properties exhaust the data. Meanwhile the analyst goes on to saturate other categories by the constant comparative method.

The interchangeability of indicators produces at the same time the transferability of the theory to other areas by the link of finding indicators in other substantive or sub-substantive areas that produce the same category or

properties of it. The theory applies to more and more structural units, (until if desired the analyst generates a formal theory). For example, by looking at attention indicators one can find similar ones in an emergency room, and intensive care unit, a waiting room and a battlefield. One can, if desired, then generate a theory of attention distribution in response to social value as perceived and learned.

This transferability of theory through comparing interchangeable indicators allows theory to apply to less obvious areas. Thus awareness context theory can also apply to a hospital neonatal ward (where no patient is aware), to comparing experts such as sub-contractors or to choosing doctors.[9] Since the interchangeability of indicators posits a universe of indicators from which the analyst never finds more than some proportion, transferability is endless even if empirical. It also means that there are always plenty of substantively varied similar indicators to use in illustrating a concept.

Needless to say, grounded theory is based on multi-indicator concepts, *not* one-indicator concepts as much conceptual theory is. A one-indicator concept is, of course, merely an illustration of what might conceivedly be a concept that could fit and work, but which the reader never knows—while the author conjectures. In contrast, saturation of the concept through interchangeability occurs long before a high proportion of all possible indicators are found. The analyst also increases his expertise in understanding and using a grounded theory in diverse contexts as he increases his knowledge of the universe of possible indicators in various social units. At the same time saturation through interchangeability reduces his need to find all indicators and discover all their meanings as is often the frantic case in studies of description and verification.

Constructing Typologies: A special case of our concept-indicator model is the construction of typologies. There are two sources of typologies, one is the construction of types by a social or psychological scientist e.g.—the high achiever. The other is the socially generated types by laymen—e.g. heroes, villains and fools.

All typologies are based on a differentiating criteria.[10] Consistent with the above statements the criteria used in grounded theory are *earned* distinctions, not received distinctions. The criteria may be internal to a concept as either dimensions of it (temporal expectations and certainty expectations)[11] or degrees of it (high producers or low producers). Or the criteria may be separate properties of the concept which are external to it (motivation—reward). The reason we use earned or grounded distinctions, is that each type must

9. Glaser and Strauss, "Awareness Context and Social Interaction," *American Sociological Review,* 29, (1964), and *Awareness of Dying,* op cit.
10. R.K. Merton, *op.cit.* chapter IX.
11. Glaser Strauss, *Awareness . . .* op.cit. Chapter 2.

make a difference in its relation to other categories. That is, say, a patient who is a high social loss must draw different attention from nurses than one with a low social loss. If each type does not make a difference, they can be collapsed, since there is no sense in using them in the theory. Thus, if degree of social loss does not matter, then a typology based on degree is a superfluous elaboration. We have found that earned distinctions have a greater probability of making a difference in their relation to other categories, *but* they do not necessarily have to.

There are two operations in constructing a typologies from a distinction. One is reduction: moving from the criteria to the typology. The other is substruction: moving from the typology to the criteria. The latter reverses the normal reduction process, by looking for the implicit criteria from which a typology had been unawarely constructed. This applies both to social types and to types that the analyst constructs without quite knowing his criteria. They just occur as he analyzes his data.

The analyst constructs a type by reduction when he cross-tabulates the internal or external distinction of a concept. In Figure 4 we see the types of death expectations that are generated by cross tabulating time and certainty. We have started out "wholistically" with the concept death expectations, discovered two dimensions of it, then reduced them to a typology.

Figure 4

Outcomes

		Certainty of Outcome +	Certainty of Outcome —
Time of Outcome	+	Certainty of time and outcome: nothing more to do	Certain time when outcome will reach turning point
	—	Certainty outcome but lingering	Wait and see

In this example the reader can see that the naming of and interpreting of the types can vary, but again should be grounded. Often nurses talked of patients who "lingered" or for whom there was "nothing more to do."

Further, not every cell has to have a meaningful type in it. Each type of the typology must fit, and in bigger tables, such as using trichotomous variables, the analyst often finds ungrounded cells. It becomes a mere logical elaboration of no worth to the theory to develop a type for such non-empirical cells. Thus in the dying study the recovery cell was of no theoretical use.

In Figure 5 we give another example of a reduction typology. It combines the two dimensions of social loss as a calculation of a social loss typological.

Figure 5
Social Loss Calculation

Apparent	Learned +	Learned −
+	Accurate calculation	To be surprised (perhaps)
−	To be recalculated with new information	No calculation

Figure 6

Local-Cosmopolitan

Profession-oriented	Local-oriented +	Local-oriented −
+	Local-cosmopolitan	Cosmopolitan
−	Local	Disinterested

In Figure 6 we see an example of the local-cosmopolitan typology.[12]

12. R.K. Merton *op.cit.* Chapter X and Barney G. Glaser, *Organizational Scientists: Their Professional Careers*, (New York, Bobbs-Merrill), Chapter 2.

In Figure 7 we see a typology based on age and degree of sophistication.

Figure 7

Age	Sophistication +	Sophistication −
old	man about town polished person	yokal
young	young sophisticate	sweet young thing

In Figure 8, we give an example of a typological substruction of a successful person and a failure and in the bargain we obtain a reduction to two other types the lucky person and the lazy or indifferent person. In trying to discover why some scientists felt successful and others felt like failures I discovered the two criteria of motivation to produce worthy research and recognition of worthy production. The successful were usually highly motivated and had recognition for the products of motivation. This is also a good example of a typology base on two separate properties external to the concept.

Figure 8

Success Typology

Motivation to produce	Social Recognition +	Social Recognition −
+	Success	Failure
−	Lucky	Indifferent to reward system

Running through the minds of people we study, no matter the substantive action area, are usually social types of people. They are phenomenologically developed and never analytically substructed. Familiar types are creeps,

geniuses, fools, heroes, stars, losers, jerks, victims, crooks, nice guy and so forth. While sociologists substruct to be analytically clear, the people of the substantive area never have to substruct their social types, for the functions of their use are different. These people usually think of others by what they are doing or did and type them according to the behavior involved. It gives the typed person a social identity. It is the basis of relating to them and establishing a social control over them when they are processed by the imputed types—which is typical, especially in occupations, family and deviant behaviors. The person is typed as a social whole, overruling all counter specifics although there are many other facets to the person. It is handy even though unjust; it works well in the world of business, and less well in intimacy and friendship relations, where responding to multi-facets of people is more appropriate.

The paradox is this. When a social scientist uses a social type in his analysis, it is considered an affrontery to the sensibility of those concerned in the substantive area. They feel it simplifies what is complex. They can think of too many other dimensions of a typed person, so they consider the analysis way off as the other dimensions are lost as well as scientifically denuding what is really going on. In our view they tend to be right. Using a social type as a type of person in analysis can be a travesty on the data.

It is unnecessary, also, in grounded theory, because our work is to *type behavior not people.* We talk of cultivating behavior, not cultivators. This allows the actors in grounded theory to walk in and out of many behavior patterns without being typed as one of them. Our actors can roam unlabeled and unclassified. They can succeed here and fail there and not be failures or successes, deviate here and conform there and not be deviants or conformists, and so forth. This does not offend people, since the emphasis is on behavioral patterns, not personal patterns. This emphasis on typing behavior not people, fits with our primary emphasis on process sociology as opposed to unit sociology (See Chapter 6). We talk of cultivating behavior as a process which occurs among milkmen, we do not engage in a descriptive analysis of milkmen as cultivators.

However, at some "cutting point" a person may be considered to engage in a type of behavior and its sub-culture to such an extent that his whole life may take on the identity and he may for analytic purposes be typed that way.[13] For example some juveniles engage, from time to time, in delinquent behavior. A few others become full time deviants in attitude and behavior and can possibly be typed that way. Alcoholics and junkies are other examples of full time behaviors. Cutting points in these cases are never clear, but they do emerge in the data, when it is appropriate to label a person a type. We are all familiar with the self-realization of a person who can define

13. John Lofland, *Deviance and Identity,* (New Jersey: Prentice-Hall, 1969).

himself as an alcoholic, when he realizes that so much of his waking hours is devoted to alcohol.[14]

In sum, while the laymen uses social typing of persons for social purposes, we do not need them for analytic purposes. Most often the actors in our analyses engage in a type of behavior without being typed by it, while they engage in many other behaviors as well.

Conceptual Dosage: As the analyst codes the data, he discovers two types of categories. Sociological constructs and *in vivo* codes. The latter are taken or derived directly from the language of the substantive field, such as composure, cultivating, legitimizing and negotiating. In vivo codes tend to be the behaviors or processes which explain how the basic problem is resolved or processed. They fracture the data directly when abstracted. They can imply theoretical codes; for example, cultivating implies looking into consequences, since anticipating consequences is why people cultivate.

Sociological constructs and in vivo codes should have two components in grounded theory: *analytic ability and imagery.* Analytic ability relates it to other codes with specified meaning and carry it forward in the theory with ease. Imagery is so that the analyst does not have to keep illustrating the code to give it meaning. Its imagery implies data with meaning sufficiently well so that the analyst does not bog down the writing with illustrations. In vivo words have a very vivid imagery with much local interpretative meaning. They have grab for people involved in the area and similar areas. They sound so meaningful; they are seldom forgotten. They also have much analytic ability, since the people of the field use them with ease and sufficiently precise meaning.

Sociological constructs are codes formulated by the sociologist, such as "social loss, awareness context, or anticipatory socialization." They are based on a combination of the analyst's scholarly knowledge and his research knowledge of the substantive field. As a result they can add more sociological meaning to the analysis than *in vivo* codes. They add scope by going beyond local meanings to broader sociological concerns. They have much analytic ability because they are constructed with clear meaning. They may have little imagery, as some analysts think that the flatter they are the more scientific and less impressionistic—e.g. role occupant or control agent.[15] But others put much imagery into them—e.g. awareness context or social loss.

The relative dosage of type of concept in a study is of course up to the analyst, but usually one or two (maybe three) sociological constructs are all a study can take along with many in vivo codes. The constructs are the core

14. Bigus Taylor, *op.cit.*

15. Howard S. Becker, *Outsiders: Studies in the Sociology of Deviance,* (Glenco, Ill: Free Press, 1963).

or close to the core variable. Too many constructs may make an analysis sound too rhetorical, contrived, airy or wordy. On the other hand in vivo codes, easily recognizable with meaning, make the reader feel familiar with the ideas of the analysis. In all 10 to 15 codes are typically enough for a monograph on a parsimonious substantive theory. One can generate considerably more, say fifty, but usually only a small set are most relevant for the theory. Over coding is unnecessary. Too many codes dilute the impact of core and near core ones, and add only minor variations. Concept manufacturing is a dangerous game and can easily go empty, when relevance is lost for the thrill of proliferation of "neat" sounding words.[16]

Mechanics of Coding: The mechanics of coding are as well known as they are varied. But it is important to note here those variations that aid and those that hinder generating grounded theory. The major goal in coding is to *use a method that facilitates high sortibility and flexibility.* Basically when sorting, the analyst sorts both memos on a code and the code with data, so he can see exactly how he grounded his memo, hence his theory. By sortibility we mean that memos on the code and the code along with the data that indicated it can be scissored out of the memos and the field notes with ease for sorting. We go into sorting extensively in Chapter 7.

By flexibility we mean that codes come fast and the mechanics of coding should accommodate this, as well as be easily used for constant comparisons of indicators or concepts to indicators. Also the mechanics should flexibly allow the analyst to stop and memo and get right back to where he was coding, and allow the analyst to keep sufficient track of indicators so he can tell when he has saturated a code, and thus no longer need code it. Lastly, the mechanics should foster references back to data through codes to be used when writing occurs months later.

We have tried many mechanics and found that the best way to achieve these goals is to code in the margin right next to the indicator. Then the analyst may cut up field notes for sorting, when sorting them with the memos is required. Otherwise the code and its indicator can be referred to quickly in the field note—by date and page. This method assumes that field notes should take up only two thirds of the page to allow ease of writing in margins. And they are done with carbons so when cutting them up, a complete set of field notes still remain. Writing in margins can be done quickly and saturation can be checked out by quick references back, when memory is vague. Index cards, McPhee cards, separate papers, IBM cards and other mechanics used in quantitative analysis and content analysis are hindering to the constant comparison of indicators and concepts which generates the codes.

---

16. Erving Goffman, *The Presentation of Self in Everyday Life,* (New York: Doubleday, 1959).

their properties and theoretical memos. They are suitable for the summing concept-indicator model, not ours: the comparing concept-indicators model.

THEORETICAL CODING

As we said in the introduction, theoretical codes conceptualize how the substantive codes may relate to each other as hypotheses to be integrated into a theory. They, like substantive codes, are emergent; they weave the fractured story back together again. Without substantive codes, they are empty abstractions. The analyst should never forget this no matter how enthralling a theoretical code may be. Yet substantive codes can be related without them since theoretical codes are always implicit, but then the relations are likely to be less clear and often to have formulation confusions. Thus, questions arise such as: is the author really talking covariances or causes or is he talking causes or anticipated consequences?

Every now and then a critic of sociology (including many sociologists) says "Everything of importance has been said by somebody who did not discover it." Perhaps so, but theoretical coding, in establishing new connections that make ideas (however recognizable) relevant, is what is so often the "new" and "original" about theory. Theoretical codes give integrative scope, broad pictures and a new perspective. This is why grounded theory is so often "new" because of its *grounded integration.*

The implicitness of theoretical coding is used purposefully in grounded theory. One talks substantively and thinks theoretically of the relationship between codes. Though one can also talk theoretical codes, they are just as easily presented to readers as implicit, while the analyst is quite explicit in his formulations. They are amazingly flexible, several may fit the same data. Thus the analyst's choice while perhaps arbitrary is still grounded in one of many useful fits. The theoretical code must earn its way like a substantive code. The choice starts determining integrative patterns which limits the freedom in further choices. In this way, theoretical codes describe the world to us theoretically and can span all current perspectives depending upon how they are chosen and combined. The analyst must be careful of logical elaboration in theoretical coding, since it occurs so easily with them. He must continually remind himself to ground. We all know, for example, how easily role theory can get elaborated with logical possibilities.[17]

It is necessary for the grounded theorist to know many theoretical codes in order to be sensitive to rendering explicitly the subtleties of the relationships in his data. Some are standard and some less so. Like our respondents, we all know many implicitly and impute them to other's ideas. Though never

17. Neal Gross, et al, *Explorations in Role Analysis,* (New York, John Wiley and Sons, Inc., 1958) and R. K. Merton, op.cit.

fully known (they are always emerging), the fullest range of theoretical coding possibilities give the grounded theorist a powerful approach to generation of theory. It sensitizes him to the myriad of implicit integrative possibilities in the data. We will present several lists of theoretical coding families in a few pages below for study and reference.

Theoretical codes are not hard to learn, but interestingly enough most theorists in sociology to date, seem to focus only on their "pet" code—the code they were indoctrinated into—and no others! Indeed, many do not know about theoretical coding. Hence they miss the best possible renderings for the need to over-focus *only* on say: a role or status problem, a process, a dimension, an interaction strategy, a subordinate—superordinate issue, a consensus-confirmity pattern and so forth. They think in terms of studies using the same code, not variations in theoretical codes. It is very easy to get stuck on a "pet" code through training, development of skill and imputations of orthodoxy in learning a code by some sociology departments. Their entrapment in paradigms is basic to science. Thus analysts of grounded theory must continually watch how they are putting the theory together and take their cues from the data. It is also worthwhile to look at the codes of other disciplines for new and sophisticated theoretical ideas. An example is the "bias-random'walk" model used in biochemistry, the idea is that all variables are constantly present in unordered covariance but each is used and ordered only as a process occurs and needs the variable to proceed, after which it is dropped back into covariance.

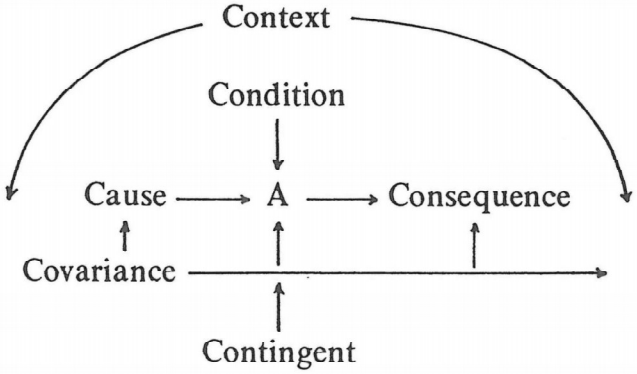
It is interesting to note a few other uses of theoretical codes. First they help the analyst maintain his conceptual level in writing about concepts and their interrelations. They prevent the analyst from dropping and bogging down in data, through the trap door of an interesting illustration and the normal propensity to talk about people, but more on this in Chapter 8 on writing. Theoretical codes are excellent to use in critiquing a sociological work. In using a theoretical coding family, the analyst can tell how the writer overlooked other aspects of the family and indeed other related coding families. The overlook is quite general since most theoretists just do not realize how finite is their repertoire of theoretical codes and that each code implies a coding family. Lastly, theoretical codes can be used, when the analyst is forced to preconceive, by logical elaboration in a grant proposal, possibilities to discover for verification. They give the applicant something to say when he does not know anything about the data to be collected!

We proceed now to list 18 coding families with comments on their use. The reader will see another property of their flexibility; they are not mutually exclusive, they overlap considerably. Also one family can spawn another. Their conceptual level in relation to each other is also arbitrary by their use. Some are different ways of saying the same thing. The reader will think of other words for each family as well as discover new families. But in the final

effort, they make the analyst very sensitive to what to write at which point, when in fact, he must take a stand and try to theoretically render an empirical pattern.

Coding Families: 1. *The Six C's:* Causes, Contexts, Contingencies, Consequences, Covariances and Conditions. This is the "bread and butter" theoretical code of sociology. It is the first general code to keep in mind when coding data. Most studies fit into either a causal model, a consequence model or a condition model. Causal has a sub-family called: sources, reasons, explanations, accountings or anticipated consequences. If one is forced to preconceive data, for a grant proposal, the six C's are good to elaborate. A subfamily of consequences is outcomes, efforts, functions, predictions and anticipated or unanticipated consequences. A causal-consequence model, depending on the focus is the independent-dependent variable model. To focus on the former is to look for its consequence, the dependents variable, and the latter its cause, the independent variable. The causal-consequence model implies an ordering which is usually temporal (see family 15. below). Causes and consequences can easily become mixed up without a clear ordering. Qualifiers is a synonym for conditions. Covariance is seldom used in sociology and it is a powerful idea. It includes connected variables without forcing the idea of cause. Ambiance is a synonym for context. A change of causes is a process. In Figure 9 we diagram the six C's as properties of A.

Figure 9

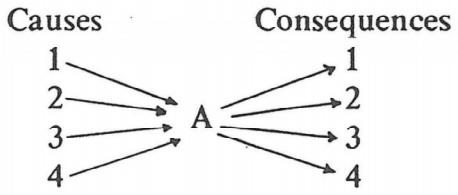


2. *Process:* Stages, staging, phases, phasings, progressions, passages, gradations, transitions, steps, ranks, careers, orderings, trajectories, chains, sequencings, temporaling, shaping and cycling. *A process must have at least two stages.* The analyst *cannot* talk of process and not have at least two stages. Process is a way of grouping together two sequencing parts to a phenomenon. Processing refers to getting something done which takes time

or something happening over time. Sociological problems are socially processed irrespective of whether they get solved or not. A causal-consequence model is a process.

3. *The Degree Family:* Limit, range, intensity, extent, amount, polarity, extreme, boundary, rank, grades, continuum, probability, possibility, level, cutting points, critical juncture, statistical average (mean, medium, mode), deviation, standard deviation, exemplar, modicum, full, partial, almost, half and so forth. The point to remember in theory generation is that, since variables vary, everything we say *implies* a matter of degree. Since it is cumbersome to constantly state a degree qualification for a code, we must always *assume* (and let emerge) the full range, even though it appears we are merely discussing one point on it. This assumption puts to rest the "one uping" that sociologists like to utter in always reminding each other that something is a "question of degree" or who point out another place on the continuum "as if" the analyst had overlooked it, usually a negative case. For example pollution implies smelling good and smelling bad, but is used for the latter, a negative case. As grounded theorists we need only let relevant degrees of our categories emerge, the rest is merely logical elaboration, not forotten or oversighted. A range model which shows the full range of outcomes or causes is a useful combination of the 6 C's and degree families.

Figure 10



4. *The Dimension Family:* Dimensions, elements, division, piece of, properties of, facet, slice, sector, portion, segment, part, aspect, section. The dimension family divides the notion of a whole into a parts. The more one learns of a category the more he begins to see its dimensions; it breaks down into "pieces of." As we said above, it is best to specify concepts with the relevant operational dimensions rather than define them wholistically or logically elaborate several possible dimensions of no relevance, which is over fracturing the concept.

5. *Type Family:* Type, form, kinds, styles, classes, genre. While dimensions divide up the whole, types indicate a variation in the whole, based on a combination of categories. The categories themselves may be dimensions, but may begin in the generation as separate categories, such as we observed in the example on success and failure above.

6. *The Strategy Family:* Strategies, tactics, mechanisms, managed, way, manipulation, maneuverings, dealing with, handling, techniques, ploys, means, goals, arrangements, dominating, positioning. This family has lots of "grab" for analysts and readers alike. Interaction sociologists especially talk a lot about how people strategy people. However, the structuralists also talk of mechanisms and arrangements that strategy people from the point of view of social organization.

The point to keep clear on is whether or not there was a conscious act to maneuver people. If not, then a behavior pattern is a *consequence* of another behavior, and it is inaccurate to impute that the behavior was the result of a conscious approach to manage others. But some theorists like to impute this management idea, even when they have not grounded it and when the best they could accurately say is that one behavior is a consequence of the other. We call this the "over-managed" view of man. Goffman is a clear example of a theorist who takes this excessive view.[18] Some theorists just do not know the theoretical code of consequential behavior, and therefore are forced to use a strategy term as the only word available to them.

7. *Interactive Family:* Mutual effects, reciprocity, mutual trajectory, mutual dependency, interdependence, interaction of effects, covariance. This code is an effort to capture the interacting pattern of two or more variables, when the analyst cannot say which comes first. Nor does it matter, probably. For example, rewards lead to motivation to do the rewarded behavior, but also motivation to work leads to seeking rewards. Thus once the ball is rolling they feed on each other. They are interactive, no matter how it started, nor how it ends. The inplay of interactive effects is clearly related in some cases to the strategy family, when one actor is purposefully trying to advantage or position himself.

8. *Indentity-Self Family:* Self-image, self-concept, self-worth, self-evaluation, identity, social worth, self-realization, transformations of self, conversions of identity. It is best to clearly state operational dimensions of self or identity when they are used in grounded theory. Such clarity is necessary, since there is at the same time much connation of meanings of self-identity words from textbooks and much vagueness on their meanings in the literature.

9. *Cutting Point Family:* Boundary, critical juncture, cutting point, turning point, breaking point, benchmark, division, cleavage, scales, in-out, intra-extra, tolerance levels, dichotomy, trichotomy, polychotomy, deviance and point of no return. This family is a variation of the degree family. Degree focuses on the full range, while here we focus on significant breaks or cutting points on the range. Cutting points are very important in theory generation, since they indicate where the difference occurs which has differential effects.

---

18. Goffman, *op.cit.*

10. *Means-goal Family:* End, purpose, goal, anticipated consequence, products. As we have just seen, this family is also a sub-family of the 6 C's and the process family.

11. *Cultural Family:* Social norms, social values, social beliefs, and social sentiments. Social norms are aggregates of rules, values aggregates of wishes or goals, beliefs aggregates of cognitions and sentiments aggregates of attitudes. The assumption is that personal characteristics are shared to a sufficient degree. Norms are of two types, behavioral and ideational; here we refer to the latter. The former is a social action pattern.

12. *Concensus Family:* Clusters, agreements, contracts, definitions of the situation, uniformities, opinions, conflict, discensus, differential perception, cooperation, homogeniety-heterogeniety, conformity, non conformity, and mutual expectation. At one point in sociology the concensus family was used a great deal. This was during the early voting and opinion studies when homogeniety of cultural elements and social units were strong. Now that so much heterogeniety of culture and units is acceptable enough to be expressed in everyday discourse and action the swing has been to talking of, not discensus, which implies concensus should prevail, but of tolerated differential perspections and expectations. Groups and people "do their own thing" now with legitimation or with little pressure to conform when no glaring negative consequences prevail. It is best to keep in mind, when using this code, its implication for its antithesis as a matter of degree. Thus concensus varies in degree to complete discensus.

13. *The Mainline Family:* Social control (keeping people in line), Recruitment (getting people in), Socialization (training people for participation), Stratification (sorting people out on criteria which rank them), Status Passage (moving people along and getting them through), Social organization (organizing the people into groups, aggregates and divisions of labor) and Social Order, (keeping the organization of life working normatively), Social institutions (clusters of cultural ideas), Social interaction, (people acting with people), Social worlds (symbolic surround of life), Social mobility (patterned paths of people movement through society) and so forth.

These codes, and there are more, represent mainline, traditional established sociology. They generate social values for action. They are concerned with large numbers of people in process. They represent matters which all sociologist of all views are concerned with. They are our traditional perspectives on sociological problems which span all sociological perspectives. They are the big rubrics that have been broken down in a myriad of ways to discrete studies. It is important to keep in mind the general rubric even in the most microscopic study. Every study, including a grounded theory study, is "of" one of these codes, and to know which one clearly infuses the study theoretically while broadening its perspective. Thus to study and analyze nursing work on an Intensive Care unit is helped by viewing it as a specific

study of social order and social organization. Perhaps the most frequent implicit rubric in studies is a problem of social order (usually disorder).

14. *Theoretical Family:* Parsimony, scope, integration, density, conceptual level, relationship to data, relationship to other theory, clarity, fit, relevance, modifiability, utility, condensibility, inductive-deductive balance and interfeeding, degree of, multivariate structure, use of theoretical codes, interpretive, explanatory and predictive power, and so forth. This family is important in generating theory and it is especially important in critiquing and judging the theory one reads and uses.

15. *Ordering or Elaboration Family:* [19] Structural, temporal and generality are the three principal ways to order data. (15. a) One approach to *structural ordering* is based on unit size: organization, division, group, subgroup, team, person. Often an analysis of structure assumes that the flow of influence and power is down the structure. But structural ordering of influence is an empirical question—whether and in what order the flow of a variable such as influence or decisions is up or down and in what order. The interrelations and interplay between structural levels is often so rampant and complex, that the critical point is to order units and to relate their properties on the basis of data, which may yield criteria much different than size. Structural ordering can build up to larger units from properties of the person, as they aggregate and develop cultural components.

The analyst must screen his work for the "ecological fallacy:" accounting for the behavior of a small unit by the properties of a large unit just because the two units exist together (covary). This is easy to infer, but must be empirically established to not be a fallacy. For example it was easy to assume in one study, that the high turnover in a neighborhood was accounted for by the high proportion of bachelors in the same neighborhood. But this coexistence proved to be a fallacy, since the turnover was caused by young marrieds looking for bigger living spaces, apartments and houses. The bachelors stayed put in the prevalent smaller apartments.

(15. b) *Temporal ordering* is the standard way to order categories. One category comes after another in a temporal sequence. It is often implicit in coding. "One thing leads to another." The basic orderings are X will cause Y, or X explains Y, or knowing X exists predicts Y, and A interprets the relationship between X and Y since it is an intervening variable (X-->A-->Y).

The intervening variable is often a psychological variable occurring between two behavioral variables, for example: Reward system----> Motivation----> Productivity. Thus in this case, to use a psychological explanation

19. Paul F. Lazarsfeld, "Interpretation of Statistical Relations as a Research Operation," in Lazarsfeld and Rosenberg (eds), *The Language of Social Research*, (Glencoe, Ill: Free Press 1955). See also *Discovery*, Chapter VIII.

such as motivation is to miss the full explanation in process. That motivation is actually an intervening variable connecting reward and production. In short beware of psychological explanations, they usually turn out to be intervening, not the prime cause. In the same example when production leads to seeking reward then the process becomes circular which it often is. Thus motivation may be seen then as a consequence, depending on how the analyst emerges the circular process. This again misses the point of its genuine "intervening" as the most central of its positions in the process.

It also must be remembered that to temporarily order variables is hard in surveys and fieldwork. If the analyst takes the position he is *generating hypotheses not facts*, he is ok, because it is on the factual level that temporal order is debatable. Hypotheses are merely suggestions.

The spurious or specious variable is a variable which occurs prior to both X and Y and causes both, instead of hypothesizing the spurious relation of X causing Y. The more the fire engines arrive the more the fire damage, is an apparently spurious relationship. Actually the prior variable that causes both is a large fire. Some spurious relationships are not so apparent such as an upset patient jumps out the window, which appears to be correct, until we find that a fire caused both the patient to be upset and the window to be the only exit. In any event, these relationships are a matter of discovery and it does little good to an analysis to talk of spurious relationships. It is more to the point to reorder the variables as they are emerged and analyze them as such: Thus a category causes covaring consequences or a chain of consequences.

(15. c) *Conceptual orderings* are the least used generally, but none the less important and powerful and used often in grounded theory, as in specifications of concepts, and in developing properties of categories. An example is the ordering for motivational generality in the socialization, of people: Achievement orientation, institutional goal, organizational value, personal motivation. Once again the flow is often assumed as down, but it is an empirical question. Society rewards achievement, the institutional goal of science is to achieve creative findings, a research organization impliments this goal in a specific way and the scientist is motivated to pursue creativity. The analyst must be sensitive to seeing the conceptual orderings in his data; they do occur as people derive motivation for acts from the values of the larger units within which they act or they see larger units effecting smaller or vice versa. These conceptual orderings can be closely linked to structural orderings: as orders flow down they get more conceptually specified and specific or as grievances flow up a structure they get glossy and generalized and rest finally on societal issues and mores.

16. *Unit Family:* Collective, group, nation, organization, aggregate, situation, context, arena, social world, behavioral pattern, territorial units, society, family, etc., and positional units: status, role, role relationship,

status-set, role-set, person-set, role partners. These are structural units which are familiar to all of us. The important thing to remember, we have found, is to keep the units to which our categories refer *clear.* Quite often analysts confuse their units by talking of properties of one as properties of another of a different order. For example, people are paranoid, not societies. If the analyst confuses units or does not make them clear, other relevant properties of the units, such as contextual conditions, may be missed in the analysis. For example if a nurse is working feverishly over a patient, we think differently of its meaning knowing whether the unit is an ICU or a medical ward. From the grounded theory point of view, units are clusters of *possibly relevant variables.* They are in process themselves, and they are where the action, behavioral pattern and process of our theory takes place (are grounded) for a time. At the same time our theories escape from time and place with their level of generality (See Chapter 6).

Two schools have formulated role theory to some extent, Harvard and Columbia, while many sociologists just use positional concepts as they see fit. In general, status is a position in a social structure, a role is the relationship between two statuses and the evaluation of status is rank. But often role is used as this meaning of status, and status is used as this meaning of rank. Thus, the analyst should always be clear on the meaning he is giving his positional terms.

We have also discovered that the notion of role-set is more a logical elaboration than relevant for grounded research. It is too gross a concept to handle behavioral patterns as we see them. If a teacher is talking to all his students in a few different patterns this is not enough to talk of as a role, and it is not a role set, and it is trite to say his role set is students, colleagues and administration. The real action takes place in his *person-set* of the student-teacher role. One can see the person set (the students) and analyze its variants, properties, and patterns. A mother has a family-set, a nurse a patient set, an accountant a client set; these are all clear person-sets to analyze and the reader can picture the units. This unit has fit and relevance and it works. It is an excellent unit for generating vital properties such as do they know each other, do they suffer from pluralistic ignorance, do they refer to each other, can they gang up on the "teacher", do they pass in and out of the person set, in solo, aggregate or concert and so forth. If on reduction the person-set resolves to a role-set fine. But do not start with role set, as it seldom (if ever) does become the most meaningful unit of analysis, and to start there means missing a lot of relevant action.

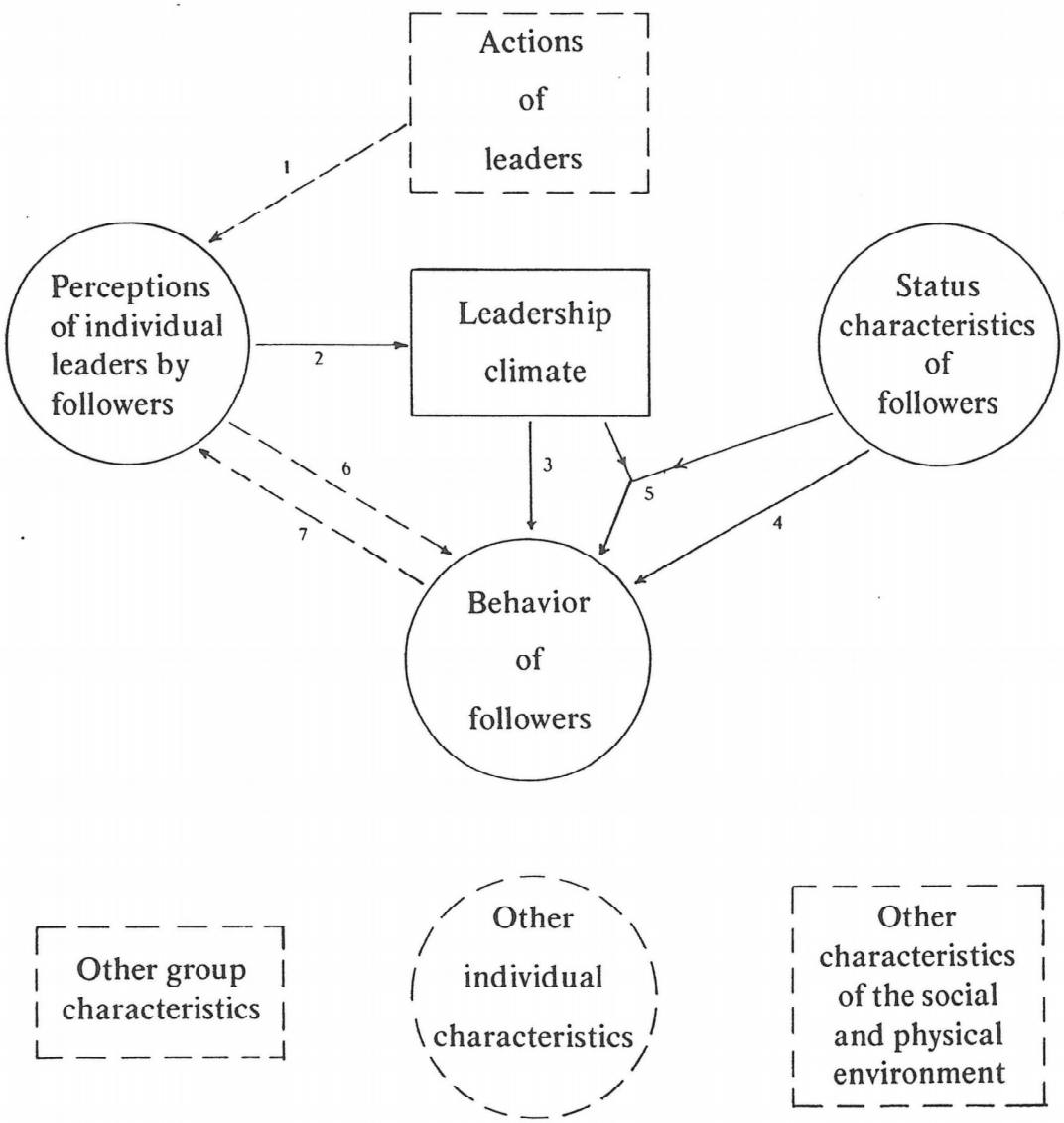
17.*Reading Family:* Concepts, problems and hypotheses. The best and simplest work in reading sociology is to underline these codes in different colors. They give one much formulation to take away from his reading. They are the stepping stone for more intense coding of reading, if called for along

the lines of other codes. They stimulate immersion in the reading and development of theoretical sensitivity.

18. *Models:* Another way to theoretically code is to model ones theory pictorially by either a linear model or a property space. Essentially any theory can be linearly drawn in the fashion of a molecule. Figure 11 is a good example.[20] These kinds of models are facilitating for writing. The analysis writes up the substantive meaning of each connection.

Figure 11

Categories and Relationships in the Study of Leadership Effects



20. From Hanan C. Selvin, *The Effects of Leadership* (Glencoe, Ill, Free Press, 1960) p 9.

Property space models can also facilitate writing, by writing up the meaning of each relevant cell and their interrelation. We showed in typological construction how to generate property spaces by reduction and subtraction. The caution in these models is that they are given to too much logical elaboration which never becomes grounded, but appears rich to write about. Thus writing up a model can become too wordy, sterile and dilute of relevance. Models can produce a fuller range than is relevant. Constraint in their use is mandatory.

By now we suspect that the reader has thought of more theoretical coding families and more words for each family. We hope to have touched on the most useful, but they continually emerge and arrive in sociology from other fields, like the bias random walk from biochemistry. The important point for us is that the reader has developed a clear notion of their conscious use and relevance in generating theory. Then he can use the above with theoretical sensitivity in putting his theory together. He can also develop his own coding families and never again be trapped into just writing about "pet" codes, as so many learn to do in training.