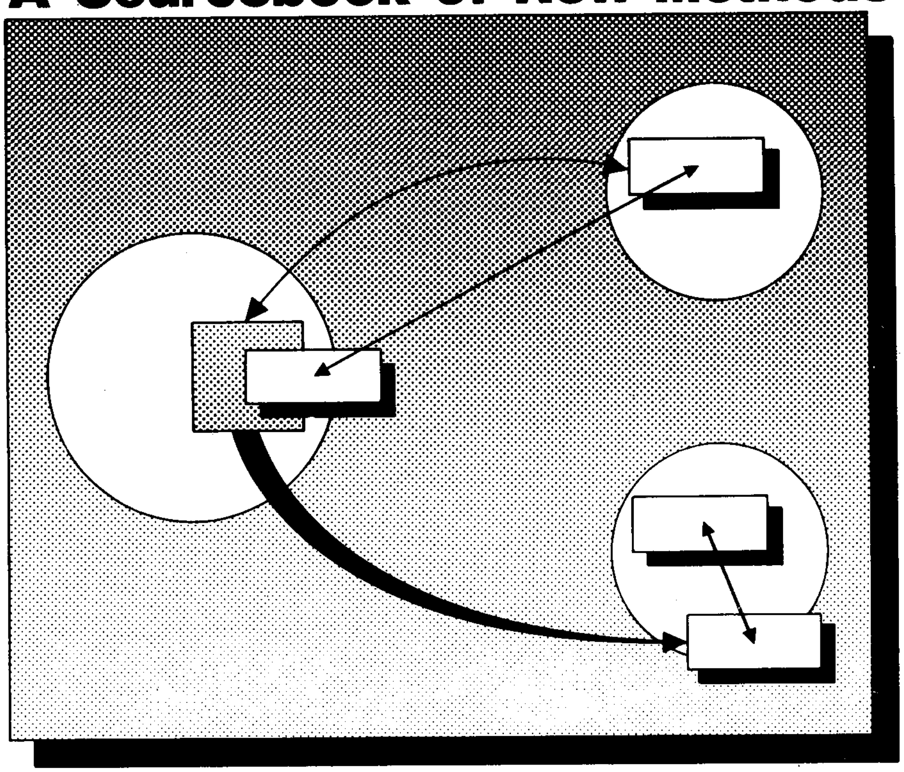
QUALITATIVE DATA ANALYSIS

A Sourcebook of New Methods



Matthew B. Miles
A. Michael Huberman

SAGE PUBLICATIONS
The International Professional Publishers
Newbury Park London New Delhi

For information address:

SAGE Publications, Inc.
2455 Teller Road
Newbury Park, California 91320

SAGE Publications Ltd.
6 Bonhill Street
London EC2A 4PU
United Kingdom

SAGE Publications India Pvt. Ltd.
M-32 Market
Greater Kailash I
New Delhi 110 048 India

Printed in the United States of America

Library of Congress Cataloging in Publication Data

Miles, Matthew B.
Qualitative data analysis.
Bibliography: p.
Includes index.
1. Social sciences—Research. 2. Education—Research.
I. Huberman, A. M. II. Title.
H62.M437 1984 300′.72 84-2140
ISBN 0-8039-2274-4

92 93 94 15 14 13

VII

Drawing and Verifying Conclusions

The many types of displays we have outlined always involve a general analysis *strategy*, a systematic approach to finding meaning in a set of data. But as we have seen, general strategies, in qualitative research as in war, are not enough. There will always be a flow of specific analysis *tactics*, ways of drawing and verifying conclusions that the analyst employs during the process.

VII.A TACTICS FOR GENERATING MEANING

In this section we discuss twelve specific tactics for drawing *meaning* from a particular configuration of data in a display. Usually, we will describe the general analysis situation being faced, explain the tactic, then give one or more examples, often referring back to previous sections. We will also refer to others' work for examples. If we can muster advice, it will be presented too. Our approach will be brisk; the test of these tactics comes in the using.

People are meaning-finders; they can make sense of the most chaotic events very quickly. Our equilibrium depends on such skills: We keep the world consistent and predictable by cognitively organizing and interpreting it. The critical question is whether the meanings found in qualitative data through the tactics outlined here are valid, repeatable, *right*. The following section (VII.B) discusses tactics for testing or *confirming* meanings, avoiding bias, and assuring the quality of conclusions.

Here is a quick overview of the tactics for generating meaning, numbered from 1 to 12. They are roughly arranged from the descriptive to the explanatory, and from the concrete to the more conceptual and abstract. *Counting* (tactic 1) is a familiar way to see "what's there." *Noting patterns, themes* (2), *seeing plausibility* (3), and *clustering* (4) help the analyst see "what goes with what." *Making metaphors* (5), like the preceding four tactics, is a way to achieve more integration among diverse pieces of data. Differentiation is sometimes needed, too, as in *splitting variables* (6).

We also need tactics for seeing things and their relationships more abstractly. These include *subsuming particulars into the general* (7); *factoring* (8), an analogue of a familiar quantitative technique; *noting relations between variables* (9); and *finding intervening variables* (10).

Finally, how can we assemble coherent understanding of data? The tactics discussed are *building a logical chain of evidence* (11) and *making conceptual/theoretical coherence* (12).

VII.A.1 Counting

In qualitative research, numbers tend to get ignored. After all, the hallmark of qualitative research is that it goes beyond how *much* there is of something to tell us about its essential *qualities*.

However, as we noted earlier, there is a lot of counting going on when judgments of qualities are being made. When we identify a theme or pattern, we are isolating something (a) that happens a number of times and (b) that consistently happens in a specific way. The "number of times" and consistency judgments are based on counting. They are not *only* counting exercises, but they are *also* counting exercises. When we make a generalization, we amass a swarm of particulars and decide, almost unconsciously, which particulars are there *more often*, matter *more* than others, *go together*, and so on. The moment we say something is "important" or "significant" or "recurrent," we have achieved that estimate in part by making counts, comparisons, and weights.[1]

So it's important in qualitative research to know, first of all, that we *are* sometimes counting, and to know *when* it's a good idea to work self-consciously with frequencies. There are three good reasons to resort to numbers: to see rapidly what you have in a large slice of data; to verify a hunch or hypothesis; and to keep yourself analytically honest, protecting against bias. Let's review each briefly.

Seeing what you have. Numbers, we noted earlier, are more economical and manipulable than words; one "sees" the general drift of the data more easily and rapidly by looking at distributions. For instance, in the school improvement study, we asked informants why they were using the new school practices we were studying. We got a mass of answers from several informants at each of twelve field sites. It *seemed* that a lot of people were saying they had been pushed, more or less gently, into these projects rather than diving in voluntarily. To see more clearly, we did a content analysis of the responses, totaled them, and derived Chart 41.

It turns out that 62 percent of the respondents mention pressure and constraint. And, counterintuitively, very few of the practices were adopted to solve problems. There also seemed to be a general "professional development/capacity enhancement" theme (challenge, shaping projects, professional growth). *Seeing* that theme, gauging the importance of the "constraint" motive, noting the infrequent problem-solving incentive, were all helpful. We saw the overall trends, got some new leads, saw some unexpected differences. All this helped in the subsequent *non*-quantitative analysis. Even within a *single site*, that kind of exercise would have been a useful one.

Verifying a hypothesis. Section V.C, Chart 32, is probably the best example of this. We reasoned that good preparation was the key to smooth initial use, so we simply created and computed a preparation "index" and set it against an estimate of smoothness of early use. Except at the extremes, we were wrong. But, by doing the counts, we saw at *which* sites we were wrong and *why* this appeared to be the case. We then set aside the numbers and followed up those leads.[2]

Keeping analytically honest. We had expected from the start that careers would be important in these school improvement projects. The more data we got, the more it seemed that "innovating" was a vehicle for moving up, in, over, or out (seldom *down*). The finding seemed important, potentially controversial, and might have been a result of our expectation. So we actually counted up the number of job moves (75 for 12 sites) and estimated how many could be attributed to the innovation (83 percent were). Afterwards, we felt far more comfortable about the claims we were making; for example, it seemed that only 35 percent of the job-related shifts were upward ones, contrary to our early impression.[3]

The last illustration is an important one. As a qualitative researcher, one works to some extent by insight and intuition. There are moments of illumination. Things "come together." The problem is that we could be wrong. There is a near-library of research evidence to show that people habitually tend to *overweight* facts they believe in or depend on, to *forget* data not going in the direction of their reasoning, and to *"see" confirming instances* far more easily than disconfirming instances (see the review in Nisbett & Ross, 1980). We do this by differentially weighting information and by looking at *part* of the data, not all of it. Doing qualitative analysis of all the data with the aid of numbers is a good way of seeing how robust our insights are.

VII.A.2 Noting Patterns, Themes

When one is working with text, or less well-organized displays, one will often note recurring patterns, themes, or "Gestalts," which pull together a lot of separate pieces of data. Something "jumps out" at you, suddenly makes sense. We have already discussed this under "pattern coding" (section III.C); such patterns can often be found under the heading of repeated themes, causes/explanations, interpersonal relationships, and theoretical constructs.

Some examples of patterns from our school improvement study:

- The frequent citing of a "miracle case" (a failing student who was rejuvenated by the innovation) as either an explanation or a justification for the project.
- "Organic coping" as a problem-solving style in a certain staff group.
- The use of "administrative latitude"—freedom to alter an innovation in return for trying it at all.

Pattern finding can be very productive as an analysis strategy when the number of sites and/or the data overload is severe. Stearns, Greene, David, et al. (1980), for example, studied 22 sites, all implementing a new law on education for handicapped children. Fieldworkers generated unstructured "pattern" statements of findings from more than one site; these were gradually reduced to a few hundred "propositions," ordered into 21 categories. This procedure used fieldworkers' knowledge of the sites well, and did not involve close coding of all field notes. See section III.D.a for more detail.

The human mind finds patterns so quickly and easily that it needs no how-to advice. Patterns just "happen," almost too quickly. The important thing, rather, is to be able to (a) see *real* added evidence of the same pattern; (b) remain open to disconfirming evidence when it appears. As Ross and Lepper (1980) point out, beliefs (in this case in the existence of a pattern) are remarkably resistant to new evidence. Patterns need to be subjected to skepticism—one's own or that of others—and to conceptual and empirical test (Does it really make sense? Do we find it elsewhere in the data where predicted?) before they represent useful knowledge.

VII.A.3 Seeing Plausibility

The faithful fieldworker Boswell (1791) reports that Dr. Johnson said, "Patriotism is the last refuge of a scoundrel." With good reason: A noble sentiment can easily be exploited for other purposes. There's a crude parallelism to the idea of "plausibility" as a last-refuge tactic for drawing conclusions.

Chart 41
Reasons Given for Adoption by Users

Reasons/Motives	*Number of Respondents Mentioning Item (N = 56)*
Administrative pressure, constraint	35
Improves classroom practice (new resources, relative advantage over current practice)	16
Novelty value, challenge	10
Social (usually peer influence)	9*
Opportunity to shape projects	5
Professional growth	5
Gives better working conditions	3
Solves problems	2
Provides extra money	1
Total	86

*Seven mentions were from one site.

It often happens during analysis that a conclusion is plausible, "makes good sense," "fits." If a colleague asks you how you came to the conclusion, or what you based it on, the initial answer is something like, "I don't really know. . . . It just feels right." Many scientific discoveries initially appeared to their authors in this guise; the history of science is full of global, intuitive understandings that, after laborious verification, proved to be true. So plausibility, and intuition as the underlying basis for it, is not to be sneered at.

But, as we have noted, people are meaning-finders, even in the most genuinely chaotic data sets. Patterns can be found even in random data, as the activities of numerologically obsessed people show. So plausibility can easily become the refuge of, if not scoundrels, analysts who are too ready to jump to conclusions. As we remark later, "Plausibility is the opiate of the intellectual."

During documentation of our own analysis efforts (as in section VII.C), we often found ourselves giving the "plausibility" basis for conclusions we drew, particularly in the early stages of analysis. Nearly always, it developed, the "plausibility" was an initial impression that needed further checking through other conclusion-drawing tactics, or through verification efforts. Plausibility in this sense was a sort of pointer, drawing the analyst's attention to a conclusion that looked reasonable and sensible on the face of it—but what was the real basis involved?

Here's a brief illustration. The analyst in our school improvement study was trying to order twelve school sites on "percentage of use" of the innovation—the number of teachers in a school or district eligible to use an innovation (such as a reading program) who in fact *were* using it. On the face of it this looks simple: the ratio of two numbers. The sites could be sorted into a few categories: full, moderate, minimal percentage of use. But as the analyst proceeded, life looked more complicated. His documentation comment was:

> The categories are OK, but they don't form an ordinal scale. Much depends on the size or scale of what is being attempted in the first place.

This was based, the analyst said, on "plausibility. . . . Seems clear on the face of it."

The next step was to get more systematic. The analyst went through several iterations of a chart displaying the data until Chart 42 appeared. In the course of this, the analyst also realized that percentage of use in the district and in the school were quite different things, and he clustered sites accordingly.

We can now see that the initial "plausibility" basis had to be supplemented by *clustering* (realizing that Astoria, Calston, and Lido were all dealing with a specialized population of students) and by *splitting a variable* (building-level and district-level percentage of use). The final four clusters of schools fall in an order that has a clear basis (substantial use in building and in district; full in building, less in district; moderate to full, but for a specialized, hence less demanding, population; and clearly minimal use). The analyst can also order the sites within each category.

So the moral is something like this: Trust your "plausibility" intuitions, but don't fall in love with them. Subject the preliminary conclusions to *other* tactics of conclusion drawing and verification.

Chart 42
Site-Ordered Meta-Matrix: Percentage of Use, by Sites

SITES by Percentage of Use	In the building: Years of use	# users in bldg.	Eligible users in bldg.	% of use	In the district: Yrs. of use	# users in dist.	Elig. users in dist.	% of use	Eligibility criteria	Remarks
Substantial										
Carson (IV-C)	3	20	20	100%	3	42	42	100%	All regular teachers in dist(1 elem, 1 HS)	Mandated as of Fall '79.
Masepa (NDN)	3	9	11	82%	4	36	43	84%	All tchrs, grades 3-7, 6 schools	Mandated as of April, 1980.
Tindale (IV-C)	4	29	36?	80%?	4	48	60?	80%?	Eng. math & sci. tchrs for lower-track students, 2 HSs. 60= max. of tchrs ever using innovation.	Not clear what maximum student population is.
Full in buildings, less in district										
Plummet (IV-C)	4	25	25	100%	4	N/A	N/A	?	Innovation is complete school; all present staff eligible.	Unknown whether school deals with all of potential target population (delinquents, 11 high schools).
Perry-Parkdale (NDN)	3	6	6	100%	3	N/A	N/A	?	Innovation is self-contained program; all present staff elig.	Program accomodates 3% of total jrs. and srs. from 2 HSs; unknown what "eligible" population is.
Banestown (NDN)	1½	3	3	100%	1½	10?	DK	?	Innovation is remedial lab; all lab staff eligible.	Used in 5 schls of large county system.
Moderate to full for specialized population										
Astoria (NDN)	1	5	5	100%	2	DK	DK	100%?	All 1st-grade tchrs & aides (92 schools).	Prog. mandated as of Fall '78; actual percentage of use throughout district unknown.
Calston (NDN)	2	2	2?	100%?	4	4	4--50	100% to 8%	Narrow definition; all intermediate teachers in schools with an interested principal (N=2)	Pct. of use for all 25 schools in district is 8%
Lido (NDN)	4	3	5	60%	4	3	5	60% to 8%	High school science teachers, in the one high school.	In princ. all 36 HS tchrs are elig. since innov. is interdisciplinary. Pct. use would thus be 8%
Minimal										
Burton (NDN)	1	1	5	20%	1	3	20?	15%?	All social studies tchrs, in 4 high schools.	Use defined as "experimental" during first year.
Dun Hollow (IV-C)	2½	2	13?	15%?	2½	3	84	3%?	All primary teachers (grades 1-3 in 7 elem. schools	Use defined as "field testing."
Proville (IV-C)	3	0	11 to 36?	0	3	0	44 to 144	0%	44 voc. ed. tchrs in 4 HSs and/or 100 classified personnel.	A discontinuation as of Spring '79.

? Missing or inconclusive data.

Incidentally, a somewhat more trustworthy tactic involves *lack* of plausibility. When a conclusion someone is advancing "just *doesn't* make sense," it's a bit safer to rule it out. But not completely safe. Counterintuitive or puzzling findings can sometimes be extraordinarily stimulating and rich, so they should be allowed their day in the sun too.

One final comment. Most conclusions drawn during analysis are *substantive*, based on the content. But the analyst is constantly drawing *procedural* conclusions along the way as well: to transpose two rows in a matrix; to add or discard a column; to collapse the data into a summary table; to change a decision rule for data entry. Our experience is that "plausibility" is often a reasonable guide for such decisions, and laboring to try other tactics, or to "verify" the procedural decision, is unnecessary. Of course, it's important, as we have said many times, to *show* the procedural decision made (the final matrix, the operative decision rules, and so on).

VII.A.4 Clustering

In daily life, we are constantly sorting things into classes, categories, bins: Something that doesn't move around but grows is called a "plant"; something that moves around and has babies is called an "animal"; something that moves around, has four wheels and an engine run by fossil fuels and carries people is called an "automobile." Most categories require *other* categories to define them: "wheel," "engine," "babies."

As we have already noted in our discussion of coding (section III.B), the qualitative analyst is looking to see, as LeCompte and Goetz (1983) put it, "What things are like each other? Which things go together and which do not?" The categories or classes used by the analyst may be preexisting (for school districts: urban, suburban, rural) or they may emerge from the data ("controlling," "directive," "facilitative," and "neglectful"), as found by Berman and Weiler in their study of districts in California school improvement projects (Degener, 1983). Typically, as Bulmer (1979) points out, they emerge from an *interaction* of theory and data.[4]

Clustering is a tactic that can be applied at many levels to qualitative data: at the level of events of acts, of individual actors, of processes, of settings/locales, of sites as wholes. In all instances, we are trying to understand a phenomenon better by *grouping*, then *conceptualizing* objects that have similar patterns or characteristics. Here are some illustrations.

Davis (1959), quoted in Lofland (1971), was studying the *acts* of cab drivers who were more interested in receiving a larger tip. They clustered this way:

fumbling in the making of change
giving the passenger a hard-luck story
making fictitious charges for services
providing a concerted show of fast, fancy driving
displaying extraordinary courtesy

At the level of *actors*, in our school improvement study we asked teachers to draw pictures of how their high school looked to them—which people and groups were involved. This led to categories of teachers such as "goners," "boy coaches," "girl coaches," and the "old new guard" (see section IV.A).

At the level of *processes*, we were able to cluster the activities involved in coping with the problems of later implementation of an innovation (Chart 28, section V.A):

reaching up
improving, debugging
refining
integrating
adapting
extending

These clusters took a good deal of summarizing and reworking before they came clear. A simpler example comes from our study of teachers' and administrators' job mobility, which fell rather easily into these categories:

moving in
moving out
moving up
moving in and up
moving out and up
moving over

It is also possible to make clusters of *settings* or locales. For example, in schools, we might sort places where people interact into these clusters:

formal instructional (classrooms, gyms)
informal instructional (library, club room)
formal adult work (meeting room, office)
informal adult association (teacher lunchroom, restroom, corridor)
mixed (cafeteria, playground)

Finally, as we have seen in the school improvement study, it's possible to sort *sites* as complex wholes into meaningful clusters. For example, in Box V.E.a, we show how twelve sites were sorted into three "families," according to the level of assistance provided and roughness/smoothness of early implementation, then the later degree of practice stabilization. Using more complex data—causal networks—as the base, section V.H shows how the twelve school sites could be clustered into four different scenarios: "casualties," "success-driven advancement," "opportunism," and "career crystallization."

We can see from these various examples that "clustering" is a general name given to the process of using and/or forming categories, and the iterative sorting of things—events, actors, processes, settings, sites—into those categories. Where lower-level, less complex things are being sorted (events, actors, and the like), the clustering tactic typically relies on aggregation and comparison ("What things are like each other/unlike each other?"), and is naturally closely interwoven with the creation and use of codes, both at the first level (section III.B) and the pattern-coding level (section III.C). As the analyst works at clustering processes, settings, and sites, the clustering operations become more and more complex and extended—just as sorting things into "animals" and "plants" is a (perhaps deceptively) simpler task than sorting various kinds of four-wheeled machines (automobiles, trucks, golf carts, airplanes, ski-lift gondolas, typewriter tables, and floor polishers) into sensible clusters.

The typical problem in making clusters at these more complex levels is that the entities being clustered have many attributes that you initially expect are relevant to the clustering task. A simple way to proceed is the "site-by-attribute matrix." Listing sites as rows and attributes as columns lets you see the whole picture. By inspection of the columns, you can find which attributes are critical in differentiating sites. Then "families" of sites can be formed by cutting up the rows of the matrix and re-sorting them. Each family shares the same set of critical attributes. See Miles, Farrar, and Neufeld (1983) for an example.

Clustering can also be seen as a process of moving to higher levels of abstraction (see section VII.B.7 on *subsuming particulars into the general*). Figure 23 illustrates this with a content-analytic "dendrogram" display for representing clusters, from Krippendorff's (1980a) excellent summary. It emerged from computer clustering of groupings made by 28 people, sorting 300 advertising appeals. Human analysts aren't so elegant, but the principle holds.

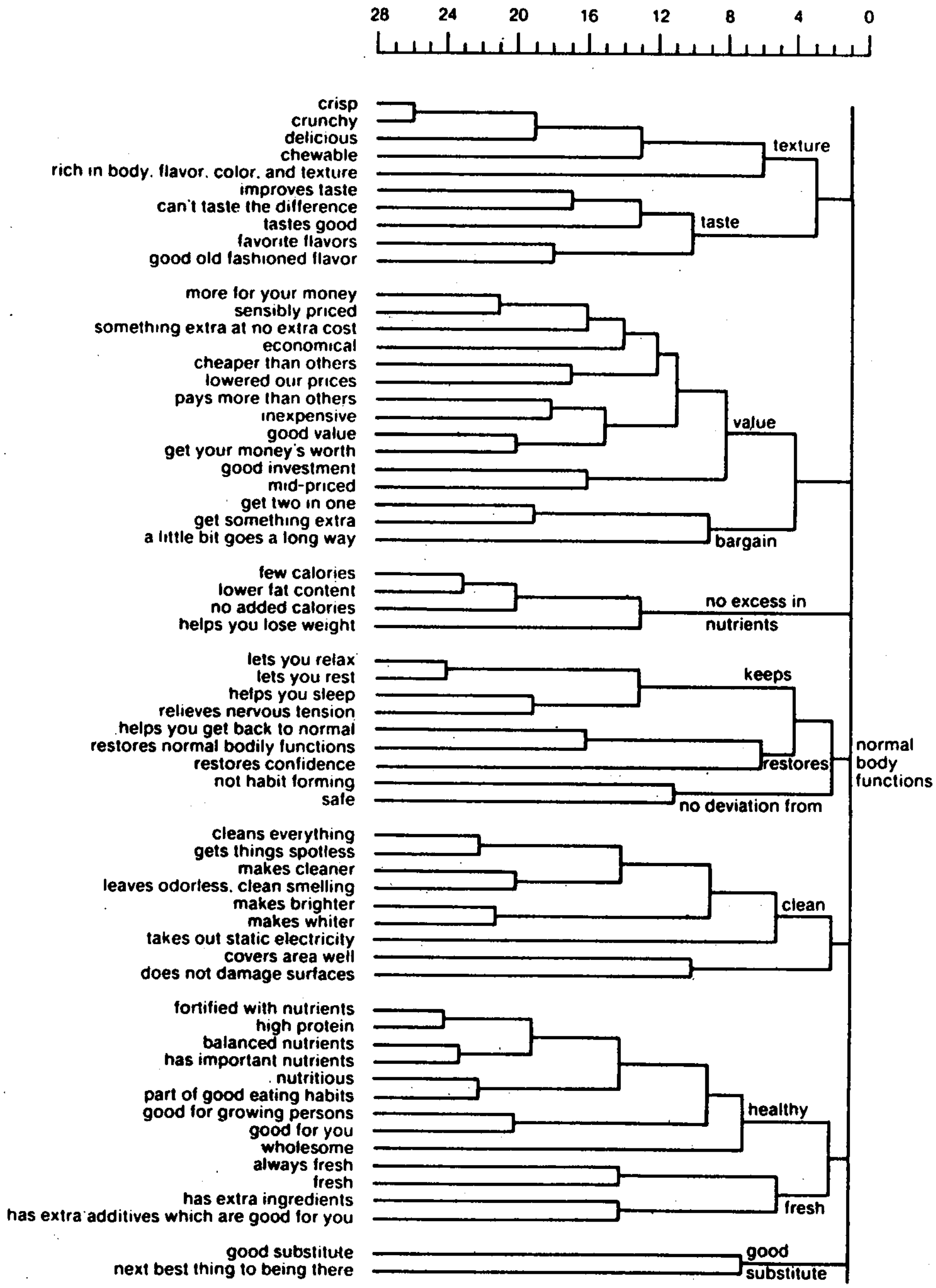


Figure 23 Illustration of Clustering, Using Dendrogram Method
SOURCE: Krippendorff (1980a). Reprinted by permission.

As this example suggests, the qualitative analyst need not assume that clustering techniques have to be completely self-invented. There is a long tradition of content-analytic techniques dealing directly with issues of coding, unitizing, and clustering qualitative data that can be very helpful (see, for example, Holsti, 1968, 1969; Berelson, 1971; Pool, 1973; Krippendorff, 1980b; Monge & Capella, 1980).

Clusters, like the results of other conclusion-drawing tactics, must be held lightly in the analyst's mind; premature closure is to be warded off. And they naturally need to be verified, as well. For example, does a cluster *"outlier"* (section VII.B.6) really belong to cluster A, or more appropriately to cluster B? Are the data being used fully *representative* (section VII.B.1) of the universe of events, actors, or settings being studied?

VII.A.5 Making Metaphors

Surprisingly, most of the case studies we have read are baldly matter-of-fact. The emphasis is put on description and, near the end, on interpretation and meaning, but both exercises tend to be taken literally, even actuarially. Most of this material is not—how shall we say?—riveting. Nor is it particularly illuminating in the way a piece of fiction can be. Though most novelists would cringe at being called "social scientists," many surpass the best qualitative researchers in communicating complex social reality. Consider Proust, Dreiser, Balzac, Austen, Nabokov, Updike, Oates, and Garcia Marquez. Their appeal is that they dramatize, amplify, and depict, rather than simply describe, social phenomena. The language itself is often figurative and connotative, rather than solely literal and denotative. Part of this has to do with the use of metaphors, analogies, symbols, and other allusive techniques of expression.

We are suggesting that qualitative researchers should not only *write* metaphorically, but also *think* metaphorically. For instance, it seems to us that "mother's separation anxiety" is a less appealing, less suggestive, *and* less theoretically powerful descriptor than is "the empty nest syndrome." Why do we say this? For one thing, the metaphor is richer, more *complete*. Instead of a clinical character trait (anxiety), we have allusions to an important environmental setting ("nest"), the idea of nurturance aching for a newly absent (but grown-up) object, and the acknowledgement that all this has taken a good deal of time up to now. Instead of focusing only on a mother's inner states, we can see new theoretical possibilities (maybe if socialization for independence was weak, the child will regress).

What else is true of metaphors? They are *data-reducing devices,* taking several particulars and making a single generality of them. For instance, the "scapegoat" metaphor pulls together facts about group norms, treatment of deviants, social rituals, and social rationalizations into one package. This is not to be sneezed at. Qualitative researchers looking at mountains of field note write-ups are grateful for *any* device that will reduce the bulk without locking out multiple avenues for analysis.

Metaphors are also *pattern-making devices*. For example, in the school improvement study, we found at one site that the remedial learning room felt like an "oasis" for the pupils sent there for part of each day. The metaphor "oasis" pulls together separate bits of information: The larger school is harsh (like a desert); not only can students rest in the remedial room, but they can get sustenance (learning); some resources are very abundant there (like water in an oasis); and so on. Such metaphors also help to place the pattern noted in the larger context (in this case, the harsh, resource-thin school).

Metaphors are also excellent *decentering devices*. You have to step back from the welter of observations and conversations at the field site and say, "What's going on here?" Since metaphors won't let you simply describe or denote a phenomenon, you have to move up a notch, to a slightly more inferential or analytical level. The remedial learning room doesn't *look* like an oasis, and perhaps no one is actually *describing* it that way, nor is anyone behaving literally like an exhausted Bedouin under a date palm.

Finally, metaphors or analogies are ways of *connecting findings to theory*. The "oasis" metaphor makes one think of how institutions develop compensating mechanisms to reduce the stress they put on role occupants, or of how they nurture as well as isolate deviants. Or one starts considering social control mechanisms more generally.

The metaphor is halfway from the empirical facts to the conceptual *significance* of those facts; it gets the analyst, as it were, up and over the particulars en route to the basic social processes that give meaning to those particulars. For instance, Glaser (1978) advises the field researcher struggling to make sense of social phenomena to attach metaphorical gerunds to them (such as servicing, bargaining, becoming). In so doing, one shifts from facts to *processes*, and those processes are likely to account for the phenomena being studied at the most inferential level.

For more on metaphor, see Ortony (1979), Lakoff and Johnson (1980), and Johnson (1981).

Now a few words of advice for metaphor-makers.

(1) Looking for metaphors too early in the study is dangerous. To begin with, it distracts one from the work at hand—collecting the information, following leads, observing closely. It also leads to making judgments on the basis of too few facts. Finally, it clamps down too quickly on the meaning of what one is studying. Psychologists call this "premature closure." You become all too quickly wedded to your metaphor; it *sounds* good, other people resonate to it (it sounds good to them, too); it makes you feel insightful. You start to look around less and to project the metaphor increasingly on things that are, at best, remotely

related to it. Better to wait until about two-thirds of the way through data collection, when you have a strong body of information to draw from, and still some time to test the validity of the image on the site during late visits.

(2) The best way to generate metaphors is to be cognitively playful, to say to oneself, "What's the gerund there?" or "If I only had two words to describe an important feature at this site, what would they be?" or "What are people doing here?" The trick is to move from the denotative to the connotative.

(3) Interaction helps. Groups stimulate their members' thinking by increasing the inputs, bringing in ideas from a new angle, making one decenter from one's cognitive ruts, and creating a contagiously playful thinking environment.

(4) Know when to stop pressing the metaphor for its juice. When the oasis starts to have camels, camel drivers, a bazaar, and a howling sandstorm, you know you're forcing things. Use it as long as it's fruitful, and don't overmetaphorize.

VII.A.6 Splitting Variables

"Westward the course of empire takes its way." So wrote George Berkeley in 1752 in *On the Prospect of Planting Arts and Learning in America*. To Berkeley, the main direction was obvious. It sometimes seems equally "obvious" that the course of qualitative data analysis is toward *integration*—ever and ever greater linkage among variables, expressed at a more and more powerful level.

But just as there was plenty of colonial push to the south and to the east, there are many times when differentiation is more important than integration. The analyst must say, "Stop!" That wonderful variable isn't really one variable, but two, or maybe even three. You must have the courage to question what might be called "premature parsimony."

Splitting variables can occur at many points during analysis. At the stage of initial conceptualization, it pays to "unbundle" variables rather than assume a monolithic simplicity. For example, the checklist matrix shown in section IV.B splits the general variable of "preparedness" to carry out an innovation into ten subvariables or components, ranging from states of the user ("commitment," "understanding," "skills") to the availability of materials and actions taken by administrators ("time allocation," "inservice").

When coding schemes are being developed and elaborated, splitting is often useful. For example (section III.B), our study originally had a code TR-ORG, used for instances of change in the school as an organization as a result of innovation use. Initial fieldwork showed clearly that this variable should be split to distinguish among organizational changes that were *practices*, such as staffing, scheduling, planning, use of resources, and an equally important, "softer" variable, organizational *climate* (norms, interpersonal relationships, power, social networks).

When matrix formats are being designed, variable-splitting is also a very useful strategy; more differentiation lets you see differences that might otherwise be blurred or buried. For example, in box IV.F.a, the analyst who wanted to study and explain the effects of various types of assistance to teachers realized very early that the variable "effects" of assistance should be separated into *short-run* effects (the user's immediate "state" after receiving assistance) and *longer-run* effects (what the user was able, or unable, to *do* as a result of the assistance).

Similarly, as conclusion drawing proceeds, the analyst will often realize that a variable needs to be split. In section V.A, where the issue was the feelings and concerns of innovation users, the analyst was struck by the fact that some concerns, as originally expected, were *individual* ones, such as discouragement, fatigue, or friction with others, and many other concerns were essentially *institutional* in nature, such as changes in district priorities, poor overall functioning of the project, or uncertainties about project continuation. This led to the summary tabulation in Chart 27.

Sometimes creating a two-variable matrix, as in Box V.B.c, helps to clarify whether splitting a variable will be illuminating or not. In this case, the analyst was struggling with the relationship between user "practice stabilization" and the "continuation" of the innovation locally. The first look at continuation emphasized users' *attitudes* to continuation as the indicator. That made sense: A stabilized innovation ought to evoke more positive attitudes on the part of its users. But the relationship was quite weak. Users from almost all sites had positive attitudes to continuation; only in very low-stabilized sites were there negative attitudes. This suggested splitting the "continuation" variable into two parts: users' *attitudes* toward continuation and their estimates of the *probability* of continuation. Adding a column to the matrix led the analyst to pay dirt. The probability variable turned out to illuminate the relationship between stabilization and continuation more clearly than the attitude one, because it took into account factors beyond individuals, such as district support, staff turnover, and funding difficulties.

In sum, when is variable-splitting a good tactic? The first answer is: Split in early stages (conceptualizing, coding) to avoid monolithism and blurring of data. The second answer is: Split a variable when it is not relating as well to another variable as your conceptual framework (or other available data) have led you to expect.

Finally, we should remark that variable-splitting is not a virtue in itself. Extreme differentiation leads to complexity and atomization, poor mapping of events and processes. When you split a variable, it should be in the service of finding coherent, integrated descriptions and explanations.

VII.A.7 Subsuming Particulars into the General

Clustering, as we have seen, involves clumping things together that "go together," using single or multiple dimensions. That process is often an intuitive, first-level process, corresponding to ordinary coding (section III.B).

A related tactic involves asking the question, "What is this specific thing an instance of? Does it belong to a more general class?" This tactic corresponds to many of the types of "pattern code" we outlined in section III.D. The analysis is taking a step up, trying to locate the immediate act, event, actor, or activity in a more abstractly defined class. That class may have been predefined, or it may have emerged as a result of memoing (section III.D). Here is an illustration.

LeCompte (1974), looking at a classroom, noted that the following events fell into a cluster:

- Teacher vacuums room twice daily.
- Students cannot leave in afternoon until they wash their desks.
- Teacher refers to personal hygiene when reprimanding students.
- Teacher says, "Children who come to school looking like that [soiled shirt] you just can't expect to do the same kind of work as the others."

At first glance, this is simply a cluster of behaviors dealing with cleanliness. But LeCompte also noticed that children's statements about the cleanliness issue tended to parallel the teacher's. Furthermore, most of the items (a) recurred and (b) had a strong regulatory emphasis. She concluded that the general class here was "cleanliness norm"—a shared set of standards about appropriate and inappropriate behavior.

Glaser (1978) uses this tactic during his "constant comparative method," looking for "basic social processes," such as negotiation or bargaining, that are a more general class into which specific behaviors (for example, bickering, arguing, refusing, offering, soothing) can be subsumed.

In our school improvement study, we noted specific statements made by teachers and administrators, such as:

- If you want to depart from the guide, ask me and also tell me why you want to do it and how it will fulfill the guide's objectives.
- The basic philosophy is there, but the use of (the innovation) is flexible, and doesn't require use of all units.
- In this program you're like a robot . . . but I learned that if I wanted to change something I would just go ahead and do it. . . . I learned to cut corners and do it just as well.

These statements can be subsumed into a more general class: The presence of high or low *administrative latitude* given to teachers to adapt or alter an innovation, a variable that turned out to be very important in explaining the amount of adaptation that occurred.

Note that the process of moving up a step on the abstraction ladder is not a mechanical or automatic process. For example, all three statements above could be seen as an instance of the more general class "adherence to objectives," if one attends to such phrases as "the guide's objectives," "basic philosophy," and "do it just as well." Here the analyst was convinced, through the presence of many other statements, that the question of administrative action restricting or permitting latitude was the more important general class.

In short, subsuming particulars into more general classes is a *conceptual* and theoretical activity (Glaser, 1978) in which the analyst shuttles back and forth between first-level data and more general categories, which evolve and develop through successive iterations until the category is "saturated" (new data do not add to the meaning of the general category).

Arbitrary movement up the abstraction ladder gets a researcher nowhere. Suppose that you observed a teacher writing his name on the blackboard on the first day of school. That specific action can be subsumed in a larger class of "written communication," then in a larger class of "information transmission," then in a still-larger class of "human action." That is a sort of taxonomic classification without useful meaning, however. Depending on the purposes of the study and the assumptions of the researchers, the action might more fruitfully be classified as an instance of "thoroughness of preparation," "reassurance/anxiety reduction," "group control through legitimacy," or "institutional perpetuation."

One cannot decide in a vacuum which of these classes is "right" or "best." There must be a clear linkage to the study's conceptual framework and research questions. *And* one must repeatedly move back down the abstraction ladder, as Korzybski (1933) always counseled, to "find the referent": the concrete instance that's being alluded to when a phrase such as "institutional perpetuation" is being used in an analysis.

VII.A.8 Factoring

"Factoring" comes from factor analysis, a statistical technique for representing a large number of measured variables in terms of a smaller number of unobserved, usually hypothetical variables. These second-order variables ("factors") may be largely uncorrelated, or have some "communality," overlapping with each other. In either case it is possible to identify general themes, giving a name to the statistically identified factor or factor cluster. What is the qualitative researcher's analogue?

Most of the tactics covered here are designed to do two things: to reduce the bulk of data *and* to find patterns in them. Clustering, making metaphors, moving up the abstraction ladder are all pattern-forcing exercises. The task is essentially that of saying to oneself, "I have a mountain of information here. Which bits go together?" When I derive a pattern code (section III.C), I am really hypothesizing that some disparate facts or words *do* something in common or *are* something in common. What they do or are is the "factor," and the process by which we generate it is "factoring." Time for an illustration.

Our study concerned interorganizational arrangements linking a university or college with a set of surrounding schools. At one successful site in a midwestern state, we collected some study-relevant characteristics of the state college. Most were coded "COLL-CHARS," so we could easily yank them out and list them. Here is the list:

- service is a central objective
- few contacts with state-level agencies
- small-scale projects
- low publishing rate by staff
- little research activity
- numerous outreach activities to surrounding schools
- concern for following up on preservice training with in-service training
- college staff active on community councils
- use of area teachers as resource people in activities
- majority of within-state college staff
- few cross-university discipline contacts

How can one "factor" the list? The reader might like to scan these items to see what factor might underlie them. Then we'll share our try.

Let's list the items again and see how the analyst, in the course of successive scans, identified the commonalities (shown between slash marks).

- service is a central objective / activist, client-centered /
- few contacts with state-level agencies / not looking beyond own region /
- small-scale projects / operating on scale of community /
- low publishing rate by staff / service over academic orientation /
- little research activity / *idem* /
- numerous outreach activities to local schools / service again; localism, too /
- concern for following up on preservice training with in-service training / keeping connected to graduates in local area /
- college staff active on community councils / service; *local* investment, again /
- use of area teachers as resource people in activities / practical focus; nonacademic orientation, again /
- majority of within-state college staff / not looking beyond region, again /
- few cross-university discipline contacts / *idem*; also nonacademic orientation /

The same themes recur: Activism, client-centeredness, localism, nonacademic orientation. We settled on a double-barreled theme we called "localism/activism." All of the characteristics fit into this theme without a conceptual shoehorn. From these eleven particulars, which most of the people we observed or interviewed had in common, we had derived a general characteristic.

We also noted that the large state university, another site in the same state, had different characteristics. It, too, was activist, but it was *not* localist. Its principal client was the state, not the city in which the university was located. There was more cross-university contact, and publishing was higher. So perhaps we had a bipolar factor in our study, "localism versus cosmopolitanism" with the effect of activism "controlled," analogically speaking. Now we are factoring at a slightly higher level of abstraction, using multiple sites to do a sort of "second-order factoring." It gets us to still fewer overarching themes or constructs that subsume bigger chunks of the data. We get there by asking the question, "What is there a *lot* of in one place that there is *little* of in another—and are they comparable things?"

Then comes the more consequential question, "Do these contrasts make any *meaningful* difference, or are they essentially decorative?" The factors have to contribute to our understanding of the case or of its underlying dynamics. Otherwise, they are no more useful than the big, gift-wrapped boxes that unpack into a succession of smaller, but equally empty gift-wrapped boxes, leaving us at the end with a shapeless heap of ribbon and cardboard.

In this case, the localism/cosmopolitanism factor, when associated with the activism factor, was very helpful in explaining why some school-university arrangements were very successful and others limped along.

VII.A.9 Noting Relations Between Variables

The idea of relations between variables has already been prefigured in our discussion of conceptual frameworks (section II.A), which we suggest can economically be thought of as sets of boxes and arrows; boxes are variables, and the arrows show relationships between them. Once an analyst is reasonably clear about what variables are in play in a situation, the natural next query is, how do they relate to each other?

What sort of relations can we envision between two variables A and B? A variable is something that *varies*. Thus we might have:

(1) A+ , B+ (both are high, or both low at the same time)
(2) A+ , B− (A is high, B is low, or vice versa)
(3) A↑ , B↑ (A has increased, and B has increased)
(4) A↑ , B↓ (A has increased, and B has decreased)
(5) A↑ then → B↑ (A increased first, then B increased)
(6) A↑ then → B↑ then A↑ (A increased, then B increased, then A increased some more)

(These don't cover all the possible permutations, of course.) Relationship 1 is a direct association: Both variables are high (or low) at the same time. For variables that are "all or nothing," this relationship can be read as follows: When A is present, B is also present, or both may be absent. Relationship 2 is "inverse." With relationship 3, we are noting that *changes* have recently occurred in A, and in B, in the same direction; 4 is the inverse. No claims are necessarily being made that the changes are linked; they are just present. In relationship 5, we verge toward causality: A has changed, *then* B changed (and—not shown—there is a reasonable belief that A "could" have caused B). If A is an evening of heavy drinking and B is a headache the next morning, there's a presumptive connection. But there's probably little likely connection—in most cases—if B is a morning headache and A is the announcement of the new city budget. (Still, if the headache belongs to the Mayor, maybe . . .)

Finally, in relationship 6 we see a mutual ("non-recursive") relation: A change in A leads to a subsequent change in B, then to a subsequent change in A. Of course the strength of these associations can vary: We can have decisive, strong, clear relationships—or feeble, weak, ambiguous ones.

The basic analysis tactic here involves trying to discover what sort of relationship—if any—there is between two (or more) variables. The important thing to keep in mind is that we are talking about variables, concepts, not necessarily specific acts or behaviors. Even when we are focusing on specific *events*, there are usually underlying or more general variables involved. The event of an evening of heavy drinking and the event of the morning headache do not quite affect each other directly. All sorts of variables are at work: the presence of certain esters in the beverage involved and the body's ability to metabolize alcohol, the amount consumed, the time intervening, and so on. So, we concur with Glaser (1978), who says that the researcher is "making theoretical statements about the relationship between concepts, rather than writing descriptive statements about people."

How are relationships detected? We argue in this book that matrix displays are an especially economical way to see them: Data bearing on two or more variables can be arrayed for systematic inspection, and conclusions drawn.

For one illustration, consider the relationship we wanted to explore between the "centrality" of an innovation to a person's interests and the initial attitude held toward it. Are people more favorable to a "central" innovation than one that is more superficial in their life space? Perhaps, indeed, we might find that people are more negative when an innovation looms large in their scheme of things. Turning to section IV.E, Chart 16c, we can see a matrix displaying data on these two variables (and others). Scanning down the columns for "centrality" and "initial attitude toward program," we can see that there is indeed some relationship, though it is not a tight one. For users, the centrality is always high and initial attitudes are neutral or favorable; only one is unfavorable. And it looks as if early apprehension is at work with the "neutral" teachers.

We cannot know what the consequences of moderate or low centrality are for teachers, because we have no such cases. There is more of a range of centrality for administrators, but no clear pattern in terms of attitudes. So we have only modest support for the idea of a positive relationship. The analyst might want to go back to the field and hunt for some low-centrality teachers to fill out the picture—and/or turn to other associations with positive attitude—such as "career relevance," in Chart 16c. There the range is wider—and it looks as if there may be a clearer relationship: For the teachers (and administrators) where there is no career relevance, the attitudes are neutral or unfavorable; in the case where career relevance grows (Weelling), attitude becomes more positive; and for two of three teachers who have high career relevance, initial attitude is positive.

That example shows how pairs of variables were associated within one site. For an illustration that uses data from multiple sites, see section V.C, Chart 32, where the analyst was trying to relate a number of variables, such as user commitment and user understanding, to another variable, "ease of early use." He concluded, for example, that user commitment had a stronger relationship with ease of early use than did user understanding. This conclusion was based on how fully "in place" the conditions were prior to early use, and on the assessment of the conditions' "facilitating" or "barrier"-like status (note that the display includes causal judgments, marked as F and B, made by the researcher, based on the field data).

For a third example, also with multiple sites, turn to section V.E, where we see a scatterplot (Figure 15).

The researcher thought that an administrator's early pressure on teachers to adopt an innovation might lead to the negotiation of a sort of compensatory bargain, where the administrator agreed to allow for adaptation—gave "latitude"—in return for the teachers' willingness to proceed. But the scatterplot in Figure 15 shows otherwise: The initial pressure to adopt seems to be accompanied by *low* latitude. It's the sites with initial *weak* adoption pressure where the administrator allows a good deal of adaptation (Perry-Parkdale, Burton, Plummet, for example).

Note, though, that the analyst picked up on the issue of how things evolved during later implementation (see arrows), pointing out that in four of six high-pressure sites, latitude increased slightly during later implementation. Thus, he concluded, "Perhaps there is something to our hypothesis . . . though the negotiation comes later than we had thought."

As we have noted at a number of points, people tend to think in causal terms. The risk in trying to understand relationships between two variables is jumping too rapidly to the conclusion that A "causes" B, rather than that A happens to be high and B happens to be high. Here it helps to shift to verification tactics, such as proposing and *checking out rival explanations* (section VII.B.10), *ruling out spurious relations* (section VII.B.8), or *using extreme cases* (section VII.B.7).

Drawing in skeptical colleagues to use one or more of these tactics can be very useful. One friend of ours says that any causal statement made about a social situation should immediately be reversed, to see whether it looks truer that way:

> "The students are late to class because they hate the teacher." (resistance driven by dislike)
>
> "The students hate the teacher because they are late to class." (lateness, caused by other reasons, leads to dislike—perhaps mediated by the teacher's reactions to tardiness)

That example may sound a little fanciful, but the reversal exercise is useful. In our school improvement study, as we note in section VII.A.12, we considered this statement:

> "Teacher involvement and commitment lead to more effort in using the innovation."

We entertained the reverse:

> "High teacher effort leads to teacher involvement and commitment."

This made good theoretical sense, in terms of cognitive dissonance theory. And the data had already shown us several examples of sites where early strong teacher effort led to later increases in commitment.

VII.A.10 Finding Intervening Variables

It often happens during analysis that two variables that "ought" to go together according to your conceptual expectations or your preliminary understanding of events at the site have only a tepid or inconclusive relation. A puzzle encountered almost as frequently is the case of two variables that *do* go together, but without making much sense. The analyst cannot quite figure out *why* they go together.

In both of these conditions, looking for *other* variables that may be in the picture is a useful tactic. Perhaps a third variable Q is confusing, depressing, or elevating the relationship between A and B, so that if you "controlled" for Q, the relationship between A and B would become clearer:

$$\begin{array}{c} Q \\ A\text{———?——}\rightarrow B \end{array}$$

Or perhaps the third variable actually fills out a reasonable chain, mediating or linking between A and B:

$$A \rightarrow Q \rightarrow B$$

Let's take the last case first. In section V.G, we found that school sites that had adopted innovations bearing large associated funding changed their organization structures and procedures more than those adopting less well-funded innovations. That leaves a great deal unexplained. As we noted, you can't change organizations just by throwing money at them. Why should it be that a well-funded innovation "induces" more organizational change?

In this case, the analyst created a site-ordered matrix of other possible correlates of organizational change, ranging from "environmental pressure" to "problem-solving orientation," "implementation requirements," and "administrative support." A careful scan showed that the original relation (Figure 24a) could be understood much more realistically when several other variables entered the picture (Figure 24b). Here we see that "size of funding" is part of a web of other variables. Larger innovations (1) carry more funds with them (2). The funds increase the support administrators give (4), but so do the heavier implementation requirements (3) of larger innovations. Organizational change (6) comes from at least three sources: the direct requirements of the implementation itself (3), administrative support (4), and the degree to which implementation is successful (5). As we can see, administrative support is a very central intervening variable.

In this case, the effort to clarify a plausible but puzzling relationship led to a much clearer—if more complex—formulation. Simpler cases of finding intervening variables also exist. For example, we found that *administrative pressure* to use an innovation was associated with its eventual *institutionalization*. That looked puzzling, until we discovered that sites where there was administrative pressure were also those

Figure 24a Two-Variable Relationship

where *organizational changes* had occurred—and these in turn supported institutionalization.

Now let's turn back briefly to the other sort of puzzle: where the relationship between A and B is tepid when it "ought" to be hot. The tactic here is also to examine a series of other candidate variables that may be "depressing" or confusing the relation between A and B. Chart 32 (section V.C) shows that *preparedness* of innovation users (commitment, understanding, resources, and so on, summed to a general variable) is not a very brilliant predictor of *smoothness* of early implementation (note, for example, the fairly high preparedness scores in the rough-starting sites of Carson, Dun Hollow, and Plummet). So getting people ready and prepared is not enough to guarantee early smoothness. What else might be in the picture?

The analyst went back to the written-up field notes, and located a series of five other variables that might affect smoothness: Were users pressured to adopt? Did the innovation fit their previous ways of working? How much did the innovation demand of them? Did they have latitude for making changes? How big was the innovation? These were displayed in a site-ordered predictor-outcome matrix (Chart 33). Presto! It came very clear that the main issue making for smoothness/roughness was the *size* of the innovation. Even if people were well prepared, if the innovation was large and demanding, early implementation was rough.

We might note that finding intervening variables is easiest when there are multiple examples of the two-variable relationship to look at, contrast, and compare. In our illustrations, these were *sites*. But the same principle can be used to examine multiple instances of *actors*, *settings*, or *events*.

VII.A.11 Building a Logical Chain of Evidence

We have been talking a good deal about patterns, metaphors, clusters, and themes. In these cases, what happens is that discrete bits of information come together to make a more economical whole that, analytically speaking, is more than the sum of its parts. How do we actually *accomplish* this? Is there some kind of heuristic or algorithm we can use?

Let's begin with an illustration. In the study of interorganizational arrangements between schools and universities, we happened upon one case that looked particularly successful. It was a "teacher center," connected to a rural state college and undertaking a variety of in-service training activities for schools within a radius of some 60 miles.

We tried to develop a logical chain of factors that could be leading to success, as seen from the state college side and from the school side. Figure 25 shows them. The logical chain of evidence goes like this. The state college might regard service and outreach activities as very central (1). Since that is in fact so, we would expect college staff to see the benefits (2) of a teacher center as high (they did). That should in turn lead to higher resource commitment (3) to the center; such commitment was found in the form of money and staff. Looking at the school side, we found few other opportunities for in-service help (5), and a shortage of good teaching materials (6); both of these should lead to high perceived benefits (7) from using the center—*if* the center did in fact give good in-service help and provide new materials. As it turned out, the high resource commitment did permit that; teacher center assets (4) and extent of use (8) were high.

This is, of course, a stripped-down version of a far more intricate case. But it does illustrate how one builds an evidential chain. To come out with such a chain some minimal conditions have to be met. For example, *several* informants with *different* roles have to *emphasize* these factors *independently*, and *indicate the causal links*, directly or indirectly (for example, "We didn't have any other facility to go to to find out about new materials, so the center looked good" = the link between 5 and 7). The researcher has to *verify the claims* (for instance, the actual funds committed, the lack of alternative resource sources, the activities actually undertaken). *Countervailing evidence* has to be accounted for.

How does building a chain of evidence differ from the "causal network" method we've already described in section IV.J? This approach is more tactically, specifically oriented. Building a chain of evidence is more painstaking at each step, more demanding of variable-to-variable logical coherence. There are more occasions of "If that were true, we should find X. We do find X. Therefore . . . "

Furthermore, *the relationships have to make sense*; the logical basis for the claim that "perceived benefits"

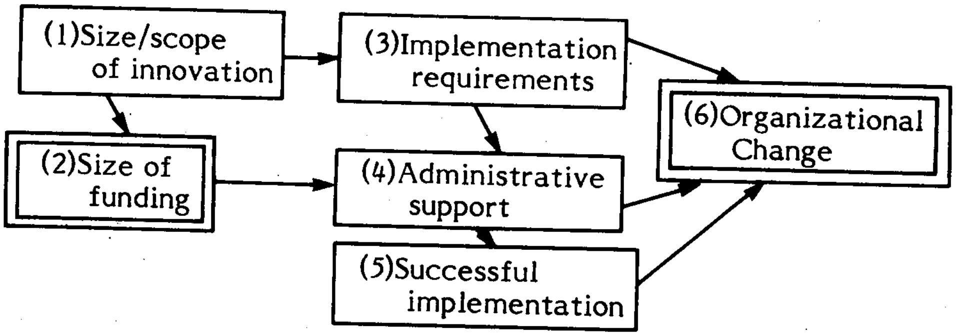


Figure 24b Two-Variable Relationship with Intervening Variables

translate into "resource commitment" must be incontrovertible. The *chain must be complete*; the stream from antecedents to outcomes should stand alone. For instance, in the figure shown, the link between 3 and 4 is not so obvious. One might have committed resources and come up with another model or another center with different characteristics. Committing resources doesn't translate automatically into, let us say, craft-oriented resource materials. Something is missing.

The field researcher constructs this evidential trail gradually, getting an initial sense of the main factors, plotting the logical relationships tentatively, testing them against the yield from the next wave of data collection, modifying and refining them into a new explanatory map, which then gets tested against new cases and instances. This is the classic procedure of analytic induction, well codified by epistemologists. It has been used in qualitative research, such as Lindesmith's (1947, 1968) celebrated studies of opiate addiction, to make a case for the necessary and sufficient causes of social behavior. See also Smith and Manning (1982, pp. 280-295) for more on this method.

At its most powerful, the method uses two interlocking cycles. One is called "enumerative induction," in which you collect a number and variety of instances all going in the same direction. The second is called "eliminative induction," in which you test your hypothesis against alternatives and look carefully for qualifications that bound the generality of the case being made. When qualitative researchers invoke "progressive focusing," they are talking about enumerative induction, and when they get into "constant comparisons" and "structural corroborations," they are switching into a more eliminative inductive mode of work. The "modus operandi" logic used in several professions as a troubleshooting device—for forensic pathologists, garage mechanics, clinicians, detectives, classroom teachers—is a good example of a back-and-forth cycling between enumerative and eliminative induction.

VII.A.12 Making Conceptual/Theoretical Coherence

When you are trying to determine what someone's behavior "means," the mental exercise involves connecting a discrete fact with other discrete facts, then grouping these into lawful, comprehensible, and more abstract patterns. With the preceding tactics, we are moving progressively up from the empirical trenches to a more conceptual overview of the landscape. We are no longer dealing just with observables but also with unobservables, and are connecting the two with successive layers of inferential glue.

The next, perilous step, is to move from metaphors and interrelationships to *constructs*, and from there to *theories*. That is, we need to tie the findings of our study to overarching, across-more-than-one-study propositions that can account for the "how" and "why" of the phenomena under study. Let's illustrate this abstraction-making process.

In the school improvement study, we came to notice that people at some of the field sites were literally exhausting themselves in the course of using new instructional practices. We also found that these people were making strong claims that the practice had substantially improved reading scores or children's attitudes toward school. The interesting part was that hard data to substantiate the claims were either nonexistent or gave little support to the claims.

These are the "facts" from which *we*, as analysts, made a pattern. Field site informants could—and did—agree with the facts, but they didn't put them together as we did. To some extent, we were only able to *see* the pattern because things were happening

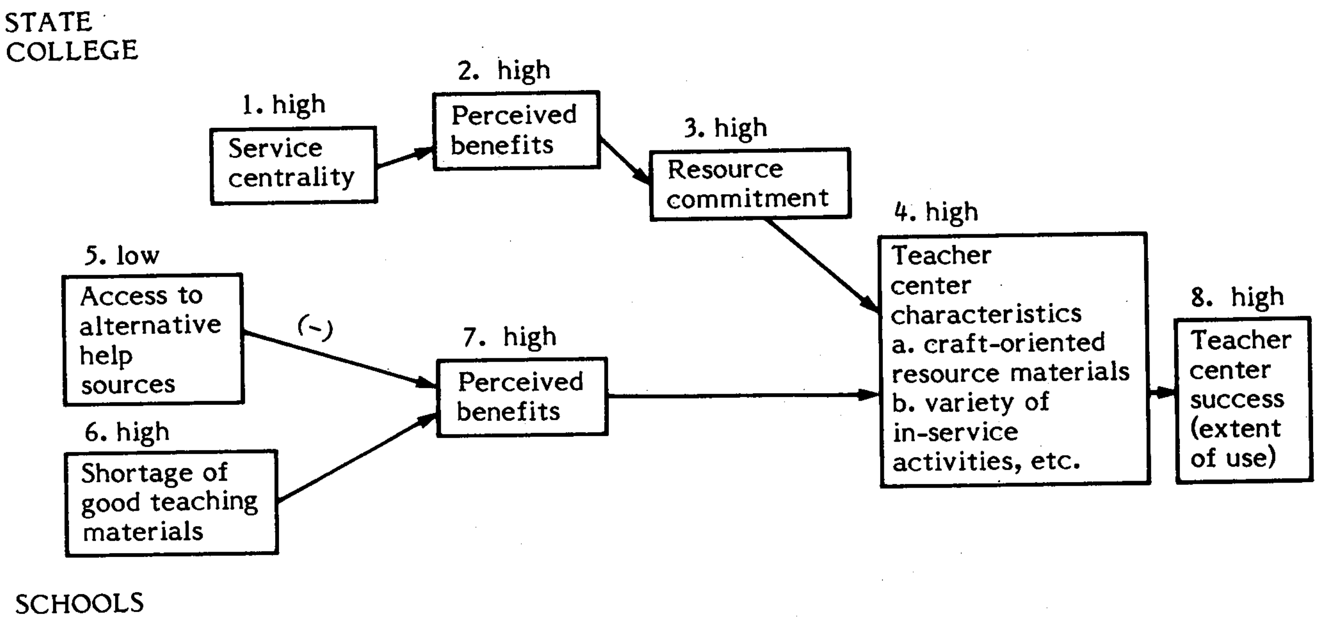


Figure 25 Example of a Chain of Evidence Supporting an Observed Outcome

NOTE: (–) = inverse influence.

otherwise at *other sites*, such as less investment, fewer claims, or more accurate claims. This, of course, is where multisite field research is so useful; it provides contrast and variance.

Let's call the pattern we saw "consensual delusion"—everyone agrees that more is happening than really is. The metaphors might be "groupthinking," or "self-deluding" or "wish fulfilling." We could also sketch out a logical flowchart like the one shown in Figure 25 to get a fix on *how* this happens at the several sites. But we're still within the confines of our study; we can't converse meaningfully with other social scientists except in the sense of reporting replication of an existing finding.

The analyst now has to ask, "Are there any broader constructs that put these facts together the way I'm putting them together?" In principle, as we noted earlier, the metaphors should help. The first one points toward group behavior, the next one toward individual cognitive process, and the third one toward motivational dynamics.

In this instance, we picked up an appropriate and powerful construct from cognitive social psychology: effort justification (Lawrence & Festinger, 1962). In order to justify the effort expended, one "sees" more results than are objectively there. This led us on into the domain of cognitive dissonance, and how people learn to love that for which they have suffered.

Where does this get us? For starters, it tells us that our finding has a conceptual analogue, which lends more plausibility to the finding *and* to the concept, which is now empirically grounded in a new context. It also helps to explain *why* such a pattern occurs. And it throws light on larger issues (such as how people, at our sites and more generally, cope with uncertainty). Finally, the construct can be trained back on our sites to explain related but puzzling phenomena. For example, we can now see *why* objective criteria (the test scores) are being systematically ignored when, in fact, they are widely available.

We have progressed here from the bottom up—from the field to the concepts. The steps are: (a) establishing the discrete findings, (b) relating the findings, (c) naming the pattern, and (d) identifying a corresponding construct. As we showed earlier in the book (Chapter II), it's perfectly legitimate, and often desirable, to work from the top down, from a conceptual framework to the collection of information testing its validity. Of course, you have to stay open to the idea that the concept is inapplicable, or has to be bent or discarded when you see the data. Concepts without corresponding facts are hollow, just as facts without concepts are, literally, meaningless.

VII.B. TACTICS FOR TESTING OR CONFIRMING FINDINGS

Having spent some time on particular ways of making and interpreting findings at different levels of inference, we now confront the issue of validity. As we have said often, qualitative analyses can be evocative, illuminating, masterful, and downright *wrong*. The data, looked at more scrupulously, don't support the conclusions drawn. Researchers double-checking the site come up with different findings. Site informants, asked to provide feedback on the findings, contest some or all of them, very plausibly. The phenomenologist chuckles, reinforced in the idea that there is no single reality out there to get "right." The psychometrician, from the polar opposite stance, concludes the nonstatistical research is an albatross.

The problem, as we have discussed in more detail elsewhere (Huberman & Miles, 1983a), is that there are no canons, decision rules, algorithms, or even any agreed-upon heuristics in qualitative research to indicate whether findings are valid and procedures robust. This is changing, but very slowly. Some recent work (Guba, 1981; Dawson, 1979, 1982) has provided general guidelines, and we are doing our best in this book to operationalize ways of testing/confirming findings.

Let's take a longer view of the problem at hand. Most qualitative researchers work alone in the field. Each is a one-person research machine: defining the problem, doing the sampling, designing the instruments, collecting the information, reducing the information, analyzing it, interpreting it, writing it up. A vertical monopoly. When we read the reports, they are most often heavy on the "what" (the findings, the descriptions) and wafer thin on the "how" (how one got to the "what"). In most cases, we don't read how, exactly, the researcher got from 500 pages of field notes to the main conclusions drawn, and we don't know how much confidence we can place in them. Researchers are not being obtuse. It's just that they have very little that is systematic to draw upon.

The other part of the problem appears when we look at research on information processing by individuals. There's a long research tradition (Meehl, 1954, 1965; Goldberg, 1970; Dawes, 1971) showing that human judgments are consistently *less* accurate than statistical/actuarial ones—and even that "expert" judges are worse than untrained ones (Taft, 1955). Oskamp (1965) found that experts felt more and more *confident* of their erroneous judgments as they got more accurate information, without modifying the initial judgments. (Think of the lone fieldworker out there, trying to collect trustworthy data, with few or no actuarial procedures to help.)

Why are such errors made? There's a long tradition of research on how people attribute causality (Heider, 1944; Kelley, 1967, 1972) and form judgments (Kahnemann & Tversky, 1972; Nisbett & Ross, 1980). The most recent findings, baldly put, are that most people are rotten scientists, relying heavily on preexisting beliefs, and making bias-ridden judgments. They don't act like scientists: They don't keep track of frequencies, make probabilistic estimates, sample representatively, or make accurate deductions.[5]

What these works don't tell us is when and how scientists behave like laypeople. The suspicion is that they do so often. For example, Kahnemann and Tversky (1972) have studied the "availability heuristic": Data that are "vivid" rather than "pallid" tend to be noticed, registered, retrieved, and used more frequently. As a result, one overweights the importance and frequency of concrete, immediate inputs. One personally experienced or witnessed dramatic event "means more" than several one has read about. Think now of the fieldworker who witnesses a crisis or a dramatic conflict on site, then makes inferences about the significance of the event—and still other inferences about what the site is like when he or she is absent.

So we need to be especially watchful in qualitative research about the multiple sources of potential analytic bias that can weaken, or even invalidate, our findings. Some of these biases have been identified in mainstream anthropological textbooks, along with some pointers for avoiding them (for example, see Wax, 1971). The archetypical ones include the following:

(1) *the holistic fallacy:* interpreting events as more patterned and congruent than they really are, lopping off the many loose ends of which social life is made
(2) *elite bias:* overweighting data from articulate, well-informed, usually high-status informants and underrepresenting data from intractable, less articulate, lower-status ones
(3) *going native:* losing one's perspective or one's "bracketing" ability, being coopted into the perceptions and explanations of local informants

It's useful to note that each of these three biases corresponds, respectively, to one of the three major judgmental heuristics identified by researchers: "representativeness," "availability," and "weighting." We recommend Nisbett and Ross (1980) as a good guide to this literature, and a help in avoiding self-delusion.[6] We'll also draw on this literature in discussing tactics for testing/confirming findings, a task to which we now turn.

The twelve tactics for confirming conclusions are numbered as before. They begin with tactics that are aimed at assuring the basic quality of the data, then move to those that check findings by various contrasts, then conclude with tactics that take a skeptical, demanding approach to emerging explanations.

Data quality can be assessed through *checking for representativeness* (1); *checking for researcher effects* (2) on the site, and vice versa; and *triangulating* (3) across data sources and methods. These checks may also involve *weighting the evidence* (4), deciding which kinds of data are most trustable.

Looking at differences tells us a lot. *Contrasts/comparisons* (5), *checking the meaning of outliers* (6), and *using extreme cases* (7) are all tactics that test a conclusion about a "pattern" by saying what it's *not* like.

How can we really test our explanations? *Ruling out spurious relations* (8), *replicating a finding* (9), *check-out out rival explanations* (10), and *looking for negative evidence* (11) are all ways of submitting our beautiful theories to the assault of brute facts, or to a race with someone else's beautiful theory.

Finally, a good explanation deserves attention from the very people whose behavior it is about—informants who supplied the original data. The tactic of *getting feedback from informants* (12) concludes our list.

VII.B.1 Checking for Representativeness

When we come up with a "finding" in a field study, we quickly assume it to be typical, representative, an instance of a more general phenomenon. But *is* it? And, if it is, *how* representative?

In their numerous studies of the "representativeness heuristic" used by individuals to calculate base rates and sample sizes, Tversky and Kahneman (1971) showed how biased we are in moving from particulars to generalities. From one or two concrete, vivid instances, we assume there are dozens more lurking in the bushes—but we don't verify whether or how many there are, and there are usually fewer than we think. People typically make a generalization, then illustrate it ("For example, my friend X . . . "), but would be hard put to come up with several more instances of a supposedly widespread occurrence. To compound the problem, people as information seekers—and processors—are far more likely to see *confirming* instances of original beliefs or perceptions than to see disconfirming instances, even when disconfirming instances are more frequent (for example, see Edwards, 1968). Still, more ominously, Kahneman and Tversky (1972) were able with little effort to catch mathematical psychologists in the act of making biased inferences from samples to populations. What, then, of the fieldworker who, Prometheuslike, is doing all the sampling, measuring, and inference-making *de novo*?

Operating alone, without any standardized or validated instruments, the field researcher runs several risks of generalizing inappropriately from specific instances. Here are some of the most common pitfalls, and their associated sources of error:

Pitfall	*Source of Error*
sampling nonrepresentative *informants*	overreliance on accessible and elite informants
generalizing from nonrepresentative *events* or activities	researcher's noncontinuous presence at the site; overweighting dramatic events
drawing inferences from nonrepresentative *processes*	nonrepresentative informants and events; heavy reliance on plausibility; good fit into emerging explanation, holistic bias

The first pitfall highlights the fact that the researcher can talk only with people who can be contacted; some people are easier to contact than others, especially when the researcher has limited time on site. This problem in itself signals something particularistic about these people. Their accessibility may be connected with work load, lack of cooperativeness, or both.

We should also emphasize that anthropologists have often warned of fieldworkers' tendency to rely too much on articulate, insightful, attractive, and intellectually responsive informants; such people often turn out to be in the local elite. For a good discussion of the atypicality of persons chosen as prime informants, see Dean, Eichorn, and Dean (1967).

The second pitfall also results from the researcher's noncontinuous presence; you have to infer what is happening when you are not there. In particular, when the researcher observes a dramatic or salient event (a crisis, an argument), there is the tendency to assume it has "built up" when one was absent, or that it "symbolizes" a more general pattern. These are plausible, but certainly not airtight, inferences.

Finally, in looking for underlying processes explaining the phenomena you have studied closely, you naturally draw heavily from the people, events, and activities you have sampled to derive the most plausible account. But if the samples are faulty, the explanation cannot be generalized beyond them.

"Plausibility" is, unfortunately, the optiate of the intellectual. If an emerging account makes good logical sense and fits well with other, independently derived analyses within the same universe, you lock onto it and begin to make a stronger case for it; this constitutes the "confirmability" bias studied by Edwards (1968) and others. Anthropologists call this the "holistic bias": The researcher sees all the main facets of the site "coming together" to form a meaningful pattern—one more meaningful and more patterned than the loose-endedness and contradictory nature of social life warrants.

Elementary safeguards: Some illustrations. The real dilemma with selective sampling and abusive generalizing is that you can sllide incrementally into these biases, with the first layer preparing the ground for the next. Gradually, you are prisoner of your emerging system of comprehending the site. There is no longer any possibility, cognitively speaking, to "stand back" or "review critically" what you have observed up to then. What you now understand has been accumulated very gradually from within, not drawn validly from without. If you want to stand back and review critically, you are going to need someone *else* to do it—or you must build in safeguards against self-delusion. We have already addressed the former approach (second readers, auditors, second fieldworkers), so let us illustrate some safeguards.[7]

In our study of linkage between universities and local school districts, we were on the lookout for—and felt we were drifting into—the three pitfalls listed earlier. For example, each researcher covered several institutions in a state, which meant discontinuous field visiting. We wondered whether there were informants we were systematically missing when we interviewed school and university staff, even though these people assured us that we had sampled broadly. Our first safeguard was to look for the outliers (see section VII.B.6), in order to check the representativeness of those with whom we had talked. As a second safeguard, we asked for the personnel rosters of the districts and the university, and randomly sampled eight informants from each one for interviewing. As it turned out, we *had* been in danger of sampling a biased group.

Here is a second example, from the same study. Informants often talked about one type of project, saying that this was the most frequent activity sponsored by the university. We began to wonder if informants were confusing frequency with impact; did they *remember* this activity type most because it seemed to have been the most practice-relevant? So we took the last two years' newsletters, extracted a calendar of activities, and counted and content analyzed them. Informants had, in fact, misled themselves and us by overweighting one activity type.

A final example from that study corresponds to the third pitfall. There were periodic policy board meetings between delegates from the university and school districts. We tried to be present at them since they represented the formal, more decisive meeting ground between the two parties. Gradually, it appeared that university staff were dominating them; the university delegates spoke more often and longer than the school people, and most decisions went their way. We thus entertained the hypothesis of a university-dominated arrangement, an idea that fitted the data we had. But to check it, we reverted once again to a wider sampling procedure and observed *other* meetings between the two parties, notably ones that (a) were less formal and did not involve representatives or delegates; or (b) were also formal, but not as prominent as the policy board. This was salutary; we ended up having to qualify our initial hypothesis. In less formal, less prominent meetings, school people held their own in the decision making.

We suggest that you *assume* you are selectively sampling and drawing inferences from a weak or nonrepresentative sample of "cases," be they people, events, or processes. You are guilty, until you prove yourself innocent. You can do this by extending the universe of study, in at least one of four ways: (1) simply *increase the number of cases*; (2) look purposively for *contrasting cases* (negative, extreme, countervailing); (3) *sort the cases systematically* (perhaps using substructing, as in Box V.C.a) and filling out weakly sampled case types; and (4) *sample randomly* within the total universe of people and phenomena under study. The last two procedures are especially recommended. They correspond, not haphazardly, with the "stratification" and "randomization" conventions used by experimental researchers to enhance internal validity. But while the experimental researcher uses the conventions *early*, as anticipatory controls against sampling and measurement error, the qualitative researcher typically uses them *later*, as verification devices. That allows you to let all the candidate people and data in, so that the most influential ones will have a chance of emerging. But you *still* have to carry the burden of proof that the patterns you ultimately pinpoint are, in fact, representative.

VII.B.2 Checking for Researcher Effects

"Outsiders" to a group influence "insiders," and vice versa. So it is with the researcher who disembarks in a field setting to study the natives, whether in a familiar culture or a more exotic setting. The researcher is likely, especially at the outset, to create social behavior in others that would not have ordinarily occurred. This, in turn, can lead the researcher into making biased observations and inferences, thus "confounding" (a nice term in this instance) the "natural" characteristics of the setting with the artificial effects of the researcher-native relationship. Unconfounding them is like threading one's way through a hall of mirrors.

So we have two possible sources of bias here: (a) the effects of the researcher on the site and (b) the effects of the site on the researcher. Both topics have been treated at length in methodological textbooks, more so by experimental or laboratory-based researchers than by field-study methodologists. The latter are less worried about the first kind of bias (bias a) because the field researcher typically spends enough time on the site to become part of the local landscape.[8] This, of course, increases the danger of the second kind of bias (bias b): Being coopted, going native, swallowing the agreed-upon or taken-for-granted version of local events.

For simplicity, we discuss these biases in terms of what is happening during site visits. It is important to remember that they influence *analysis* deeply, both during and after data collection. The researcher who has "gone native" stays native during analysis. The researcher who has influenced the site in un-understood ways suffers unaware from that influence during analysis.

Bias (a) occurs when the researcher disrupts or threatens ongoing social and institutional relationships. People now have to figure out who this person is, why he or she is there, and what might be done with the information being collected. While they are figuring that out, informants will typically switch into an on-stage role or special persona, a presentation of self to the outsider. (They have *other* personae, of course, for fellow insiders, as Goffman, 1959, has so graphically described.)

Even after this preliminary dance, informants will typically craft their responses in such a way as to be amenable to the researcher and to protect their self-interests. For some analysts (for example, Douglas, 1976), local informants' interests are fundamentally in conflict with those of the researcher, who might penetrate to the core of the rivalries, compromises, weaknesses, or contradictions that make up much of the basic history of the site and that insiders don't want outsiders to know about—either because *other* outsiders aren't meant to find out or because the social equilibrium among local actors depends on those facts being kept private. So the researcher, who is usually interested in uncovering precisely this kind of information, must *assume* that people will try to be misleading and must shift into a more investigative mode.[9] (For more detail on useful tactics, see section VII.B.4, on *weighting the evidence.)*

It is probably true that, fundamentally, field research is an act of betrayal, no matter how well intentioned or well integrated the researcher. One makes public the private and leaves the locals to take the consequences.[10] But that is not the only way bias (a) can occur. In some instances bias (a) can team up with bias (b) to create artifactual effects, as a result of the complicity between researcher and local actors. This is the famous "experimenter" effect, studied intensively by Rosenthal (1976).

We have been caught napping several times on this one. For instance, we studied one field site in the school improvement project that was about to phase out the project we had come to see. For some mysterious reason, the phase-out decision was suspended during our time on site, The reasoning, which we unraveled after several more days on site, was that the practice *had* to be better than it appeared since people had come from so far away to see it. Mixed in also was the desire to avoid a public indictment; the researcher, and the public reading the research or talking with her, might convey the impression that the school had botched the project.

Bias (a) can take still other forms. For example, local site informants can implicitly or explicitly boycott the researcher, who is seen variously as a spy, a voyeur, or a pest. Or the researcher can inhibit local actors. After several days on site and multiple interviews with informants, people aren't sure any more how much the researcher has found out and assume—wrongly in most cases—that the researcher knows *too* much. This then triggers bias (b). The researcher accordingly becomes more reassuring or, alternatively, moves into the investigative-adversarial mode. Both strategies are likely to affect the data being collected.

Assuming then, that the researcher has only a few months, weeks, or even days on site, how may these two interlocking forms of bias be countered? Below is a short shopping list of suggestions, many of which are treated in far more detail in the mainstream methodological literature (for example, see Pelto & Pelto, 1978; Lofland, 1971; Adams & Preiss, 1960; Wax, 1971).

(a) Avoiding biases stemming from researcher effects on the site:

- Stay as long on site as possible; spend some of that time simply hanging around, fitting into the landscape, taking a lower profile.
- Use unobtrusive measures where you can. (Webb, Campbell, Schwartz, & Sechrest, 1965; McCall & Simmons, 1969).
- Make sure your mandate is unequivocal for informants: why you are there, what you are studying generally, how you will be collecting information, what you will do with it.
- Consider coopting an informant—asking that person to be attentive to your influence on the site and its inhabitants.
- Do some of your interviewing off site, in a congenial social environment (cafe, restaurant, informant's home), by way of reducing both your threat quotient and your exoticism quotient for informants.
- Don't inflate the potential problem; you are not *really* such an important presence in the lives of these people.

(b) Avoiding biases stemming from the effects of the site on the researcher:

- Avoid the "elite" bias by spreading out your informants; include people not directly involved in the focus of your study (peripheral actors, former actors).
- Avoid cooptation or going native by spending time *away* from the site; spread out site visits (see Whyte, 1943, on "temporary withdrawal").
- Be sure to include dissidents, cranks, isolates—people with different points of view from the mainstream, people less committed to tranquility and equilibrium in the setting.
- Keep thinking *conceptually*; translate sentimental or interpersonal thoughts into more theoretical ones.
- Consider coopting an informant who agrees to provide background and historical information for you and to collect information when you are off-site (the cooptation may be more useful, in bias-reduction terms, than the information provided).
- Triangulate with several data collection methods; don't overly depend on *talk* to make sense of the setting.
- If you sense you are being misled, try to understand, and focus on *why* an informant would find it necessary to mislead you. Follow that trace as far upstream as you can.
- Don't casually show off how much you *do* know; this is a covert plea for confirmation that deludes only the person making it.
- Show your field notes to a second outside reader. Another researcher is often much quicker to see where and how a fieldworker is being misled or coopted.
- Keep your research questions firmly in mind; don't wander too far from them to follow alluring leads, or drop them in the face of a more dramatic or momentous event.

As with all such lists, following some items gets you in trouble on others. For instance, if a researcher has only eight days on site, spending much of it in off-site interviewing may be too costly. Or, one may be coopted by the informant one is trying to coopt.

Supposedly, bias detection and removal take time. The more time you have, the more layers you can peel off the setting to get to the meatiest, explanatory factors, and the less subject you are to either of the biases described above.

However, we take that with a grain of salt. Long exposure can just push up bias (b) and make bias (a) harder to see. We reiterate that people who are discreet, savvy in the environment under study, and conceptually ecumenical are often able to get to the core of a site in a matter of days, sidestepping both types of researcher bias and coming away with higher-quality data than others could have compiled after several months' work, if at all. In that sense, it's possible that the methodologists demanding months or years on site before valid data can be obtained are confusing time with competence.

VII.B.3 Triangulating

In psychological testing, an important part of the internal validation process is checking a new item or test against other, already validated, measures of the same skill or construct. If they concur—overlap, correlate strongly—the new item or test has good "concurrent validity." Until recently, qualitative research had no comparable all-purpose term. One spoke of "corroboration" or, somewhat loosely, of "cross-validation" or of "multiple validation procedures," (see Becker, 1958) to ensure the dependability of a field-study finding. In all instances, since there was typically no *external* measure to check the new finding against, one looked to other internal indices that should provide convergent evidence. Webb et al. (1965) coined a term for this procedure that stuck: triangulation. In addition, their depiction of the process was an apt one; they spoke of validating a finding by subjecting it to "the onslaught of a series of imperfect measures."

Stripped to its basics, triangulation is supposed to support a finding by showing that independent measures of it agree with it or, at least, don't contradict it.[11] The measures are imperfect in that the researcher usually invented them on the spot, and we know little about their validity or reliability. They are also imperfect because they usually come from the same "instrument," that is, observations made or conversations recorded by the researcher alone. When the same instrument—in this case the same person—is both establishing and corroborating a finding, we have what amounts to a potential cognitive conflict of interest.

Bias, however, is not inevitable. Detectives, car mechanics, and general practitioners all engage successfully in establishing and corroborating findings with little elaborate instrumentation. They often use a *modus operandi* approach, which consists largely of triangulating independent indices. When the detective amasses fingerprints, hair samples, alibis, eyewitness accounts, and the like, a case is being made that presumably fits one suspect far better than others. Diagnosing engine failure or chest pain follows a similar pattern. All the signs presumably point to the same conclusion. Note the importance of having different *kinds* of measurements, which provide repeated verification.

How can the qualitative researcher apply a *modus operandi* approach to the testing of field study findings? Essentially, we have been saying throughout this book that this is precisely how you get to the finding in the first place—by seeing or hearing multiple instances of it from different sources, and by squaring the finding with others it needs to be squared with. Analytic induction, once again.

But it is important to make certain that the several indices chosen are indeed independent, sturdy, of different types and sources, and congruent. Let's look at an example. At one field site in the school improvement study, we found what looked like a highly successful practice. Virtually all the people we talked with made this claim, and the test scores looked good. To make certain of the finding, we assembled the most likely sources of evidence:

(1) test scores for first-year and second-year pupils whose teachers were and were not using the practice
(2) testimony of teachers using the practice
(3) testimony of teachers *not* using the practice
(4) testimony of pupils
(5) observations of the practice
(6) samples of pupils' work
(7) hands-on work with the practice in the classroom
(8) testimony of local administrators
(9) observation of classrooms *not* using the practice
(10) analysis of the program manual and materials

Notice that we have compiled different *sources* of evidence, using different *methods* and operating at different *levels* of the school. Most indices were *corroborative* or verificatory indicators of success, while some were more *contrasting* and *inferential* indicators. (We looked at comparable test scores for nonusing pupils, observed nonusing classrooms, and talked with nonusing teachers to see whether different or lesser results were obtained when other methods were used.)

Finally, putting some of the indicators together yielded some *multiple* and more *causally linked* evidence. For example, we analyzed the program manual and materials (10 on the list) to determine whether, in fact, the results could flow conceptually and logically from the features of the program itself. We then conducted observations in the classrooms to make sure that teachers were putting into practice the same program described in the manual. Looking at work samples provided another check on this, as did a check on pupils scoring high on the tests in relation to low-scoring pupils: Were they further ahead in the program? Triangulation here consisted of retracing the most plausible causal chain from program design to execution to interim outcomes (work samples) to ultimate outcomes (test scores), trying to get more than one type of measure from more than one source for each link in the chain.

The process sounds more obsessive and sophisticated than it really is. It is easy to sit down and imagine where one can find or double-check sources of corroborative, contrasting, and causally linked information. It is also relatively easy to get data from multiple *sources* (people with different roles, deviant and mainstream informants) using multiple *methods* (such as talking with people *and* observing routine life at the site). See also the thoughtful discussion by Jick (1979).

It also helps to be on the lookout for a *new* source of data—a new informant or class of informants, another comparable event or setting. A new source forces the researcher to "replicate" the finding in a place where, if valid, it should reoccur (see also section VII.B.9).

Finally, we can triangulate with different *researchers*. They can be taking parallel mesures at the same time, or following up a finding to confirm it.

Perhaps our basic point is that triangulation is a state of mind. If you *self-consciously* set out to collect and double-check findings, using multiple sources and modes of evidence, the verification process will largely be built into the data-gathering process, and little more need be done than to report on one's procedures.

VII.B.4 Weighting the Evidence

Any given preliminary conclusion is always based on certain data. Maybe we should use the word some historians have employed: "capta." There are events in the real world, from which we "capture" only a partial record, in the form of raw field notes, from which we further extract only certain information in the form of write-ups, which we then call "data." There is in turn further reduction, selection, and transformation as these data are entered into various displays.

Some of these data are "better" than others. Fortunately, the qualitative analyst can exploit that fact beautifully in verifying conclusions. If the data on which a conclusion is based are known to be stronger, more valid than the average, then the conclusion is strengthened. Stronger data can be given more weight in the conclusion. Conversely, a conclusion based on weak or suspect data can be, at the least, held lightly, and, optimally, discarded if there is an alternate conclusion with stronger data back of it.

Basically, there is a very large range of reasons that certain data are stronger or weaker than others—essentially, the question is one of validity (Dawson, 1979, 1982). We cannot be encyclopedic here, but will suggest a number of markers the analyst can use in deciding whether to give more weight to some data than to others.

First, data from some *informants* are "better." The informant may be articulate, thoughtful, and reflective, and may enjoy talking about events and processes. Or the informant may be knowledgeable, close to the event, action, process, or setting with which you are concerned. In our study, for example, we gave more weight to school superintendents' judgments about forward budget categories than we did to those of teachers.

Second, the *circumstances* of the data collection may have strengthened (or weakened) the quality of the data. Here is a partial list (see also Sieber, 1976; Becker, 1970; Bogdan & Taylor, 1975):

Stronger Data	*Weaker Data*
Collected later, or after repeated contact.	Collected early, during entry.
Seen or reported firsthand.	Heard secondhand.
Observed behavior, activities.	Reports or statements.
Fieldworker is trusted.	Fieldworker is not trusted.
Collected in official or formal setting.	Collected in informal setting.
Volunteered to fieldworker.	Prompted by fieldworker question.
Respondent is alone with fieldworker.	Respondent is in presence of others, or in group setting.

Finally, data quality may be stronger because of a fieldworker's *validation* efforts. These may be of several varieties:

- checking for researcher effects (section VII.B.2) and biases
- checking for representativeness (section VII.B.1)
- getting feedback from informants (section VII.B.12)
- triangulating (section VII.B.3)
- looking for deception
- looking for ulterior motives

We might comment briefly on the last two, since they haven't been attended to in other sections. Douglas (1976) emphasizes the idea that, regardless of the degree of trust a fieldworker may feel has been developed, people in field sites nearly always have some reasons for omitting, selecting, or distorting data, and may have active reasons for *deceiving* the fieldworker (not to mention deceiving themselves). If the fieldworker has actively entertained such a view of particular respondents, and of a particular set of data from them, *and* has done something to validate the data, more confidence is justified. The interested reader should consult Douglas (1976) for a wide range of specific interventions. Here are a few:

- Check against "hard facts."
- Check against alternative accounts.
- Look for "the trapdoor"—what's going on beyond the obvious.
- Share own personal material to open up the respondent.
- Assert your knowledge of "what's going on" and see if respondent buys it.
- Summarize a state of affairs and ask the respondent to deny it.
- Name possible ulterior motives, and see respondent response (denial, acknowledgment).

Fieldworkers who rely mainly on "trust" may quail at such interventions or dismiss them as too intrusive. Nevertheless, Douglas makes a good case for such validating tactics when the object of investigation has good reasons for being evasive and/or self-deceiving (some of his studies have included people who had attempted suicide, clients of massage parlors, habitues of nude beaches, and police who work in emergency rooms). And even for less dramatic settings, we have found that it pays to be suspicious, to expect to be lied to sometimes, to look for respondent self-deception, and to push respondents from time to time on such matters.[12]

Two added suggestions: First, we have found it useful to keep a running log of data quality issues (often in the form of reflective or marginal remarks on the field notes; see Boxes III.B.a and III.B.b) together with recurrent efforts to improve data quality in subsequent site visits.

Second, when approaching final write-up of a site analysis, it is useful to summarize one's views of data quality. Here is an example from a site report in our school improvement study, which appeared after the researcher summarized the number of interviews (46), informal talks (24), and observations (17) held during three site visits:

> The data base is probably biased toward administrators and central program personnel (3-6 interviews apiece), and may underrepresent those of normal program users, and certainly those of peripheral (and more disenchanted) people. So the information may be fuller about the ins and outs of operation as seen by key operators, and thinner on what day-to-day life in the Carson schools is like, with the IPA program as a feature in that life.
>
> Though some interviews were brief, I had no difficulty in re-interviewing people whose opinions seemed crucial or especially illuminating.
>
> I have moderately good retrospective data on the first and second years of the program, except for assistance provided. Data on the current year's operations are quite full, except that I observed no actual student-parent-teacher conferences in the high school. The only key informant I missed talking to was the former IPA director, Helena Rolland. It could also be argued that I should have interviewed Mark Covington, as the archetypal coach. But with these exceptions the coverage was thorough.
>
> Dictation from field notes was done 14-32 days after Visit #1, 28-49 days after Visit #2, and 10-30 days after Visit #3. Interim phone calls were usually written up immediately after the call, with one or two exceptions. In spite of the write-up delay, I experienced little decay from notes. Where puzzling or meaningless notes could not be reconstructed, this was marked in the field notes and noted during dictation; such indications appeared, however, for only 3-5 percent of field notes.

Editorial comment: The researcher's confidence about "little decay" is almost surely self-deluding. The probable loss of detail will need to be compensated for by triangulation and by looking for repeat examples from the same respondent. And comments from "disenchanted" people should probably be given more weight in conclusion verification.

VII.B.5 Making Contrasts/Comparisons

A time-honored, classic way to test a conclusion is to draw a contrast or make a comparison between two sets of things—persons, roles, activities, sites as a whole—that are known to differ in some *other* important respect. This is the "method of differences," which goes back to Aristotle, if not further. (The contrast between experimental and control groups was not invented by R. A. Fisher.) A few examples from our work:

- When we looked at preparedness (section IV.B, Chart 13a), the comparison showed that administrators were enthusiastic about the innovation, but users were bewildered—a picture that fit with other aspects of their roles: Administrators press for adoption, users have to do the actual implementation work.
- Comparing *sites* that had many negative effects of implementation with those that had few (section V.B, Chart 31a) showed us that such sites were also ones where demanding innovations were being tried.
- Comparing *program sponsorship* (NDN versus IV-C, section V.B, Chart 31a) showed us that there were *no* differences in final student impact, a finding that fit with prior expectation.
- Contrast tables (Box V.C.b) comparing *sites* on, for example, amounts of user change, made it clear that "change pervasiveness" might be causing user change.
- Predictor-outcome matrices (section V.C) array sites by high and low *outcomes*, and use that leverage to examine the impact of a wide range of possible predictors. In our example, the comparison was between roughness and smoothness of implementation; which predictors were present in smooth sites but not in rough ones (Charts 32, 33)?
- Comparing job mobility during the middle, early, and later portions of projects showed (section V.D, Chart 36) the effects of *timing* and project development; more people moved *in* at the beginning and *out* toward the end.

The method of comparisons is so pervasive throughout social science that we won't introduce other examples. Some notes, however:

(1) Mindless comparisons are useless. The trick is to be sure that the comparisons being made are the right ones, and that they make sense.

(2) The results of a comparison should themselves be compared with what *else* we know about the roles, persons, groups, activities, or sites being compared.

(3) Take a minute before you display a comparison, and think, "How big must a difference be before it makes a difference? And, how do I think I know that?" You don't have a significance test to fall back on.

VII.B.6 Checking the Meaning of Outliers

For any given finding, there are usually exceptions. The temptation is to smooth them over, ignore them, or explain them away. But *the outlier is your friend.* A good look at the exceptions, or the ends of a distribution, can test and strengthen the basic finding. It not only tests the generality of the finding, but protects against self-selecting biases.

For example, in the school improvement study, we happened on one site where the new practice was seen by many teachers as a *miraculous cure* for local ills. Although teachers found it hard to get on top of, the project eventually led to dramatic increases in reading and composition scores. Enthusiasm was high.

To test the generality of the finding, we asked about people who either hadn't adopted the practice or had used it and found it wanting. After some thought, our key informants came up with one each.

Our interviews with these two people were instructive. First, we found that the reasons given for *not* adopting were opposite to—and thereby coherent with—the reasons given by the other informants *for* adopting. And we found that the dissident user had *not* really mastered the innovation in the way the contended users had. We already had good evidence linking technical mastery to positive results. So our findings were strengthened, and we understood far better *why* deviant cases were deviant.

So, was the innovation a "miracle cure"? Perhaps—but only if it were technically well carried out. Furthermore, these dissidents told us there were more people like them around than advocates had admitted.

We realized then that we had oversampled contented users and, in a sense, had been "sucked in" to the taken-for-granted version of events among local actors. In widening the sampling of discontented users thereafter, we got a somewhat different, and more intricate, picture of the site.

The second part of the illustration suggests that, very often, there are *more* exceptions or deviant cases than one realizes at first, and that one has to go looking for them. They don't come calling, nor do we think spontaneously of sampling for them. After all, they are inconvenient—not only hard to reach or observe, but, more fundamentally, spoilers of the artfully built, coherent version of site dynamics at which the researcher has arrived.

Remember, too, that outliers are not only people; they can also be discrepant *sites*, atypical *settings*, unique *treatments*, or unusual *events*. You need to find the outliers, then verify whether what is present in them is absent or different in other, more mainstream examples (see also the discussion of *using extreme cases*, section VII.B.7).

A good illustration appears in Section V.E, Figure 15, where the Astoria site proved to be an outlier. Most other sites had *high* pressure to adopt the innovation, and gave *little* latitude to users, or vice versa—little pressure and lots of latitude. Only in Astoria was there high pressure *and* high latitude. Why should this be? A look back at the site report showed that some careful administrator-teacher bargaining had been going on: The administrator said the innovation was mandated, but agreed that it could be flexibly adapted. Astoria was also the only parochial school in our sample, suggesting that authority was less questioned. This exploration gave us more confidence that the basic finding was right.

Finding outliers is easier when you have good displays. Sites, settings, events, people can be shown along a distribution. If you are still collecting data, display what you have and go for the outliers. If things are closely clumped (no apparent outliers), consider where you might go to find some outlying persons, events, settings. And, on following up, be cognitively open to the eventuality that the exceptional cases are, it turns out, the modal ones.

VII.B.7 Using Extreme Cases

We have just described the use of *outliers* (section VII.B.6) in deepening preliminary conclusions. Outliers of a certain type, which we'll call extreme cases, can be very useful in verifying and confirming conclusions.

Here we can use two illustrations from Sieber (1976). The first involves extreme *situations*. Sieber asks us to imagine a situation in which an educational innovation failed, then to look at the possible antecedents of that failure. If we found that in a particular site there were many *positive* factors—such as high motivation to innovate, access to resources, and implementation skills—but that *administrative support* was lacking, we could argue persuasively that in this site we have found the key factor responsible for the failure. In effect, Sieber suggests, this is a tactic of "holding everything else constant"—looking for the most extreme case, where there *should* have been success, but there wasn't. Note that this tactic requires conceptual and/or empirical knowledge of the variables involved; it cannot be done in a vacuum.

The second sort of extreme case Sieber mentions is *persons* known to have a strong bias. For example, suppose that you are talking to a very conservative administrator, whom you know from past contact is inclined to be rather defensive. You ask him why the teachers he works with are reluctant to try innovations. He answers that it's due to a lack of cooperation and support on his part. That answer is very persuasive because you wouldn't expect this particular administrator to make such a statement at all. So its truth value is probably high.

To put this another way: Look for the person in a site who would have most to gain (or lose) by affirming or denying something, and pop the question. If you get a surprising answer (for example, the person who has much to gain by denying the statement/question in fact affirms it), then you can be more confident. Of course this requires that you have a good prior understanding of the person's typical stance and biases, not a superficial impression or stereotype.

In a way, this is another style of differentially *weighting evidence* (section VII.B.4). For example, if you are interviewing people you know to be enthusiastic proponents of an innovation, their comments on the innovation's warts and trouble spots should be taken quite seriously.

VII.B.8 Ruling Out Spurious Relations

Suppose that you have been able through assorted ingenious tactics to establish that variable A is indeed related to B, perhaps causally. Before breathing a sigh of relief and proceeding to the next conclusion, it pays to consider that the picture you are drawing:

$$A \rightarrow B$$

may in fact be more accurately portrayed as

$$\begin{array}{ccc} & X & \\ \swarrow & & \searrow \\ A & & B \end{array}$$

where some third factor is in play, causing both A and B to occur.

This is an old problem, which statisticians have dealt with well. We can draw a nice example from Wallis and Roberts (1956), describing a study from the *Journal of the American Medical Association*. Researchers noted that polio patients who traveled longer distances (average, 85 miles) to a hospital were more likely to die than patients who traveled little (average, 7 miles) and were more likely to die sooner (50 percent died within 24 hours, versus 20 percent). They concluded:

> The greater mortality in the transported group, occurring shortly after admission to the hospital, is a manifestation of the effect of long transportation during the acute stage of illness.

Wallis and Roberts suggest that another, third variable may be influencing both A (transportation) and B (mortality). It is *seriousness of the initial attack*. All the patients were seen in a certain hospital, Willard Parker, a noted center for treatment of contagious diseases. Polio patients who lived farther away were probably only brought to Willard Parker if their conditions were

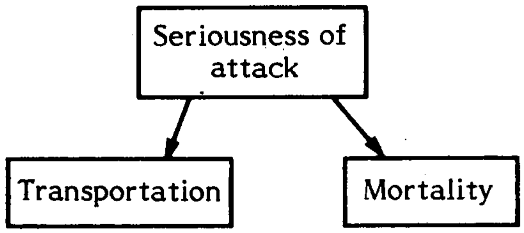


Figure 26a Possible Explanation of Spurious Relationship

serious; milder cases would be treated nearer their own homes. Thus the picture develops as shown in Figure 26a. This interpretation can be checked out through the sort of display found in Figure 26b, with Ns and mortality rates entered in the cells.

And, as Wallis and Roberts faithfully point out, even if the reanalysis could be done, and it supported "seriousness of attack" as the real issue, one would have to do additional ruling out. Perhaps those coming from a distance had poorer basic health to begin with. Perhaps they came from an area where a particularly virulent strain of polio was widespread. And so on.

Finding a candidate third variable is not always easy, particularly if the original explanation "makes sense," as the transportation-mortality link seemed to at first glance. The Willard Parker researchers did think of one third variable—length of prior illness—which showed no real difference. Then they stopped, not realizing that the "Willard Parkerness" of Willard Parker was probably in the picture. Had they been able to recruit Wallis and Roberts as "friendly strangers" part way through their analysis, the story might have been different.

The moral for qualitative researchers is the same. When two variables look correlated, especially when you think they are causally associated, wait a beat, and consider whether some third variable might be underlying/influencing/causing them both. Use a knowledgeable but detached colleague to help in the search. Then consider new displays that will give you a clean look at such third variables and their effects.

Doing this in a more than cursory way takes time, so it is worth it mainly when you have a major (but perhaps surprising) conclusion, or one on which a lot is riding, in practical terms.

VII.B.9 Replicating a Finding

As we showed in section VII.B.3, findings are more dependable when they can be buttressed from several independent sources. Their validity is enhanced for having been confirmed by more than one "instrument" measuring the same trait or outcome.

Still, the fact that in most cases one person is doing all this measuring with homemade instruments is grounds for precaution. Once a researcher has latched onto a hypothesis that makes powerful sense of the site, it's the dickens to get rid of it. Confirmation seems, almost magically, to come from all quarters. New interviews, observations, and documents all appear to bring verification, and to fit together coherently. Disconfirming evidence is absent or feeble. This is a heady and extremely dangerous moment, and it usually means that the researcher is knee-deep in the "holistic fallacy": putting more logic, coherence, and meaning into events than the inherent sloppiness of social life warrants. How to protect against this?

One line of attack is to think in terms of *replication*, which is the bedrock of science. If I can reproduce the finding in a new context or in another part of my data base, it's a dependable one. If someone else can reproduce it, better still.

There are several ways of doing this. At the most elementary level, the fieldworker is replicating in the simple act of collecting new information from new informants, from new settings and events. New data bolster or qualify old data; in fact, one begins to collect new data *in order to* test the validity and generality of the old.

At a notch higher in the confidence scale, one can test an emerging hypothesis in another part of the site or data set. This amounts to "if/then" reasoning: If I find this *here*, then I should find it—or something coherent with it—*there*, too. For instance, in one study we gradually got the picture that school administrators in the district office were aggressively promoting changes in three local schools, bypassing the school principals. This looked important, because it could realign the entire power-influence network within the district. Teachers, for example, could short-circuit

	Local patients	Distant patients
Mild attack		
Severe attack		

Figure 26b Display for Testing Explanation in Figure 26a

their principals and go directly to the central administrative office. We reasoned that, if this were true, the principals would have relatively little detailed understanding of the practices being so implemented, even though they were happening in that principal's own building. So we conducted open-ended interviews with the principals, simply asking them to tell us about the practice. As hypothesized, there was a good deal of hemming and hawing, even in the very small schools that principals could cover easily.

Such a test is more rigorous; it's harder to bootleg researcher bias into it. Even stiffer tests can be made. For example, in another study, we came up with the hypothesis that a "localist" orientation at a college or university was more promising than a more "cosmopolitan" orientation when the college undertook collaborative work with surrounding school districts. To test this, we had "saved" one college in the sample for study later on. We operationalized what we meant by localism and cosmopolitanism (see section VII.A.8), described for the other sites just how the two variables were associated with successful collaborative projects, then went to the new (most localist) site and tried to determine whether our hypothesis was viable. (It was, but needed some qualifying.) Others (see Stake and Easley, 1978) have used this kind of staggered replication device even more methodically.

Some brief words of advice:

(1) In cross-site studies, replication is an important part of the basic data-collection effort. Emerging patterns from one site need to be tested in others. They usually surface in some of the methods we described earlier: pattern codes, memos, site analysis meetings, interim site summaries. One should therefore be prepared, in the course of fieldwork, to do the kind of corroborative and replicative testing described here and in section VII.B.3.

(2) If provisions aren't made in advance for replications later in the study, they won't happen; there is inevitably too little time and too much information still to compile.

(3) Doing replication at the end of fieldwork, during final analysis and write-ups, is very difficult and less credible. To test a hypothesis in another part of the data set assumes that all the requisite data are there for the test to be made. They usually aren't, unless the researcher has made sure to collect them beforehand in anticipation of just such an exercise.

VII.B.10 Checking Out Rival Explanations

Thinking that there may be rival explanations to account for the phenomenon one has carefully studied and masterfully unraveled is a healthy exercise in self-discipline and hubris avoidance. But that thought often gets lost in the shuffle. During data collection, you are often too busy making minimal sense of the welter of stimuli. Later on, you tend to get married to your emerging account, and usually opt for investing the scant time left to buttress, rather than to unhorse, your explanation. Then, during data analysis, it is often too late to "test" any other explanation than the one arrived at; the data necessary for doing that in any but a cosmetic way just aren't there.

So, in qualitative research, there appears at first blush to be little of the kind of rival hypothesis-testing that Platt (1964) praises so much when he speaks of "strong inference." Platt, drawing on fields such as molecular biology, emphasizes (a) developing alternative hypotheses; (b) specifying critical experiments, the outcomes of which will exclude one or more hypotheses; (c) carrying out the experiments; and (d) recycling until a satisfactory conclusion is reached. But in most social settings, we cannot easily conduct the series of carefully controlled, elegantly scaffolded critical experiments that theoretically—but seldom practically—do away with equivocal findings from competing studies. Maybe we shouldn't worry too much about trying rival hypotheses on our data, since our rivals will cheerfully do it for us afterwards. (Somehow, though, *their* conclusions don't match the ones *we* would have reached had we done the same exercise.)

We claim that the search for rival explanations is usually more thorough in qualitative research than in survey research or in most laboratory studies, and that it is relatively easy to do. The competent field researcher is looking for the most plausible, empirically grounded explanation of local events from among the *several* competing for attention in the course of fieldwork. You are not looking for *one* account, forsaking all others, but for the best of several alternative accounts. This is rendered beautifully, we think in Umberto Eco's novel, *The Name of the Rose* (1982, pp. 311-312), in the following exchange between the protagonist and his foil:

> *Guillaume:* I have a number of good hypotheses [for explaining a series of events], but no overriding fact to tell me which is the best . . .
>
> *Adso:* So then you must be far from finding the solution?
>
> *Guillaume:* I'm very close, but I don't know to which one.
>
> *Adso:* So you don't only have one solution?
>
> *Guillaume:* Adso, if that were the case, I'd be teaching theology at the Sorbonne.
>
> *Adso:* So in Paris they always have the right solution?
>
> *Guillaume:* Never, but they're very confident of their errors.
>
> *Adso:* What about you? Don't you make errors too?
>
> *Guillaume:* Often. But instead of imagining one, I try to imagine several. That way I don't become a slave of any one error in particular. (translated from the French)

The trick, then, is that of holding on to several possible (rival) explanations until one of them gets increasingly more compelling as the result of more, stronger, and varied sources of evidence. Looking at it from the other end, you give each rival explanation a good chance. Is it maybe *better* than your main love? Do you have some biases you weren't aware of? Do you need to collect any new data?

Foreclosing too *early* on alternative explanations is a harbinger of bias, of what psychometricians call "systematic measurement error." One locks into a particular way of construing the site and selectively scans the environment for supporting evidence. Discounting evidence is ignored, underregistered or handled as "exceptional"—and, as such, *further* increases one's confidence in one's thesis.[13]

On the other hand, we should note that closing *too late* on alternative explanations builds too weak a case for the best one. It also adds enormous bulk to the corpus of data. So, rival explanations should be looked at fairly promptly in fieldwork, and sustained until they prove genuinely inviable—or prove to be better. This should happen if possible before the bulk of fieldwork is done. The same principle applies to analysis done *after* fieldwork. Check out alternative explanations early, but don't iterate forever.

It is usually difficult for the analyst who has spent weeks or months coming up with one explanation to get seriously involved with another one. The idea may have to come from someone else who is not on the site and has the cognitive distance to imagine alternatives, to be a devil's advocate. You can ask a colleague (preferably from another discipline), "Here's how I see it and here's *why* I see it this way. Can you think of another way to look at it?" We also heartly commend the use of what Platt (1964) calls "The Question": "What could disprove your hypothesis?" Or, "What hypothesis does your analysis *dis*prove?"

It also helps to fasten on to discrepant information—things that "don't fit" or are still puzzling. The trick is *not* to explain them away in light of one's favorite explanation—that's a piece of cake—but rather to run *with* them, to ask yourself what kind of alternative case *these* bits of information could build, then check them out further.

See section V.E, Figure 15, for an illustration of checking rival explanations. Explanation 1 was that high pressure to adopt an innovation will be accompanied by plenty of latitude to change it, which administrators grant in a bargain with users. Explanation 2a turned up when the analyst cross-tabulated the data: Pressure and latitude are *inversely* related. Explanation 2b noted that *some* bargaining was going on, but much later during the relationship; the inverse finding was essentially upheld. Note that in this case the rival hypothesis wasn't dredged up from the depths of the analyst's mind or presented by a "stranger" colleague, but was forced on him by the data.

A useful rule of thumb is this: During final analysis, first check out the merits of the "next best" explanation you or others can think of as an alternative to the one you preferred at the end of fieldwork. "Next bests" have more pulling power than fanciful alternatives. For more on rival explanations, including many fascinating problem exercises for practice, see Huck and Sandler's (1979) text.

VII.B.11 Looking for Negative Evidence

This tactic is easy to describe, but given people's pattern-making proclivities, not naturally come by. Essentially, when a preliminary conclusion is in hand, the tactic is to say, "Are there any data that would oppose this conclusion, or are inconsistent with this conclusion?" This is a more extreme version of looking for *outliers* (section VII.B.6) and for *rival explanations* (section VII.B.10); you are actively seeking disconfirmation of what you think is true.

Einstein is supposed to have said, "No amount of evidence can prove me right, and *any* amount of evidence can prove me wrong." That is right, in the abstract, but most of us act as if the converse were true. Our beautiful theories need little data to convince us of their solidity, and we are not eager to encounter the many brutal facts that could doom our frameworks.

As Glaser and Strauss (1967) have remarked, "There are no guidelines specifying how and how long to search for negative cases," so it's never quite clear what "enough" effort is. For a good case study of such an effort, see Kidder's (1981) account of Cressey's (1953) classic study of embezzlers.

The easiest way to proceed is to commission a skeptic to take a good cut at the conclusion at hand, avoiding *your* data display, and seeking data back in the written-up field notes that would effectively disconfirm your conclusion. If such evidence is found, do your best to rejoice, and proceed to the formulation of an alternative conclusion that deals with the evidence. If such evidence cannot be marshaled in, say, half the time it took you to do the analysis, more confidence is justified.

But note what might be called the "delusional error." Absence of negative evidence can *never* be decisive as a confirmatory tactic. As in this exchange:

> *Therapist:* Why do you have that blue ribbon on your little finger every day?
>
> *Patient:* It's to keep elephants from following me.
>
> *Therapist:* But there are no elephants here.
>
> *Patient:* See? It's working.

VII.B.12 Getting Feedback from Informants

One of the most logical sources of corroboration is the people with whom one has talked and whom one has observed. After all, an alert and observant actor in the setting is bound to know more than the researcher ever will about the realities under investigation (see Blumer, 1969). In that sense, local informants can act as a panel of judges, evaluating singly and collectively the major findings of a study (Denzin, 1978).

Feeding findings back to informants is a venerated but not always executed practice in qualitative research. It dates back at least to Malinowski's fieldwork in the 1920s, and has been used in numerous field studies since then. More recently, Bronfenbrenner (1976) classified feedback to informants as a source of "phenomenological validity," and Guba (1981) built it into his repertoire of devices for assuring the "confirmability" of findings, using the sociologists' term "member checks." Other researchers (such as Stake, 1976) have made the case for feedback a quasi-ethical one—informants have the right to know what the researcher has found—and still others, more and more numerous, are feeding back findings because people at the site are making this a precondition for access.

Earlier (in sections IV.J and IV.K) we showed two techniques for feedback to site informants. The first one required the reader to comment on a short summary of findings, then to evaluate the accuracy of a causal network encapsulating the higher-inference findings. In the second, the researcher generates predictions that should play out if the findings are valid, then submits them to site informants for verification a year later.

There is an important distinction to be made about the *timing* of such feedback. In a more or less deliberate way, researchers are *continuously* getting feedback in the course of data collection. When a finding begins to take shape, the researcher checks it out with *new* informants and/or with *key* informants, often called "confidants." The check-out process may be more indirect with the former than with the latter, who are often built into the study as sources of verification (for example, Becker, Geer, Hughes, & Strauss, 1961; Lofland, 1971; Whyte, 1943). The delicate issue here, of course, is that of introducing bias (see section VII.B.2). Feeding things back in the course of a study may change informants' behaviors or perspectives.

There are good reasons for conducting feedback after final analysis instead of during data collection. For one thing, the researcher knows more. You also know better what you know—are less tentative, have more supporting evidence, can illustrate it. In addition, you can get feedback at a higher level of inference: on main factors, on causal relationships, on interpretive conclusions. Finally, the feedback process can be done at this point in a less haphazard way. You can lay out the findings clearly and systematically, and present them to the reader for careful scrutiny and comment. As we showed in sections IV.J and IV.K, it is crucial that the reader be able to connect to the feedback—understand it, relate it to local experience and perceptions, *do* something with it (draw on it, cross out parts and add others, and so on). So *formatting* the feedback is crucial. Sending back an abstract, an executive summary, or the concluding chapter, without transforming it into the language of the site—which the researcher has come to learn in the course of field research—is of little value if one is seriously after verification.

Some advice:

(1) If you don't plan deliberately for this exercise—setting aside the time, doing the transforming of

write-ups into site-comprehensible language and formats, leaving time and space to incorporate the results of the exercise into one's final write-up—it probably won't happen. Once again, there will be too many competing claims on your time.

(2) Think carefully about displays. As with analysis, matrices and figures work much better than text alone to help local site informants access the information. Better still, people will find it easier to get an overview, to see how the pieces fit together.

(3) Providing information at more macroanalytical levels of inference (such as main factors and relationships, plus causal determinants) has to be done very carefully, by working up from particulars. If this isn't done, informants may discount the whole exercise because the overarching findings look wrongheaded or incomprehensible. Or they may swallow these macro findings whole, because they read so "scientifically." As the symbolic interactionists have shown convincingly (see Blumer, 1962), people don't act toward social structures or institutions or roles; they act toward *situations*, and are likely to understand meta-situational language only to the extent that it is directly connected to these situations.

Still, as we have remarked, people in sites *do* make maps of their reality.[14] Those maps may not coincide with the researcher's, but if we can cast our pictures, and feed them back, in an informant-friendly way, they can at the minimum be acknowledged. Beyond this, there is the possibility of mutual enlightenment.

(4) Counterpoint: Think very carefully before feeding back *any* specific incident. Will anyone's self-esteem, job chances, or standing in the organization be damaged by the report? (One of us once fed back first-draft site reports to people in five sites and was threatened with lawsuits in four of the five, because of specific incidents in the reports—even though the incidents were accurately reported.)

(5) Don't expect that informants will always agree with you or with one another. If they always did, life at the site would be more conflict-free than you probably found it to be. People have wildly varying perceptions of the same phenomenon—popularized in the so-called Rashomon effect. One should be aware that there are several reasons informants might reject the conclusions or interpretations a field researcher submits to them. Guba and Lincoln (1981, pp. 110-111) have reviewed them succinctly and well. To recapitulate:

- The informant is not familiar with the information.
- The informant doesn't understand it (jargon, lack of sophistication or awareness).
- The informant thinks the report is biased.
- The information conflicts with the informant's basic values, beliefs or self-image.
- The information is in conflict with the informant's self-interest.
- This is not the way the informant construes or puts together the same information.

General implication: The occasion of data feedback is an occasion to learn more about the *site*, not only about your feedback.

VII.C DOCUMENTATION AND AUDITING

The Problem

Most journals require authors of empirical studies to report on their procedures as an integral part of the article. The formats are often so familiar that the author can almost fill in the blanks when writing sections on sample, methods, and data analysis.

These conventions serve two purposes. First, they are a *verification* device the reader can use to track the procedures used to arrive at the findings. Second, the reporting procedures furnish details that *secondary analysts* can use to double-check the findings using other analytic techniques, to integrate these findings into another study, or to synthesize several studies on the same topic. Such secondary analyses depend upon the author's reporting such basic details as means, standard deviations, error terms, derived scores and scales, marginals, and various validity and reliability coefficients.

There is, in other words, a *technology* for reporting on empirical research and a corresponding technology for verifying the report. But these tools seem confined to statistical studies. Most qualitative researchers are uncomfortable in the reporting straitjacket we have described, but they don't have an alternative to fall back on. There is no corresponding technology for qualitative studies. Lofland (1974, p. 101) says it well:

> Qualitative field research seems distinct in the degree to which its practitioners lack a public, shared and codified conception of how what they do is done and how what they report should be formulated.

On the face of it, this is a curious state of affairs. One of the strengths of qualitative research is precisely its capacity to *describe* in detail the empirical phenomena under study. Qualitative studies are rich in descriptions of settings, people, events, and processes, but they usually say little about *how* the researcher got the information, and almost nothing about *how* conclusions were drawn.

The problem, of course, is that we can't verify or do secondary analysis of a study in which procedures