4

FUNDAMENTALS OF QUALITATIVE DATA ANALYSIS

CHAPTER SUMMARY

This chapter reviews fundamental approaches to qualitative data analysis with a particular focus on coding data segments for category, theme, and pattern development. Other analytic strategies include jottings, memos, and the formulation of assertions and propositions. Within-case and cross-case analysis are then compared for their unique advantages and contributions to the research enterprise.

CONTENTS

INTRODUCTION

In this chapter, we describe fundamental methods for qualitative data analysis while data collection progresses. They help organize data for later, deeper analyses, such as those using the displays described in Chapters 5 through 10. The appendix provides recommended titles for additional analytic guidance.

Some qualitative researchers put primary energy into data collection for weeks, months, or even years and then retire from the field to "work over their notes." We believe this is a big mistake. It rules out the possibility of collecting new data to fill in gaps or to test new hypotheses that emerge during analysis. It discourages the formulation of rival hypotheses that question a fieldworker's routine assumptions. And it makes analysis into a giant, sometimes overwhelming task that frustrates the researcher and reduces the quality of the work produced.

We strongly advise analysis concurrent with data collection (see Display 1.1). It helps the fieldworker cycle back and forth between thinking about the existing data and generating strategies for collecting new, often better, data. It makes analysis an ongoing, lively enterprise that contributes to the energizing process of field work. Furthermore, early analysis permits the production of interim reports required in most evaluation and policy studies.

For the methods in this and following chapters, we assume that the fieldworker has collected information in the form of handwritten or typed field notes, audio or video recordings of interviews or other events in the field setting, and documents or other print/digital artifacts. We further assume that the basic, raw data (scribbled field notes, interview recordings) must be processed before they are available for analysis. Expanded field note write-ups, interview transcripts, and so on can be read easily, edited for accuracy, commented on, coded, and analyzed using several of the methods we later describe.

We begin with First Cycle *coding*, then second cycle or *pattern codes*, and the process of deriving even more general themes through *jottings* and *analytic memoing*. We then discuss *assertion and proposition development*, and conclude this chapter with a section on *within-case* and *cross-case analysis*. Our presentation here addresses only the fundamentals of analysis; Chapters 5 through 10 include additional methods and specific examples.

FIRST CYCLE CODES AND CODING

Description

Codes are labels that assign symbolic meaning to the descriptive or inferential information compiled during a study. Codes usually are attached to data "chunks" or units of varying size and can take

the form of a straightforward, descriptive label or a more evocative and complex one (e.g., a concept, value, or metaphor). Saldaña (2016) defines a code as

> most often a word or short phrase that symbolically assigns a summative, salient, essence-capturing, and/or evocative attribute for a portion of language-based or visual data. The data can consist of interview transcripts, participant observation field notes, journals, documents, open-ended survey responses, drawings, artifacts, photographs, video, Internet sites, e-mail correspondence, academic and fictional literature, and so on. The portion of data coded during First Cycle coding processes can range in magnitude from a single word to a full paragraph, an entire page of text or a stream of moving images. In second cycle coding processes, the portions coded can be the exact same units, longer passages of text, analytic memos about the data, and even a reconfiguration of the codes themselves developed thus far. Charmaz (2001) describes coding as the "critical link" between data collection and their explanation of meaning. . . .
>
> In qualitative data analysis, a code is a researcher-generated construct that symbolizes or "translates" data (Vogt, Vogt, Gardner, & Haeffele, 2014, p. 13) and thus attributes interpreted meaning to each individual datum for later purposes of pattern detection, categorization, assertion or proposition development, theory building, and other analytic processes. Just as a title represents and captures a book, film, or poem's primary content and essence, so does a code represent and capture a datum's primary content and essence. (p. 4)

Some research methodologists believe that coding is merely technical, preparatory work for higher level thinking about the study. But we believe that coding is deep reflection about and, thus, deep interpretation of the data's meanings. In other words, coding *is* analysis.

Codes are primarily, but not exclusively, used to retrieve and categorize similar data units so the researcher can quickly find, pull out, and cluster the segments relating to a particular research question, hypothesis, concept, or theme. Clustering and the display of condensed units then set the stage for further analysis and drawing conclusions.

For example, let's assume you were interested, as we were in the school improvement study, in the reasons why a new educational practice is adopted. You might begin by asking participants why they or others decided to try the practice. A piece of the formatted field notes might look like this:

> I asked the principal what the need for the new program was, and he responded that the students coming into the 9th grade were two years below grade level and that the old curriculum was ineffective. Through testing [the Nelson Reading Test] it was determined that students were growing academically only 5 or 6 months during the 10-month school year.

Assuming that you found it possible to apply a single summarizing notation or code to this unit, it might be MOTIVATION FOR ADOPTION (other codes could be applicable). That code would appear capitalized in the right-hand margin beside the segment (the left-hand margin might be used for a *jotting*, explained later):

[1] I asked the principal what the need for the new program was, and he responded that the students coming into the 9th grade were two years below grade level and that the old curriculum was ineffective. Through testing [the Nelson Reading Test] it was determined that students were growing academically only 5 or 6 months during the 10-month school year.	[1] MOTIVATION FOR ADOPTION

Other units of field notes or interview transcripts that also relate to MOTIVATION FOR ADOPTION would receive the same code.

Coding is a *heuristic*—a method of discovery. You determine the code for a unit of data by careful reading and reflection on its core content or meaning. This gives you intimate, interpretive familiarity with every datum in the corpus.

Rationale

As soon as the field researcher begins to compile information, challenges appear. A big one comes from the multiplicity of data sources and forms. Some information comes from structured or informal observations. More, if not most, comes from interviews. There are also everyday or special documents, archival records, and physical artifacts. In some studies, there can be information from questionnaires and surveys, photographs, videos, or statistical records.

All of this information piles up exponentially. In the early stages of a study, most of it looks promising. But if you don't know what matters more, everything matters. You may never have the time to condense and order, much less to analyze and write up, all of this material. That's why we think conceptual frameworks and research questions are the best defense against overload. They also reflect a point we made earlier: that data collection is inescapably a *selective* process and that you cannot and do not "get it all," even though you might think you can.

You may need roughly three to five times as much time for processing and ordering the data as the time you needed to collect them. Just one substantive week at a field site often can result in something like hundreds of pages of typed-up field notes, interview transcripts, documents, and ancillary materials. Not everything in a database has to be coded—just those units that relate directly to answering the research questions of interest, and those units that merit a place in the condensed empirical materials for their emergent importance to the study. Codes are *prompts* or *triggers* for deeper reflection on the data's meanings. Coding is thus a *data condensation* task that enables you to retrieve the most meaningful material, assemble units of data that go together, and further condense the bulk into readily analyzable units.

Codes are first assigned to data units in order to detect recurring *patterns*. From these patterns, similar codes are clustered together to create a smaller number of categories, themes, or *Pattern Codes*. The categories' interrelationships with each other are then constructed to develop higher level analytic meanings for assertion, proposition, hypothesis, and/or theory development.

FIRST CYCLE CODING EXAMPLES

Saldaña (2016) divides coding into two major stages: First Cycle and Second Cycle coding. First Cycle coding methods are codes initially assigned to the data units. Second Cycle coding methods generally work with the resulting First Cycle codes themselves.

First Cycle coding methods include over 25 different approaches, each one with a particular function or purpose. You do not need to stick with just one approach for your coding efforts; some of these can be compatibly "mixed and matched" as needed. Below is a review of some of the most pertinent ones that apply to the particular analytic approaches profiled in this book. See Saldaña's (2016) *The Coding Manual for Qualitative Researchers* for a fuller description of each method.

Newcomers to coding can get easily overwhelmed by the numerous approaches available to them. It can even induce a sense of analytic paralysis: "I don't know which coding method to use!" Each

profile below includes a few recommendations for when a particular coding method may be appropriate for your particular data, research goals, and methodological approach. Qualitative inquiry defies standardization, and that includes algorithmic prescriptions for codes and coding. You (in consultation with a peer or mentor) are the best judge for which method(s) is/are appropriate for your unique study.

First, there are four *Elemental Methods* that serve as foundation approaches to coding: Descriptive, In Vivo, Process, and Concept Coding.

Descriptive Coding

A descriptive code assigns labels to data that summarize in a word or short phrase—most often as a noun—the basic topic of a passage of qualitative data. These eventually provide an inventory of topics for indexing and categorizing, which is especially helpful for ethnographies and studies with a wide variety of data forms (field notes, documents, artifacts, etc.). Descriptive codes are perhaps more appropriate for social environments and settings rather than interviews. An example comes from field notes about a lower-middle-class neighborhood:

[1] As I walked toward the school, there was a 7-11 convenience store 1 block away, next to a small professional office building: an optometrist, podiatrist, and other medical/health-related clinics. Directly across the street was an empty lot, but next to that stood a Burger King restaurant.	[1] BUSINESSES

An analyst would extract all passages coded BUSINESSES from various field notes to compose a more detailed inventory of the case and to construct a narrative describing the business climate in the area.

In Vivo Coding

This is one of the most well-known qualitative coding methods. In Vivo Coding uses words or short phrases from the participant's own language in the data record as codes. It may include folk or indigenous terms of a particular culture, subculture, or microculture to suggest the existence of the group's cultural categories (e.g., in a hospital, you may hear unique terms such as "code blue," "sharps," and "scripts"). Phrases that are used repeatedly by participants are good leads; they often point to regularities or patterns in the setting.

In Vivo Coding is appropriate for virtually all qualitative studies, but particularly for beginning qualitative researchers learning how to code data, and studies that prioritize and honor the participant's voice. In Vivo Codes are placed in quotation marks to differentiate them from researcher-generated codes. Examples below are taken from a coded interview transcript about an adolescent girl's experiences with school:

I [1] hated school last year. Freshman year, it was awful, I hated it. And [2] this year's a lot better actually. I, um, don't know why. I guess, over the summer I kind of [3] stopped caring about what other people thought and cared more about, just, I don't know.	[1] "HATED SCHOOL" [2] "THIS YEAR'S A LOT BETTER" [3] "STOPPED CARING"

Process Coding

This coding method uses gerunds ("-ing" words) exclusively to connote observable and conceptual action in the data. Processes also imply actions intertwined with the dynamics of time, such as things that emerge, change, occur in particular sequences, or become strategically implemented. Process Coding is appropriate for virtually all qualitative studies, but particularly for grounded theory research that extracts participant action/interaction and consequences. An example comes from an interview transcript about an adolescent girl explaining how rumors get spread:

Well, that's one problem, that [my school is] pretty small, so [1] if you say one thing to one person, and they decide to tell two people, then those two people tell two people, and in one period everybody else knows. [2] Everybody in the entire school knows that you said whatever it was.	[1] SPREADING RUMORS [2] KNOWING WHAT YOU SAID

Concept Coding

Concept Codes assign meso- or macrolevels of meaning to data or to data analytic work in progress (e.g., a series of codes or categories). A concept is a word or short phrase that symbolically represents a suggested meaning broader than a single item or action—a "bigger picture" beyond the tangible and apparent. A concept suggests an idea rather than an object or observable behavior. For example, a clock is something you can touch and see as it changes minute to minute, but its conceptual attribution is *time*. One can see and touch a church building, but not the concepts of *spirituality* or *religion*. Note that these concepts are nouns.

Concepts also refer to processes such as *surviving* or *coping*. Conceptual processes consist of smaller observable actions that add up to a bigger and broader scheme. For example, I can spend an afternoon at a shopping mall window shopping, buying things, and eating lunch at a food court, but the overall concept is *consuming*. Visiting Starbucks five mornings a week, for some, is not just for purchasing and drinking coffee (the observable actions), but for *performing a morning ritual*, *maintaining a workday habit*, or *feeding an addiction*. Note that these concept phrases begin with gerunds ("-ing" words) to describe broader processes at work.

Concept Codes tend to be applied to larger units or stanzas of data—an analytic "lumping" task that harmonizes with the bigger picture suggested by a concept. Concept Coding is an appropriate method when the analyst wishes to transcend the local and particular of the study to more abstract or generalizable contexts. The method bypasses the detail and nuance of other coding methods to transcend the particular participants of your fieldwork and to progress toward the *ideas* suggested by the study.

Concept Coding is illustrated here as a stand-alone method, but it is worth noting that most First Cycle methods such as Process, Values, and Holistic Coding can utilize codes that are conceptual in nature. A series or categorized collection of First Cycle codes can be condensed even further into a Concept Code.

Below is an example of an Irishman speaking to an American tourist about his perceptions of America. Note that the Concept Code attached to the datum is an evocative choice and, in part, employs an In Vivo Code:

[1] What I love about America is the excess. You know, I'll walk into a coffee shop here [in Ireland] and all there'll be is milk. I go into a coffee shop in America, and there's 1%, 2%, skim milk, whole milk, half and half, you know? And the highways! There's two lanes at most here goin' one way, and over there you've got five, six lanes of highway all goin' in the same direction. We just don't have the cars or the land for that much highway. And the malls—you'll never see things like that over here. It's usually one shop at a time, but there you've got it all under one roof. Big stores, too, big stores.	[1] AMERICAN "EXCESS"

Next, there are three *Affective Methods* that tap into the more subjective experiences we encounter with our participants: Emotion, Values, and Evaluation Coding.

Emotion Coding

Perhaps obviously, this method labels the emotions recalled and/or experienced by the participant, or inferred by the researcher about the participant. Emotion Coding is particularly appropriate for studies that explore intrapersonal and interpersonal participant experiences and actions. It also provides insight into the participants' perspectives, worldviews, and life conditions. Note that a participant himself or herself may sometimes label the emotion, and thus, it should be In Vivo Coded in quotation marks. An example comes from an interview transcript about a middle-aged man complaining about one of his work colleagues:

[1] I just hated it when he got awarded with the honor. [2] I mean, we're praising mediocrity now. Never mind that what you've accomplished isn't worth squat, it's all about who you know in the good ol' boys network.	[1] "HATED IT" [2] BITTERNESS

Values Coding

This is the application of three different types of related codes onto qualitative data that reflect a participant's values, attitudes, and beliefs, representing his or her perspectives or worldview. A value (V:) is the importance we attribute to ourselves, another person, a thing, or an idea. An attitude (A:) is the way we think and feel about ourselves, another person, a thing, or an idea. A belief (B:) is part of a system that includes values and attitudes, plus personal knowledge, experiences, opinions, prejudices, morals, and other interpretive perceptions of the social world. Values Coding is appropriate for studies that explore cultural values, identity, intrapersonal and interpersonal participant experiences and actions in case studies, appreciative inquiry, oral history, and critical ethnography. An example comes from an interview transcript about a female university student discussing her political beliefs:

[1] Government regulation of women's health issues has gotten out of hand. It's not about "protecting" us, it's about their need to control and dominate women [2] through covert religious ideology. White Christian men are deciding what's law and what's moral and what's, how it's supposed to be. [3] They can say, "It's *not* a war on women" all they want, but trust me—it's a war on women.	[1] B: GOVERNMENT CONTROL [2] B: COVERT RELIGIOUS MOTIVES [3] A: MISOGYNIST MOTIVES

Evaluation Coding

This method applies primarily nonquantitative codes onto qualitative data that assign judgments about the merit, worth, or significance of programs or policy. Evaluation Coding is appropriate for policy, critical, action, organizational, and evaluation studies, particularly across multiple cases and extended periods of time. Selected coding methods profiled thus far, such as Descriptive or In Vivo Codes, can be applied to or supplement Evaluation Coding, but the methods are customized for specific studies. A + symbol before a code tags it as a positive evaluation. Second-order codes that follow a primary code and a colon are called Subcodes (discussed later). An example comes from an interview transcript about an elementary school teacher assessing an artist-in-residency program:

[1] The artist-in-residency program was pretty successful this year. [2] The arts agency did a great job at selecting qualified candidates this time around. [3] We were pretty impressed at how they integrated math and geometry with art-making without the teachers telling them to. I think they knew the score and that it was pretty important that they cover those subject areas. And they did it in a way that made it [4] interesting for the kids. For the teachers, too! We learned some things that we can integrate into our own curriculum next year.

[1] + RESIDENCY: "SUCCESSFUL"

[2] + CANDIDATES: QUALIFIED

[3] + CURRICULUM: INTEGRATION

[4] + CURRICULUM: "INTERESTING"

One *Literary and Language Method*, Dramaturgical Coding, explores human action, reaction, and interaction through strategic analysis of people's motives.

Dramaturgical Coding

This method applies the terms and conventions of character, play script, and production analysis onto qualitative data. For character, these terms include items such as participant

- objectives (OBJ)
- conflicts (CON)
- tactics (TAC)
- attitudes (ATT)
- emotions (EMO)
- subtexts (SUB)

Dramaturgical Coding is appropriate for exploring intrapersonal and interpersonal participant experiences and actions in case studies, power relationships, and the processes of human motives and agency. An example comes from an interview transcript about a community college instructor's dilemmas with her unit's budget cuts:

[1] There was a lot of pressure this year to "do more with less." And that always [2] frustrates me, because you don't "do more with less"—you do *less with less*. So [3] if they're expecting me to do more with less money and less resources, they're not

[1] CON: LESS RESOURCES
[2] EMO: FRUSTRATION
[3] TAC: RESISTANCE

going to get it. And it's not because I'm being snotty or passive-aggressive about this; [4] it's simply that you can't squeeze blood out of a turnip. There's only so much you can do with what you have. [5] And yes, I'm spending some of my own money this year on classroom supplies because we don't have enough to last us through the end of the year. [6] That's just the way it is these days.

[4] ATT: LIMITATIONS

[5] TAC: SACRIFICING

[6] ATT: ACCEPTING "THE WAY IT IS"

Three *Exploratory Methods*—Holistic, Provisional, and Hypothesis Coding—make preliminary or global coding assignments, based on what the researcher deductively assumes may be present in the data before they are analyzed.

Holistic Coding

This method applies a single code to a large unit of data in the corpus, rather than detailed coding, to capture a sense of the overall contents and the possible categories that may develop. Holistic Coding is often a preparatory approach to a unit of data before a more detailed coding or categorization process through First or Second Cycle methods. The coded unit can be as small as one-half a page in length, to as large as an entire completed study. Holistic Coding is most applicable when the researcher has a general idea of what to investigate in the data. An example comes from field notes by a researcher observing how new, tenure-track faculty become oriented to academia:

[1] The chair of the committee debated whether to start on time or to wait for latecomers to join the meeting. "We all made the effort to be here at 8:00 a.m., so let's start," he said. The network meeting began with obligatory self-introductions of the 6 people seated around a large table designed to hold 12. Most attendees were newcomers to academia at the assistant professor or faculty associate level, primarily from midwest and east coast institutions. Each one appeared to be in his or her late 20s or early 30s. "You're the new guard of the college," said the chair, "and we're here to find ways to network and support each other as you begin your teaching careers."

[1] THE "NEW GUARD"

Provisional Coding

This approach begins with an a priori or "start" list of researcher-generated codes, based on what preparatory investigation suggests might appear in the data before they are collected and analyzed. Provisional Codes can be revised, modified, deleted, or expanded to include new codes. This method is appropriate for qualitative studies that build on or corroborate previous research and investigations. For example, a researcher about to interview people who successfully quit smoking may develop the following Provisional Codes of smoking cessation methods beforehand:

PRESCRIPTION MEDICATION

NICOTINE PATCHES

NICOTINE GUM/LOZENGES

ELECTRONIC CIGARETTES

PROFESSIONAL COUNSELING

PEER SUPPORT SYSTEM

"COLD TURKEY"

Hypothesis Coding

This is the application of a researcher-generated, predetermined list of codes onto qualitative data specifically to assess a researcher-generated hypothesis. The codes are developed from a theory/ prediction about what will be found in the data before they have been collected or analyzed. Statistical applications, if needed, can range from simple frequency counts to more complex multivariate analyses. This method is appropriate for hypothesis testing, content analysis, and analytic induction of the qualitative data set, particularly the search for rules, causes, and explanations in the data.

Hypothesis Coding also can be applied midway or later in a qualitative study's data collection or analysis to confirm or disconfirm any assertions, propositions, or theories developed thus far. For example, it is hypothesized that the responses to a particular question about language issues in the United States will generate one of four answers (and thus coded responses) from participants:

RIGHT = We have the right to speak whatever language we want in America

SAME = We need to speak the same language in America: English

MORE = We need to know how to speak more than one language

NR = No Response or "I don't know"

Two *Procedural Methods* utilize specific rather than open-ended ways of coding data: Protocol and Causation Coding.

Protocol Coding

This is the coding of qualitative data according to a preestablished, recommended, standardized, or prescribed system. The generally comprehensive list of codes and categories provided to the researcher are applied after his or her own data collection. Some protocols also recommend specific qualitative (and quantitative) data analytic techniques with the coded data. Protocol Coding is appropriate for qualitative studies in disciplines with previously developed and field-tested coding systems. It can also be used to code data collected across multiple sites for cross-case comparison. For example, a selected list of codes from a protocol used to determine causes of family violence include the following:

ALCOH = alcoholism or drinking

DRUG = drug use

EDUC = lack of education

MONEY = lack of money or financial problems

Causation Coding

This method extracts attributions or causal beliefs from participant data about not just how but why particular outcomes came about. The analyst searches for combinations of *antecedent conditions* influenced and affected by certain *mediating variables* that lead toward certain *outcome* pathways, and attempts to map a three-part process as a CODE 1 → CODE 2 → CODE 3 sequence.

Causation Coding is appropriate for discerning motives, belief systems, worldviews, processes, recent histories, interrelationships, and the complexity of influences and affects on human actions and phenomena. This method may serve grounded theorists in searches for causes, conditions, contexts, and consequences. It is also appropriate for evaluating the efficacy of a particular program, or as preparatory work before diagramming or modeling a process through visual means such as decision modeling and cognitive networks. For example, a survey respondent describes in writing what challenges she faced when she took speech classes in high school. The + symbol refers to a combination of conditions, variables, or outcomes that are mentioned by the participant as connected parts of the causation sequence; the → symbol means "leads to":

[1] Without a doubt, it was a fear of speaking in front of others. My ultimate career as an adult was in the field of journalism. Early fears I had about approaching strangers and speaking in front of a group of people were overcome due to involvement in speaking events. As I mentioned above, I think speech class and the events that I participated in due to taking that class, probably led directly to my choosing journalism as a career. My success in the field of journalism would have never come about without those speech classes in high school.

[1] "FEAR OF SPEAKING" → SPEAKING EVENTS + SPEECH CLASS → JOURNALISM CAREER + SUCCESS

Four *Grammatical Methods* play a role in the mechanics of coding: Attribute Coding, Magnitude Coding, Subcoding, and Simultaneous Coding.

Attribute Coding

This method is the notation of basic descriptive information such as the field work setting, participant characteristics or demographics, data format, and other variables of interest for qualitative and some applications of quantitative analysis. This is appropriate for virtually all qualitative studies, but particularly for those with multiple participants and sites, cross-case studies, and studies with a wide variety of data forms. Attribute Coding provides essential participant information for future management, reference, and contexts for analysis and interpretation. Examples from a data set about an educational study include the following:

CASE: Martinez School

PARTICIPANT: Nancy (pseudonym)

INTERVIEW: 2 of 5

INTERVIEW TOPICS:

    Evaluation of School Day

    Salary Issues

    Principal–Teacher Relationship

    Upcoming Extracurricular Activities

    Upcoming Fundraising Project

Magnitude Coding

Magnitudes consist of supplemental alphanumeric or symbolic codes or Subcodes to existing coded data or a category to indicate their intensity, frequency, direction, presence, or evaluative content. Magnitude Codes can be qualitative, quantitative, and/or nominal indicators to enhance

description. They are most appropriate for mixed methods and qualitative studies in education, social science, and health care disciplines that also support quantitative measures as evidence of outcomes. Examples used in the school improvement study include the following:

MAJOR

MODERATE

MINOR

√√ = Yes, clearly

√ = Possibly, in part

0 = No

++ = Very effective

+ = Effective

± = Mixed

We argue that although words may be more unwieldy than numbers, they render more meaning than numbers alone and should be hung on to throughout data analysis. Converting words into numbers and then tossing away the words gets a researcher into all kinds of trouble. You are assuming that the chief property of the words is that there are more of some than of others. Focusing solely on numbers and quantities shifts attention from substance to arithmetic, throwing out the whole notion of their qualities or essential characteristics. A solution to this problem, as we will see in later sections and displays, is to keep words and any associated magnitudes (LOW, EFFECTIVE, √) *together* throughout the analysis.

Subcoding

A Subcode is a second-order tag assigned after a primary code to detail or enrich the entry. The method is appropriate for virtually all qualitative studies, but particularly for ethnographies and content analyses, studies with multiple participants and sites, and studies with a wide variety of data forms. Subcoding is also appropriate when general code entries will later require more extensive indexing, categorizing, and subcategorizing into hierarchies or taxonomies, or for nuanced qualitative data analysis. It can be employed after an initial yet general coding scheme has been applied and the researcher realizes that the classification scheme may have been too broad, or added to primary codes if particular qualities or interrelationships emerge. This example comes from a set of field notes describing a school's facilities:

[1] The school's multipurpose room functions as a cafeteria, auditorium, assembly space, meeting space, and study hall. Its portable tables with attached seating fold up easily for somewhat quick transformation and cleaning of the space. [2] The adjoining media center houses books, a computer lab with 26 stations, study "nooks" for small groups, and various tables and chairs. A large screen and LCD projector suspended from the ceiling make the space look like a private movie theater.

[1] SCHOOL - MULTIPURPOSE SPACE

[2] SCHOOL - MEDIA CENTER

Simultaneous Coding

This is the application of two or more different codes to a single qualitative datum, or the overlapped occurrence of two or more codes applied to sequential units of qualitative data. The method is

appropriate when the data's content suggests multiple meanings (e.g., descriptively and inferentially) that necessitate and justify more than one code. Simultaneous Coding should be used sparingly, however; too many codes suggest a vague handle on the research questions and study's purpose. An example of the method comes from field notes about an organizational study of a community theatre program:

[1 & 2] The board of directors struggled with ways to keep the community theatre program going for another full season. It had been a staple in the area for almost 40 years, but now faced (like many comparable programs) the end of its existence. Less financial contributions and lower box office revenue had put the theatre severely in the red. Long-time volunteers and members were thinking with their hearts more than with their heads as they claimed that the "traditions" of this program could not be ended. The board felt otherwise, for none of its members wanted the liability of debt.	[1] FINANCIAL LOSS [2] END OF TRADITION

One stand-alone method, *Themeing the Data*, elaborates on what may be perceived as the terse short-hand of codes and coding.

Themeing the Data

In addition to codes, extended thematic phrases can be applied to units of data. Themeing is perhaps most applicable to phenomenology, but the method also lends itself to the analysis of qualitative data for comparison to comparably themed quantitative data for mixed methods studies.

Overall, a theme is an *extended phrase* or *sentence* that identifies what a unit of data is *about* and/or what it *means*. To Rubin and Rubin (2012), themes are statements *qua* (in the role of) ideas presented by participants during interviews that summarize what is going on, explain what is happening, or suggest why something is done the way it is (p. 118). It is important to note that coding and themeing are not either/or procedures. Analysts, if they wish, can code their data first, cluster the codes according to commonality, and construct an extended thematic statement rather than a short category label from the assemblage.

In the interview transcript below, a health care researcher examines the concept of "recovery" by assigning extended phrases or brief sentences to portions of the participant's story about a knee injury.

[1] I injured my right knee in a sports accident in college. It was just a bunch of guys playing football, not a college team or anything. But I squatted down as I was running and I guess I twisted my leg in a direction it wasn't supposed to go and it just hurt like hell. The pain was in the back of my knee, here. [2] I went to the emergency room, had to be on crutches for a few weeks, it hurt like a bitch but eventually healed up. [3] But I've still got a scar or something inside me, a friend told me it was a flap of some kind inside me, that can go out of whack if I move my leg the wrong way. And if it does, my knee freezes up, it locks, and I have to bend it back in place carefully by twisting my leg this way. [4] It's a pain to be so careful with how I move now because it can go out on me unexpectedly.	[1] INJURY FROM UNNATURAL BODY MOVEMENT [2] RECOVERY IS PAINFUL BUT EVENTUAL [3] RECOVERY LEAVES INTERNAL SCARS [4] RECOVERY MEANS ATTENTIVE CAUTION

CODING CRAFT AND DESIGN

Before we proceed to Second Cycle Coding, we need to discuss a few technical matters related to the craft and design of codes and coding.

Levels of Coding Detail

How detailed should coding be? That depends on the study and your goals. More typically, codes get applied to larger units—sentences, monothematic "chunks" of sentences in interview transcripts, or full paragraphs in written-up field notes. If you are coding by hand on hard copies of data and the margins get piled up with multiple codes, you are in for heavy sledding when the codes need to be clustered for Second Cycle Pattern Coding. This problem is not critical when data have been coded with CAQDAS (Computer Assisted Qualitative Data Analysis Software), or even with Microsoft Office programs like Word and Excel (see Display 2.10).

Finally, not every portion of the field notes or interview transcripts must be coded. There are things such as trivial, irrelevant data. Most field notes and selected portions of transcripts usually contain much dross—material unrelated to the research questions, either prespecified or emerging. And, if done carefully, coding of later material can be more sparing.

A Priori Codes

One method of creating codes is developing a provisional "start list" of codes prior to field work—*deductive* or a priori coding. The list comes from the literature review, conceptual framework, list of research questions, hypotheses, problem areas, and/or key variables that the researcher brings to the study.

In our (Miles and Huberman) school improvement study, for example, we conceptualized the innovation process, in part, as one of RECIPROCAL TRANSFORMATIONS. Teachers change the characteristics of new practices. Those practices, in turn, change the teachers and modify working arrangements in the classroom, which, in turn, influence how much of the innovation can be used, and so on.

We began with a master code—TRANSFORMATION or TRANS for short—to indicate the transformational process we had hypothesized, plus some Subcodes—TRANS-USER, TRANS-CLASS (classroom changes), TRANS-ORG (organizational changes), TRANS-INN (changes in the innovation)—to mark off segments of data in each class of variables. The list was held lightly, applied to the first sets of field notes, and then examined closely for fit and utility. Quite a few codes were revised, but the conceptual orientation seemed to bear real fruit—to fit and account well for what we saw and heard.

A start list can have from a dozen or so up to 50 codes; that number can be kept surprisingly well in the analyst's memory without constant reference to the full list—*if* the list has a clear structure and rationale. It is a good idea to get that list on a single sheet for easy reference. Most CAQDAS programs can retain these Provisional Codes before data are entered into their programs.

Still other codes emerge progressively during data collection—*inductive* coding. These are better grounded empirically and are especially satisfying to the researcher who has uncovered an important local factor. They also satisfy other readers, who can see that the researcher is open to what the site has to say rather than determined to force-fit the data into preexisting codes. Most field researchers, no matter how conceptually oriented, will recognize when an a priori coding system is ill molded to the data or when a rival perspective looks more promising.

Revising Codes

For all approaches to coding, several codes will change and evolve as field experience continues. Researchers with start lists know that codes will change; there is more going on out there than our initial frames have dreamed of, and few field researchers are foolish enough to avoid looking for these things.

Some codes do not work; others decay. No field material fits them, or the way they slice up the phenomenon is not the way the phenomenon appears empirically. This issue calls for doing away with the code or changing its type (e.g., transforming a noun-based Descriptive Code such as COUNSELING CENTER into an action-oriented Process Code such as REHABILITATING CLIENTS). Other codes flourish, sometimes too much so. Too many segments get the same code, thus creating the familiar problem of bulk. This problem calls for breaking down codes into Subcodes.

With manual coding, revision is tedious: Every unit you have coded before has to be relabeled. But the search-and-replace facility of your text-based software and most CAQDAS programs can accomplish this easily.

STRUCTURE AND UNITY IN CODE LISTS

Whether codes are created and revised early or late is less important than whether they have some conceptual and structural unity. Codes should relate to one another in coherent, study-important ways; they should be part of a unified structure. Incrementally adding, removing, or reconfiguring codes is certainly permissible, so long as some sense of "belonging" is maintained.

Display 4.1 exhibits the coding system developed by Westerman, Cobham, and McDermott (2017) for their *Qualitative Health Research* article, "Trauma-Focused Cognitive Behavior Therapy: Narratives of Children and Adolescents." Young people with PTSD after a local flood disaster were interviewed by therapists. The transcripts were coded using concepts from the professional literature and the researchers' emergent design, which resulted in three primary categories with their related codes. The co-authors explain that

> The three concepts of *coherence*, *elaboration*, and *evaluation* formed a functional matrix with which to describe the patterns within the stories. *Coherence* was divided into structure (beginnings, sequences, and endings), orientation (time, place, and persons), and normative referents. *Elaboration* was divided into description (sensory and kinesthetic information), emotion, and sense of self. *Evaluation* was divided into causality, insight, and integration. (Westerman, Cobham, & McDermott, 2017, p. 228; emphasis added)

DISPLAY 4.1 ⬡ Details of the Coding System

Coherence: General

   A coherent narrative is one that can be understood within an exchange

Coherence: Structure

   How beginnings are constructed

   The sequencing of information

   How endings are constructed

(Continued)

(Continued)

Coherence: Orientation

> An awareness of time, setting, and persons

Coherence: Normative

> The cultural context evident within the story

Elaboration: General

> The complexity of detail in the stories (stories may be sparse or complex)
>
> The amount of detail may indicate accessibility, or avoidance, of memories

Elaboration: Description

> Includes sensory details (sight, sound, touch, taste, smell, kinesthetic)

Elaboration: Emotion

> The level of articulation of emotional domains
>
> Intensity of expressed emotion
>
> Both internal feelings and awareness of the feelings of others

Elaboration: Sense of self

> The use of words that indicate thinking, knowing, remembering, intentions, perspective taking or positioning, relatedness, and values

Evaluation: Causality

> Explanations, links between information, and plausibility
>
> May relate to action, emotion, cognitions, or social relationships

Evaluation: Insight

> Self-reflectivity and criticality
>
> Flexibility and a tolerance of contradictions and uncertainties
>
> Immediate lessons or global understanding that are gained from the traumatic experience

Evaluation: Integration

> Meaning making in sense of self and identity in relation to information, agency, and predictive ability

Source: Westerman, Cobham, & McDermott (2017), p. 229.

The co-authors sought a balance between simple and more complex coding systems that had been developed in the narrative psychology literature. With this elegant coding system, the research team was able to provide a concrete data example for each code, which included the percentage of stories that were assigned the code.

There is a "good fit" between the data and the emergent codes, but also between and among the categories and codes themselves in Display 4.1. Their efficacy was also noted by the coauthors, who concluded, "The coding system developed for this study elicited a wealth of information about the changes that occurred within the trauma narratives as therapy progressed. Our results suggest that narratives within this clinical intervention do have the potential to yield useful and specific information about change processes for clinicians" (p. 230).

An operative coding scheme is not simply a list or index of disjointed descriptors but rather a conceptual web, including larger meanings and their constitutive characteristics. CAQDAS is especially helpful in displaying the structure of coding schemes, either in hierarchical or taxonomic form or in a network.

DEFINITIONS OF CODES

Whether codes are prespecified or developed along the way, clear operational definitions are indispensable so they can be applied consistently by a single researcher over time, and multiple researchers will be thinking about the same phenomena as they code. A First Cycle code can consist of a single term—for example, TRANSFORMATION—that can easily suggest different meanings to different analysts. Because codes will drive the retrieval and organization of the data for analysis, they must be precise and their meaning shared among analysts. Defining them helps on both counts.

McHugh, Horner, Colditz, and Wallace (2013) explored through focus group interviews with diverse youth their "Bridges and Barriers: Adolescent Perceptions of Student-Teacher Relationships." This *Urban Education* article features as an appendix a rarely seen abbreviated codebook that includes each code's definition employed for the study's analysis (see Display 4.2).

DISPLAY 4.2 ⬡ Abbreviated Codebook

Barriers

   Back-Channel Communication: incident in which two or more people are talking about another; most frequently, topic person is not present (though this may occur), and such communications are inferred by or later related to topic person.
   Boundaries: participant expresses that someone is expected to act or speak in some way based on rank, role, or relationship.
   Carelessness: instances or occurrences in which the participant or the person about whom the participant is relating failed to adequately take into consideration another person's needs, desires or likes/dislikes; may refer to verbal, physical, engagement, emotive, or relational.
   Consequence: a repercussion of a given action, behavior, statement or incidence.
   Inattention: instances or occurrences in which the participant or the person about whom the participant is relating failed to adequately attend to another person's actions or communicated thoughts or feelings; may refer to verbal, physical, emotive, or relational.
   Stereotype: when a person is presumed to behave, believe, think, or experience some emotion simply because of some external categorization; typified by an intensified, overgeneralized, and frequent reinterpretation of a perceived attribute to an extreme degree and applied regardless of contrary fact/observations.
   Unspecified Barriers: processes that act as barriers to a relationship growing closer, but that do not meet the definitional criteria of another specific barrier process.

Bridges

   Connecting: participant refers to or describes the process by which someone transitions into "being known" by the participant, or by which the participant is "known" by them; the act of connecting with another person, or growing closer.
   Effortful Engagement: an instance in which one person actively and deliberately engages another on an interpersonal level.
   Help Seeking: the act of requesting and/or receiving assistance from another person; may be educational, verbal, financial, physical, or advisory; the request for one to enact one's resources for the benefit of another.
   Personal Disclosure: participant describes or relates an occurrence in which they either shared something personal with another person, or when another person shared something personal with the participant; may be positive, neutral, expression of uncertainty or vulnerability, or any of these in combination with negative.
   Prediction and Interpretation: when a person demonstrates the ability to predict another person's reaction (verbal, active, emotive, etc.) to an action (verbal, active, or emotive) performed by the predictor, and to accurately gauge their intention underlying communication, actions undertaken, etc.

(Continued)

(Continued)

(Understanding) Constraints: temporal, monetary, physical, familial/relational, etc. features of a participant's life which, per participant report, may make connecting to others or completing a given task a more complicated and/or difficult proposition than would otherwise be.

Support

Unspecified Support: the act of providing or sharing emotional assistance, reassurance or bolstering, or other form of support that does not meet the definitional criteria of another specific support process.

Catharsis: mutual emotion sharing or emotional expression between two or more persons.

Companionship: the desire to share time, physical space, or communication with another person, typically in a friendly or emotionally positive manner.

Conflict Resolution: the act between two or more people to work collaboratively on resolving a point of disagreement, conflict, misunderstanding, etc.; primarily verbal, though may incorporate resource allocation.

Empathizing: an attempt to share in or re-create another's emotional experience.

Leniency/Willingness to Work Out Difficulties: the emotional and cognitive intention and commitment to resolving conflicts, differences of opinion, etc., that occur between two or more persons; this implies the intention to apply effort.

Problem Solving: request and/or receipt of assistance from one person by another in dealing with a real or hypothetical problem, conflict, tumultuous situation, or issue; primarily verbal in nature, and tends to be focused on social situations or circumstances.

Protection: the act of buffering or guarding another person; may be verbal (e.g., verbal defense), physical, legal, or emotional/relational.

Resource: various different types or degrees of assistance that one person can or does offer another person; may be financial, emotional, temporal, protective, work/education, etc.

Sharing of Difficulties: discussion of the revelation of personal problems, concerns, or issues to another person; participant may be discloser, or disclosed-to, but specific subject matter regards a negative occurrence in the discloser's life.

Soothing: the desire of a person for the consideration, kind words, or soothing actions of another person; the act of providing the above.

Structural Codes

Teacher Does: incidents in which a participant relates a specific instance or example of behavior that was specifically done by the teacher.

Teacher Does Not: incidents in which a participant relates a specific instance or example of behavior that was specifically NOT done by the teacher; these incidents are most typically paired with a value judgment.

Teacher Should: incidents in which a participant relates a specific instance when they feel that a teacher SHOULD have done something (coded regardless of actual occurrence of the behavior), or example of (potential or actual) teacher behavior that the student relayed feelings (or an affirming value judgment) that such behavior SHOULD happen.

Teacher Should Not: incidents in which a participant relates a specific instance when they feel that a teacher SHOULD NOT have done something (coded regardless of actual occurrence of the behavior), or example of (potential or actual) teacher behavior that the student relayed feelings (or a negative value judgment) that such behavior SHOULD NOT happen.

Source: McHugh, Horner, Colditz, & Wallace (2013), pp. 35–38.

The co-authors detail their meticulous coding procedures and the difficulties they encountered with overlapping codes applied to ever-shifting topics in focus group transcripts (see Chapter 11). Their emergent system eventually honed in on the central purpose of the study, which enabled them to

deliberately label the perceived effect of the interaction on the closeness of the interpersonal relationship. As such, we were able to explore the driving question in this coding process, "Did the interaction *bridge* the two people to bring them closer together, or did it separate them or create a *barrier* between closeness developing?" (McHugh et al., 2013, p. 18)

The codebook in Display 4.2 is an exemplar of fine-grained, nuanced definitions essential for analyzing complex relationship dynamics in multicultural settings.

Definitions become sharper when two researchers code the same data set and discuss their initial difficulties. A disagreement shows that a definition has to be expanded or otherwise amended. Time spent on this task is not hair-splitting but reaps real rewards by bringing you to a common vision of what the codes mean and which units of data best fit which code.

Team coding not only aids definitional clarity but also is a good reliability check. Do two coders working separately agree on how big a codable block of data is? And do they use roughly the same codes for the same blocks of data? If not, they are headed for different analyses and need to reconcile their differences for more credible and trustworthy findings.

Similarly, each coder is well-advised to code the first dozen pages of field notes, once right away and again (on an uncoded copy) a few days later. How good is the internal consistency? Eventually, intra- and/or intercoder agreement should be within the 85% to 90% range, depending on the size and range of the coding scheme. There are more robust statistical measures of agreement available such as Pearson's *r* and the Kappa coefficient for those who must provide quantitative evidence to readers. We believe, however, that a simple high percentage of agreement is sufficient for most qualitative studies.

SECOND CYCLE CODING—PATTERN CODES

Description

First Cycle coding is a way to initially summarize segments of data. Pattern Coding, as a Second Cycle method, is a way of grouping those summaries into a smaller number of *categories*, *themes*, or *concepts*. For quantitative researchers, it's an analog to the cluster-analytic and factor-analytic devices used in statistical analysis. Its parallel to grounded theory coding methods is Focused Coding or Axial Coding.

Pattern codes are inferential or explanatory codes, ones that identify a "bigger picture" configuration. They pull together a lot of material from First Cycle coding into more meaningful units of analysis. They are a sort of meta-code.

Applications

For the qualitative analyst, Pattern Coding has four important functions:

1. It condenses large amounts of data/codes into a smaller number of analytic units (e.g., categories, themes, concepts).

2. It gets the researcher into analysis during data collection, so that later fieldwork can be more focused.

3. It helps the researcher elaborate a cognitive network—an evolving, more integrated schema for understanding local incidents and interactions.

4. For multiple-case studies, it lays the groundwork for cross-case analysis by surfacing common themes and directional processes.

These four functions can be clarified as we discuss how Pattern Codes are generated, what they look like, and what the field researcher does with them in the course of data collection.

Examples

Generating Pattern Codes

During initial fieldwork, the researcher is looking for threads that tie together bits of data. For example, if two or three participants say independently that they "resent" a decision made by their boss, we may be on to several different phenomena—a conflict, an organizational climate factor, or a disgruntled subgroup of employees. Any of these interpretations involves chunking and sorting data (Function 1, above). For starters, is there anything else in common between these participants or in the site given for resenting the decision? Is there a different or opposing perspective among participants who are *not* resentful?

These first bits of data and the review of the coded segments being pulled together are leads; they suggest important variables to check out—factors that may account for other local perceptions and behaviors (Function 2, above). Seeing the RESENTMENT data (a First Cycle Emotion Code) in any of these alternative ways also helps the researcher make sense of puzzling or surprising observations. These several bits come together into an initial plot of the terrain (Function 3). Finally, if a colleague in a multi-case study comes across a similar profile of resentment or, alternatively, finds no resentment of decisions at all in a place otherwise similar to the more "resentful" case, we have the first threads of cross-case comparison (Function 4).

Patterning happens quickly because it is the way we instinctively and naturally process information. The danger is getting locked too quickly into naming a pattern, assuming you understand it, and then thrusting the name onto data that fit it poorly. The trick here is to work with loosely held units of meaning, to be ready to unfreeze and reconfigure them as the data shape up otherwise, to subject the most compelling themes to cross-checking, and to lay aside the more tenuous ones until other participants and observations give them better empirical grounding. In a more inductive study, it helps to look for recurring phrases (i.e., In Vivo Codes) or common threads in participants' accounts or, alternatively, for internal differences that you or participants have noted. Typically, those differences will bring forth a higher level commonality.

What Pattern Codes Look Like

Pattern Codes usually consist of four, often interrelated, summarizers:

1. Categories or themes

2. Causes or explanations

3. Relationships among people

4. Concepts or theoretical constructs

Below are some concrete examples of Pattern Codes in capital letters, followed by their brief definitions.

Categories or Themes

RULES: You don't "shop talk" in the staff lounge; the unspoken understanding is that social small talk to decompress is OK; complaining is also acceptable but without generating solutions to problems.

TRAJECTORIES: The metaphor of career "trajectories"—people are using these projects to get away from some jobs and places to other ones.

Causes or Explanations

DYSFUNCTIONAL DIRECTION: Staff perceptions of and interactions with ineffective leadership that influence workplace morale and effectiveness.

BEST PRACTICES: The best projects are ones that put together the best practitioners' tested recipes for success.

Relationships Among People

LEADERS' NETWORK: The unofficial collective of individuals seen as key leaders at their respective sites: A. Becker, P. Harrison, and V. Wales.

NEW GUARD: The new, thirtyish generation of faculty members with an aggressive yet socially conscious edge to them.

Concepts or Theoretical Constructs

BARGAINING: Bargaining or negotiating, most often covertly, seems to be the way decisions get made; a conflict model is a more plausible account of how actions get initiated, rather than through cooperative teamwork.

SURVIVAL: A defeatist, mostly negative attitude that suggests one is working on a day-to-day basis with minimal resources and support to accomplish much against sometimes overwhelming odds.

Pattern codes can emerge from repeatedly observed routines, rituals, rules, roles, and relationships; local meanings and explanations; commonsense explanations and more conceptual ones; inferential clusters and "metaphorical" ones; and single-case and cross-case observations.

Using Pattern Codes in Analysis

It may be useful at some point to "map" the Pattern Codes—that is, to lay out the component codes that got you the pattern—along with segments from the field notes. It helps to do it visually, in a network display, seeing how the components interconnect. The mapping is a new take on your conceptual framework. Although it is not hard to do this by hand, mapping by computer and CAQDAS has some powerful advantages and does this well.

Next, the most promising codes to emerge from this exercise are written up in the form of an analytic memo (discussed later) that expands on the significance of the code. This process helps the writer become less fuzzy about the emergent category, theme, concept, and so on, and gets cross-case and higher level analytic energy flowing.

Usually, a Pattern Code does not get discounted, but rather gets *qualified*: The conditions under which it holds are specified. For example, the RULE of "no 'shop talk' in the lounge" can be bent in cases of conflict, crisis, or socializing of new members. This clarification provides more precise parameters for the pattern and strengthens its validity.

If a general Pattern Code (such as RULES) is used a good deal, it is helpful to create Subcodes that explain the content and enable easy retrieval:

RULES-INDIVIDUAL: Rules about individual participant behavior

RULES-PUBLIC: Rules about behavior in public settings

RULES-WORK: Rules that specify how formal work tasks are to be carried out

Finally, Pattern Codes get checked out in the next wave of data collection. This is largely an inferential process. The analyst tries out the code on a new participant or during an observation in a similar setting; engages in **if-then** tactics (if the pattern holds, other things will happen or won't happen); or checks out a **rival explanation**.

(The **boldface** terms refer to specific tactics of drawing and verifying conclusions, which are discussed in detail in Chapter 11. We use this convention as a way of pointing to tactics as they occur in

later chapters.) See Saldaña (2016) for additional Second Cycle coding methods, particularly those designed for grounded theory and longitudinal studies.

FROM CODES TO PATTERNS

Your initial or First Cycle coding of data generates an array of individual codes associated with their respective data units. Let's take a look at a fictional and extended example of how First Cycle codes transform into Second Cycle Pattern Codes and then get inserted into three display forms: matrices, networks, and graphics.

A selected series of codes related to the first month of withdrawal symptoms described by a participant voluntarily participating in a smoking cessation treatment program, in random order and with their First Cycle code types indicated, are as follows:

1. ANXIETY [Emotion Code]

2. NERVOUSNESS [Emotion Code]

3. "HURT SOMEONE BAD" [In Vivo Code/Emotion Code]

4. RESTLESSNESS [Emotion Code]

5. DEEP BREATHING [Process Code]

6. THROAT BURNING [Process Code]

7. "FELT LIKE CRYING" [In Vivo Code/Emotion Code/Process Code]

8. ANGRY [Emotion Code]

9. "EATING A LOT MORE" [In Vivo Code/Process Code]

10. WANDERING AROUND [Process Code]

11. HABITUAL MOVEMENTS [Descriptive Code]

12. MEMORIES OF SMOKING [Descriptive Code]

13. SMELLING NEW THINGS [Process Code]

There are several ways to approach the categorizing or patterning of these 13 codes. One possible way is to pattern them by code type, such as the following:

- **EMOTIONS** (ANXIETY, NERVOUSNESS, "HURT SOMEONE BAD," RESTLESSNESS, "FELT LIKE CRYING," ANGRY)

- **PROCESSES** (DEEP BREATHING, THROAT BURNING, "FELT LIKE CRYING," "EATING A LOT MORE," WANDERING AROUND, SMELLING NEW THINGS)

- **DESCRIPTORS** (HABITUAL MOVEMENTS, MEMORIES OF SMOKING)

Since negative and strong emotions seem to play a critical role in withdrawal symptoms from smoking, EMOTIONS as a Pattern Code choice makes sense. One can even enhance the code further with the adjective NEGATIVE EMOTIONS.

The PROCESSES and DESCRIPTORS labels, however, seem to lack the "oomph" needed for a Pattern Code. Recall that Pattern Codes usually consist of four, often interrelated, summarizers: categories or themes, causes or explanations, relationships among people, and concepts or theoretical constructs. There are several ways of recategorizing the remaining codes, first by reassembling them into particular clusters because they seem to go together. The analyst proposes the following:

- **Cluster 1**: DEEP BREATHING, THROAT BURNING, "EATING A LOT MORE," SMELLING NEW THINGS
- **Cluster 2**: WANDERING AROUND, HABITUAL MOVEMENTS
- **Cluster 3**: "FELT LIKE CRYING," MEMORIES OF SMOKING

First, what do the four codes in Cluster 1 have in common? They seem to be all upper body functions: respiratory, sensory, and digestive. The analyst reflects on what the four codes have in common; they seem to have a PHYSICAL CHANGES theme that unifies them, and thus get that Pattern Code assigned to them.

The codes of Cluster 2 (WANDERING AROUND, HABITUAL MOVEMENTS) seem to evoke a metaphoric RESTLESS JOURNEY of some sort. Cluster 3's codes ("FELT LIKE CRYING," MEMORIES OF SMOKING) suggest a conceptual Pattern Code of REGRETFUL LOSS. Where did the Pattern Code labels of RESTLESS JOURNEY and REGRETFUL LOSS come from? They came from researcher reflection on what their constituent codes seemed to have in common.

Notice that these four Pattern Codes—NEGATIVE EMOTIONS, PHYSICAL CHANGES, RESTLESS JOURNEY, and REGRETFUL LOSS—are one person's analytic proposals. Other researchers reflecting on and clustering the First Cycle codes might have developed different Pattern Codes altogether. Thus, an important principle to note here is that Pattern Coding—indeed, *all* coding—is not a precise science; it's primarily an interpretive act.

The researcher can now use these four Pattern Codes in various ways, according to the needs of the study. Basic narrative description is one approach; visual displays such as matrices, networks, and graphics are another primary way of analyzing data for fresh perspectives.

Narrative Description

The researcher can compose a section that identifies and elaborates on the Pattern Code, weaving its component First Cycle codes into the narrative and supporting it with field note data:

> Smoking withdrawal symptoms during Month 1 include a *restless journey* for the individual: "I found myself just wandering around the house, just walking from room to room because I couldn't smoke, so I didn't know what to do with myself." The ex-smoker also continues to replicate habitual movements related to smoking, such as reaching for a cigarette pack in a shirt pocket, or leaving an indoor office to go outside to smoke. These physical actions interrelate with, and may even be caused by, several of the *negative emotions* induced by nicotine withdrawal: anxiety, nervousness, and restlessness.

In this case, the storyline function of narrative enables the researcher to outline the plot points of human action and how participants (or "characters") changed throughout the course of the study. Prosaic representation and presentation of our findings are essential ways to communicate to readers how the social action we witnessed and synthesized unfolded and flowed through time. But matrices, networks, and graphics are other ways of representing and presenting those observations.

Matrix Display

Matrix displays will be discussed more fully in the next six chapters, but they are briefly described here for illustrative purposes. Matrix displays chart or table the data—including codes—for analytic purposes. They organize the vast array of condensed material into an at-a-glance format for reflection, verification, conclusion drawing, and other analytic acts.

| DISPLAY 4.3 ⬡ Matrix of Smoking Cessation Patterns at Months 1 and 6		
Initiating Smoking Cessation Patterns	**Month 1**	**Month 6**
NEGATIVE EMOTIONS	Anxious, nervous, angry, aggressive	Occasionally anxious
PHYSICAL CHANGES	Gained 5 pounds, felt "burning" sensation in throat and lungs	On weight loss program after gaining 20 pounds, heightened sense of smell
RESTLESS JOURNEY	Wandering and habitual movements	Habitual movements
REGRETFUL LOSS	"Felt like crying," hyper-conscious of cessation	Nostalgic for smoking, "hangs around" smokers

Suppose that the smoking cessation study was interested in how withdrawal symptoms change across *time*. Display 4.3 charts a participant's data at 1 month and 6 months after quitting. The Pattern Codes are placed in one column, and related First Cycle codes or other data summarizers are placed in the respective columns. A simple matrix such as this enables you—and a reader of your report—to take in the salient findings of your analysis. For example, in the NEGATIVE EMOTIONS row, you can see that such phenomena decrease across a 6-month period, but *anxiety* is still present, albeit in milder form. Each cell of this matrix does not have to include the kitchen sink of withdrawal symptoms, only some of the most relevant exemplars from coding and analysis.

Network Display

This particular withdrawal symptom example describes a *process*, and thus a network of how things act or transform across time (or other aspects such as relationship dynamics or organizational patterns) can be mapped (see Display 4.4). The codes in matrix cells now become possible labels for bins. Lines and arrows indicate connections and flows between the clusters of action they represent.

The analyst has shown how the Pattern Code REGRETFUL LOSS and its constituent codes of "FELT LIKE CRYING" and MEMORIES OF SMOKING have changed from Month 1 through Month 6. Follow-up interviews with the participant suggested that the impulse to cry transformed through time into mild anxiety, while the deeply embedded memories of years of smoking changed into nostalgic reflection on past habits. An interesting track, however, appeared in the interview with the participant 6 months after he quit smoking:

DISPLAY 4.4 ⬡   A Network Model of Smoking Cessation Loss Transformation

Month 1 ⟶ Month 6
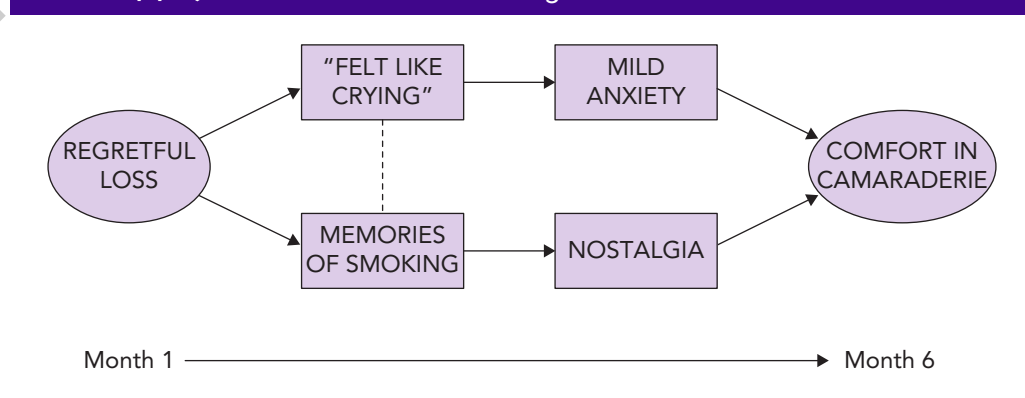
It's still hard, but I find myself hanging around smokers on campus whenever I can, just to smell the cigarette smoke as I smoke on my electronic cigarette. It's comforting just to hang around with smokers even though I don't smoke any more. I still feel like I'm connected to smoking in some way. And I can talk to them about my new habit and stuff. We're still partners in crime.

This interview excerpt, combined with other related coded units of data, led the analyst to compose the evolutionary Month 6 Pattern Code COMFORT IN CAMARADERIE. The participant's ever-present mild anxiety becomes alleviated when he "hangs with" current smokers; his nostalgia for a former habit can be fed by physically placing himself among those who currently smoke. But don't let the elegance and symmetry of the Display 4.4 network fool you into thinking that social life is always linear, balanced, and smooth flowing, and that it can be reduced to a few key variables.

Graphic Display

A graphic display utilizes artistic principles to convey what words alone cannot. The researcher literally *renders* an account through line, shape, mass, color, and other design elements to evoke within its viewers an impressionistic yet research-based summary of the research story. Pie charts and bar graphs are some of the most well-known forms of graphic quantitative data display. But qualitative researchers have at their disposal more innovative forms and formats (discussed further in Chapter 5).

For the smoking cessation data, imagine if we asked the participant himself to *draw* what his emotional states were like at Month 1 and later at Month 6. This form of arts-based inquiry, shown in Display 4.5, displays an aggressive and heavy use of diagonal lines (as smoke) emerging from a broken cigarette in Month 1, while Month 6 displays a snuffed cigarette butt sticking out of an ashtray with a small curve of wistful smoke emanating from it. Interpreting the drawing, in consultation with the participant, the smoker no longer smokes at Month 6, yet the images of a cigarette and smoke are still present. He no longer engages in the habit, but the memories of it still remain in his mind. In contrast with Month 1, the anxiety is lessened; nostalgia has taken its place. "MONTH 1" is written in/with an angry hand; "MONTH 6" appears more controlled and relaxed.

DISPLAY 4.5 ● Arts-Based Graphic Representation of Smoking Cessation at Months 1 and 6
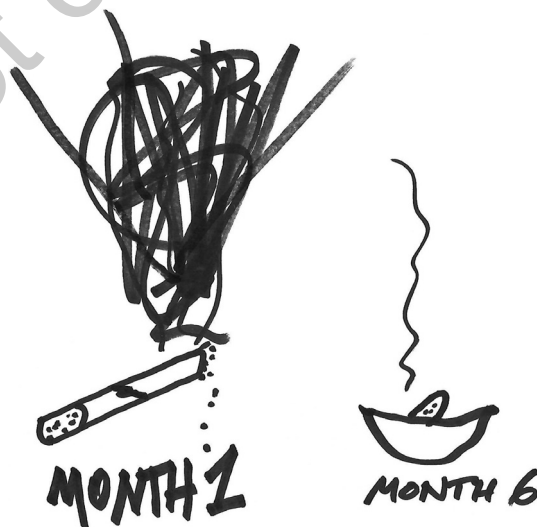
These simple examples of representing smoking cessation were intended to illustrate how Pattern Codes can become grist for narratives, matrices, networks, and graphics, to be more fully explicated in Chapters 5 through 10.

Coding Advice

Coding is not just something you do to "get the data ready" for analysis; rather, as we have said several times, it is a form of early and continuing analysis that drives ongoing data collection. It typically leads to a reshaping of your perspective and instrumentation for the next round.

Every project needs a systematic way to store coded field data and a way to retrieve them easily during analysis. Three-ring notebooks, file folders, half-sheets of paper, index cards, sticky notes, and summaries on poster-size paper taped to a wall are "old school" but time-honored methods for qualitative data analysis. Yet, as we note, good computer software is far ahead of them when it comes to data organization and management.

Perhaps the more important point is this: The ultimate power of field research lies in the researcher's emerging map of what is happening and why. So, any method that will force more differentiation and integration of that map, while remaining flexible, is a good idea. Coding, working through iterative cycles of induction and deduction to power the analysis, can accomplish these goals.

But coding can tire you; it often feels longer than it really is. So, it helps to intersperse coding with jottings and analytic memos.

JOTTINGS

Think of a jotting (Emerson, Fretz, & Shaw, 2011) as an "analytic sticky note"—a piece of writing that could literally fit onto the space of a small square piece of paper. Adobe's PDF document reader has this function; Microsoft Word's "New Comment" feature is an equivalent. CAQDAS programs enable the user to insert "annotations" or "comments" that can be attached to particular units of data. If you're working on hard copy, you could use actual sticky notes, but these can get easily detached if you're not careful. Handwriting notes in the margins will suffice and, in a text file, a separate paragraph (in a different font or color to distinguish it from the rest of the data) will serve for jottings.

So, what is a jotting and what goes into it? A jotting holds the researcher's fleeting and emergent reflections and commentary on issues that emerge during field work and especially data analysis. As you work on a project, reflections of several sorts typically swim into awareness. For example, consider the following:

- Subtextual inferences on the meaning of what a key participant was "really" saying during an exchange that seemed somehow important

- Personal reactions to some participants' remarks or actions

- What the relationship with participants feels like

- Doubts about the quality of some of the data

- Second thoughts about some of the interview questions and observation protocols

- A mental note to pursue an issue further in the next contact

- Cross-reference to material in another part of the data set

- Elaboration or clarification of a prior incident or event that now seems of possible significance

When something like any of these examples arises, it's useful to jot your mental note directly into field notes or somewhere else in the data corpus. It may or may not be fodder for a deeper analytic memo (discussed later), but at least it's in print. One convention is to distinguish the remark with italics to signal that it is of a different order from the data it comments on. Here's a field note example with a jotting:

> The administrative assistant speaks in a sincere voice over the phone: "Well, thank you so much for your help, I really appreciate it. Good-bye." Then she slams the handset into the carriage. *I find it almost amusing to hear the juxtaposition of a "professionally nice" voice followed by a hard, hang-up slam of the phone. She's probably masking a lot of on-the-job tension or frustration.*

Remarks such as these add substantial meaning to the analysis and perhaps even the write-up.

Jottings can strengthen coding by pointing to deeper or underlying issues that deserve analytic attention. Coding, as we have noted, can become tedious if you treat yourself as a sort of machine scanning the pages methodically, picking out small segments of data and assigning labels to them. The sensation of being bored is usually a signal that you have ceased to think. One way of retaining mindfulness in coding is occasional jotting (see Display 4.6 for jottings on a Word document through its New Comment function).
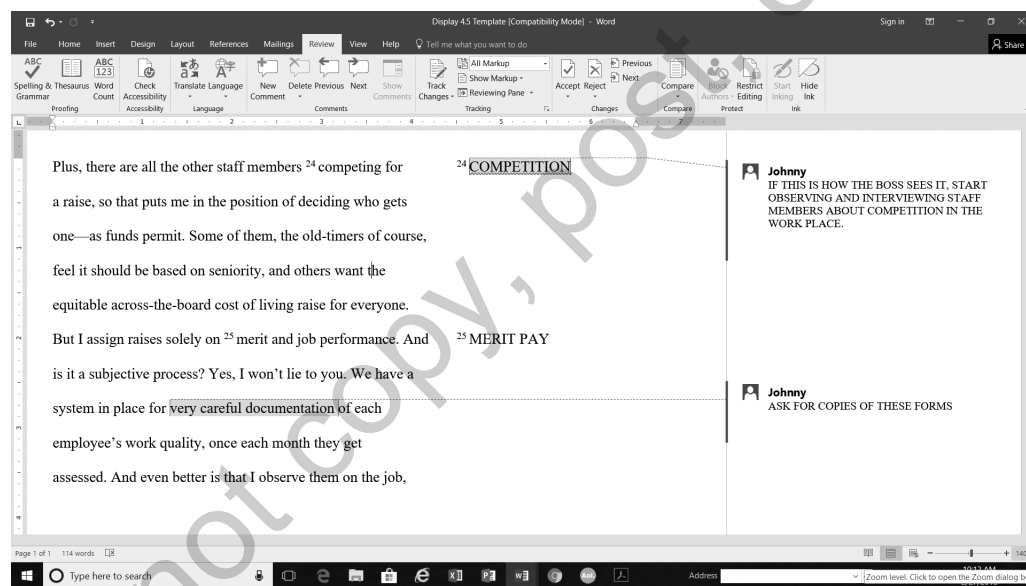
As coding proceeds, if you are alert about what you are doing, ideas and reactions to the meaning of what you are seeing will well up steadily. These ideas are important; they suggest new interpretations, leads, and connections with other parts of the data; and they usually point toward questions and issues to look into during the next wave of data collection and to ways of elaborating some of these ideas. These marginal remarks also point to important issues that a given code may be missing or blurring, suggesting revisions to the coding scheme.

Jottings in the form of reflective remarks can be added while you are writing or expanding on raw field notes. You are simultaneously aware of events in the site and of your own feelings, reactions, insights, and interpretations. Bogdan and Biklen (2007) call these sections "Observer's Comments" or, in field notes, "OCs." These can be separated from the main field notes through a different font style and/or indent:

The committee chair recommended that they take a ten-minute break, but Carla recommended that they continue. "We've got just three more evaluations to get through, we can do that in half an hour." The chair noted, "Well, one of them's going to take a while to get through," so it would be best to come back to the task refreshed and alert.

OC: *The chair gave his reason with a matter-of-fact tone of voice, but one could sense that there were a lot of sticky issues to get through with the one new candidate. Carla seems oblivious to the political infighting I've observed at other meetings, and I feel the rest of the committee wants her to stay that way.*

Marginal notes can be perceived as important "bread crumbs" that are dropped to the ground for later collection by the analyst for expansion through memoing, discussed next.

ANALYTIC MEMOING

Description and Rationale

An analytic memo is a brief or extended narrative that documents the researcher's reflections and thinking processes about the data. These are not just descriptive summaries of data but attempts to synthesize them into higher level analytic meanings. They are first draft self-reports, of sorts, about the study's phenomena and serve as the basis for more expanded and final reports.

Memos are typically a rapid way of capturing thoughts that occur throughout data collection, data condensation, data display, conclusion drawing, conclusion verification, and final reporting. Later in the study, however, memos can be more elaborate, especially when they piece together several strands of the data or look across multiple measures of a concept.

Saldaña (2016) notes that analytic memos can be developed along the following topics:

- A descriptive summary of the data

- How you personally relate to the participants and/or the phenomenon

- The participants' actions, reactions, and interactions

- The participants' routines, rituals, rules, roles, and relationships

- What you find intriguing, surprising, or disturbing (Sunstein & Chiseri-Strater, 2012)

- Your code choices and their operational definitions

- Emergent patterns, categories, themes, concepts, and assertions

- The possible networks and processes (links, connections, overlaps, flows) among the codes, patterns, categories, themes, concepts, and assertions

- An emergent or related existent theory

- Any problems with the study

- Any personal or ethical dilemmas with the study

- Future directions for the study

- The analytic memos generated thus far (metamemos)

- Tentative answers to your study's research questions

- Passage drafts for the final report for the study

Analytic memos don't just report data; they tie together different pieces of data into a recognizable cluster, often to show that those data are instances of a general concept. Analytic memos can also go well beyond codes and their relationships to any aspect of the study—personal, methodological, and substantive. Memo writing often provides sharp, sunlit moments of clarity or insight—little conceptual epiphanies. They are one of the most useful and powerful sense-making tools at hand.

Examples

Analytic memos should be dated for reference to the analytic history and progress of your study, titled with its memo type (e.g., CODE DEFINITION, ASSERTION, THEORY, ETHICS) and subtitled with its more specific content (e.g., WHAT CASES HAVE IN COMMON, PARTICI-PANT LEERINESS, AGENDA FOR NEXT SITE VISIT). Most CAQDAS programs can create and maintain memos (also called "comments" or "annotations" in selected programs), but they can also be kept as a running diary of sorts in a separate research journal. It is recommended that analytic memos not be embedded within field notes, transcripts, or other data but be kept as a separate document.

Here are a few examples of analytic memos from McCammon et al.'s (2012) "Lifelong Impact" study, which surveyed adults about their high school arts education experiences. This first memo documents a pattern observed by the analyst after the codes had been arranged to contrast the younger respondents with older ones:

January 17, 2011

PATTERN: INTRINSIC AND EXTRINSIC

One of the most striking contrasts between survey respondents who graduated in the 2000s and in the 1950s–1970s is what they seem to value about the experiences. More recent graduates wrote about those intangible, intrinsic outcomes such as "camaraderie," "self-discovery," and identity, while the older generations put more stock into awards, specific roles they played, and what they've accomplished over their life course. I wonder if just being recently graduated from high school means that the memories are fresher about those internal experiences, and so it's going to be more in your head. As someone who's from the older generation, I know that I myself put a lot of stock into my own legacy, those tangible things that are evidence of what I've accomplished.

 Ironically, I would have thought the older generations would have been more reflective and internal about those memories, more nostalgic, while the younger "me" generation would have valued awards, letter jackets, etc. Maybe it has something to do with human development—when you're in your late teens and early twenties, you're still trying to figure out "Who am I?" So, you're still looking within and exploring what's really important to you.

Below is an example of how a rough analytic memo eventually transformed into a finished narrative for a final report. First, the memo:

November 2, 2010

METAMEMO: ANTHROPOLOGICAL METAPHORS

It's interesting to see how this study has ties to the social sciences, and how the psychological, sociological, and anthropological disciplines have lenses to offer to the analysis. As for the anthropological, I was struck by a recent reading that used the phrase "the mastery of sacred

texts" as a condition for growing up and becoming part of the culture. In theatre and speech, the mastery of sacred texts is memorizing the script—becoming "one" with the play and taking deep ownership of a character. Theatre and religion have been long explored for their parallels, but I don't think the "rite of passage" theme through performance has been tackled (but check Victor Turner's writings on this just to be sure).

After reviewing the memos to date on major categories, labels such as "community," "tribe," and "family" appear frequently (though theatre people themselves are more likely to use the term "ensemble"—occupational term, I guess). Even better is when respondents told stories about feeling "lost" in performance—a kind of journey taken to achieve spiritual knowledge. The journeys of these participants are both internal and actual (out-of-town speech tournaments, especially)—getting lost to find yourself, leaving home to compete and either win or lose, but coming back stronger than when you left.

And now, these paragraphs from the unpublished technical report illustrate how the ideas from the memo were adapted and woven into the analytic portions of the text:

From social science perspectives (Lancy, Bock, & Gaskins, 2010), there is public expectation and prejudice, if not stigma, toward those who participate in theatre. Adolescent outcasts find their niche among a community tribe of like-minded kin. In these demarcated spaces of classrooms and performance venues, there are operant local school and national cultures of high school educational theatre programs. The adolescent cultural member assumes and adopts the ethos—the values, attitudes, and beliefs—of the social environment in which he/she participates, but with the prerequisite that the young person feels a sense of belonging in that culture. Cognitive maps for survival and safety, emotional and moral socialization, plus individual personality formation occur in these safe spaces through observation, interaction, and challenge. The rote learning and memorization of play scripts and speeches is comparable to the mastery of "sacred texts," valued as "acts of piety, discipline, personal transformation, and cultural preservation" (p. 212). These literal and community narratives contribute to identity, belonging, and expression.

The inherent demands of theatre and speech accelerate adult preparedness. There is high risk for high status. Achievement through awards, placements, and competitions harkens back to initiation rituals and rites of passage to progress toward a higher level of adulthood. Travel to another world, such as the spiritual one of performance and the spatial one of an out of town speech tournament, is comparable to the classic hero's journey in which trial must precede triumph in order to return to the tribe stronger than before. (McCammon & Saldaña, 2011, p. 103)

Memoing Advice

Here we draw on the recommendations of grounded theory's creators, Barney Glaser and Anselm Strauss, and Strauss's later collaborator, Juliet Corbin. Our advice is an amalgam of their experiences and ours. Also see Saldaña (2016) for an extended chapter on analytic memo writing:

1. When an idea strikes, stop whatever else you are doing and write the memo. Don't worry about prose elegance or even grammar. Include your musings of all sorts, even the fuzzy and foggy ones. Give yourself the freedom to think; don't self-censor.

2. Memoing should begin as soon as the first field data start coming in and usually should continue right up to production of the final report. Just as codes should stabilize

reasonably well by one-half or two-thirds of the way through data collection, the ideas raised in memos usually will start settling down then or shortly afterward, as the analyst approaches what grounded theorists call "saturation" (no significantly new explanations for data). Memoing also contributes strongly to the development/revision of the coding system.

3. Keep memos sortable; caption them by basic content. Like coded data, memos can be stored and retrieved using a wide variety of methods.

4. Once again, memos are about *ideas*. Simply summarizing or recounting data examples is not enough.

ON VISUAL DATA

There is much now in qualitative inquiry about the analysis of visual materials, especially since accessible and ubiquitous digital tools enable researchers to document field work with ease, and the availability and amount of Internet resources proliferate daily. Good ethnographers have always documented the visual elements of social life in one way or another. It's just that the media we have nowadays permit us to archive the visual *as* images rather than just through descriptive and evocative writing. And when both text and image work together in integrated concert, as we promote through data display, the result is a more coherent whole (Ledin & Machin, 2018, p. 29).

Analyzing the visual has its own unique repertoire of methods, and we do not have the space to outline them here in detail. We do, however, advocate that interpretation of what we see as still visual documentation—in a magazine, on a website, as a digital photograph, and so on—is more of a holistic, interpretive venture rather than a systematic one. Analytic memoing of your impressions about the frozen, captured image is a more appropriate form of exploration than detailed breakdowns of components such as color, angle, and composition.

But moving images and lengthier passages of television, film, YouTube clips, and other digital streams are more complex and might rely on more traditional content analytic methods such as counts and categories for nuanced analysis. We muse, however, that the analysis of moving images is not all that different from the observation and analysis of real life in real time in real field settings. One advantage of video is that researchers have the opportunity to replay slices of life repeatedly to discern more nuance and patterns of human action such as gaze, touch, gesture, posture, and spatial positioning (Rapley, 2018, p. 69).

Paradoxically, "a picture is worth a thousand words" must contend with "images don't speak for themselves." You as the researcher must interpret the visual and determine whether the task merits analytic methods and strategies not applicable to language-based data. To us, the visual has *always* been a vital part of fieldwork investigation. It is simply the forms and formats—the representation and presentation—of visual data that have evolved over these decades.

As an example, ethnographer and visual artist Paul J. Mack photographed street scenes from urban and inner-city areas. Display 4.7 exhibits a digital photo taken of a residential area situated in the west side of Chicago. Saldaña and Omasta (2018, Chapter 3) offer several guidelines for the analysis of visual materials, ranging from inference-making of the values, attitudes, and beliefs suggested by an image, to a breakdown of its manifest and latent contents, to an analysis of an image's symbols and evocation.

DISPLAY 4.7 ● A Residential Exterior From the West Side of Chicago, 2018

Source: © 2018, Paul J. Mack; original in color.

A content analysis of Display 4.7 might first begin with a bullet-pointed list of specific and related elements:

- A recycling trash container filled with gardening and lawn care tools

- Plywood, warped boards stacked along a wooden staircase guard railing in slight disrepair

- A child's red wagon with a missing wheel and handle, propped on its side next to a large trash can (which has the missing wagon wheel and presumably the handle on top)

- A rusty gutter and drain pipe (which leads to a newer drain pipe section at ground level)

- Cracks, holes, missing sections, discoloration, occasional vegetation, minimal patched repairs on the building's exterior concrete surface

- Older air conditioning window unit and new DIRECTV satellite dish protruding from the building

- Utility connections (cable, satellite, telephone, electricity) visible on building exterior

- Left side residence's doors and window frames in varied stages of decay; right side residence's window frames not as decayed

- Porch/landing area of one residence contains a barbeque, a rectangular folding table, one folding chair, two green metalwork chairs, two plastic buckets, one trash can, one patio umbrella

- Etc.

With this narrative documentation of the visual, the researcher now expands on the image through *inferences*, *interpretations*, and *implications* of the image through analytic memoing:

August 18, 2018

INTRIGUING: ADDITIVE EVOLUTION

I am struck by the age and weather-beaten decay of the structure and its miscellaneous debris collected in its small yards, juxtaposed with three "modern" conveniences: a barbeque pit, an air conditioning window unit, and a DIRECTV satellite dish. A late-19th-/early-20th-century dwelling is occasionally punctuated with 20th- and 21st-century technology and wiring. Old-world electric/gas lamps are juxtaposed with modern security lights. The hardness of gray concrete is juxtaposed with natural greenery, forcibly growing where it probably shouldn't.

I sense that the debris in this yard is not collected carelessly and haphazardly out of negligence, but stored purposefully for future needs of some kind: salvageable wood for building materials, miscellaneous metal and plastic pieces for contingency repairs or repurposing. I smile at the use of the blue recycling trash can for the storage of tools

rather than for its intended purpose. Things at this dwelling are not recycled for the good of the environment, but for the good of these residents' immediate and possible future needs.

Without knowing anything about the people who populate these residences, I infer that at least one of them has younger children, and that there is either low-wage employment, freelance work (e.g., landscaping/lawn care), and/or state/federal assistance. It may be a rental property, owned and operated by an exploitative landlord.

Some social artifacts evolve additively rather than renewably or transformatively. Over time, this entire structure seems *pieced* rather than *integrated* with technology as it has been developed. Its debris, like the repurposed recycling bin, consists not of unattended discards but purposely stored materials for future additive piecing as needed.

Analyzing the visual can range from systematic methods for content analysis, discursive readings for critical analysis, inferential readings for interpretative analysis, and so on (see Richardson, Pickus, & Parks, 2017, for an exemplar). There are several diverse approaches that require visual analysis to have its own methods books devoted exclusively to the subject. We will address the visual in later chapters, but we offer only a small slice of what is possible here. Refer to the appendix for excellent resources available for supplemental guidance.

HYPOTHESES, ASSERTIONS, AND PROPOSITIONS

Coding triggers analytic thought, and jotting and memoing capture the thoughts of the analyst "writing out loud," so to say. But as a study proceeds, there is a greater need to formalize the researcher's thinking into a more condensed and coherent set of explanations. One way to do that is to generate *assertions* and *propositions*, statements that reflect the summative findings and conclusions of the study, both of which can be tested as *hypotheses* before formalization. It will make more sense if we discuss the outcomes before we discuss how to develop them.

To us, an *assertion* is a declarative statement of summative synthesis, supported by confirming evidence from the data, and revised when disconfirming evidence or discrepant cases require modification of the assertion (e.g., "The workers at Adco Incorporated were not self-motivated to achieve excellence"). A *proposition* is a statement that puts forth a conditional event—an if-then or why-because proposal that gets to prediction or theory (e.g., "When employees work in a dysfunctional environment, their individual workplace skills may decay from lack of motivation to achieve excellence").

Assertions and propositions are ways of condensing and synthesizing a vast number of individual analytic observations. They are like "bullet points" of major patterns, themes, trends, and findings that you feel you can confidently put forth about your study. These points can range from descriptive, broad brushstroke facts (e.g., "Overall, the children seemed engaged with the new, experimental learning program") to higher level interpretations about the meanings of the study (e.g., "Experimental learning programs can be high-risk ventures for educators already demoralized by low public opinion of their status and efficacy").

As an example, Kell (1990) conducted a multiple-case study of the effects of computers on classroom instruction. At the first analytic meeting, field researchers recorded their case-specific assertions and propositions, keyed to the research questions. The statements then were clustered thematically, and evidence was sifted for each case. In this study, *the propositions took the form of*

emerging hypotheses—that is, statements to be tested by the data. Here are two illustrations from project data charts:

- Teachers' preferences for different software programs are greatly influenced by their theoretical orientations to reading—that is, phonics or whole-language.

- Individualized learning and self-direction, as well as cooperation and peer teaching, are promoted through computer use, and some transfer of these learning styles to other class activities may occur.

The degree of support for the hypothesis/proposition in each case was then rated as "strong," "qualified," "neutral," or "contradictory."

After the next wave of data collection, which attended to missing data, the propositions *qua* [in the role of] hypotheses were revisited. For a matrix with rows showing each teacher at each site, column entries included data that supported the hypothesis/proposition and data that did not. As it turned out, the second hypothesis/proposition (above) was not supported. At the end, the hypotheses/propositions were tested further with other data sources, and cases that did not fit the patterns were reexamined carefully.

Although this illustration describes proposition generation in the later stages of a study, it can be used productively much earlier—even after the first round of site visits. Keep a running list of bullet-pointed assertions and propositions as a study progresses—that is, hypotheses—and revise them as field work continues and evidence appears that confirms and/or disconfirms them. These statements in progress can also be used as a guide for next-step analysis and further data collection. Eventually, organize the bullet points into a sequential outline format and/or narrative that tells the story of your analysis.

Display 4.8 illustrates the process of moving from a hypothesis to an assertion or proposition. If the data support or confirm the hypothesis, the statement moves forward as a trustworthy finding. But if the data disconfirm the hypothesis, the statement is revised until the data support its "truth value." Also, to clarify the sometimes muddied terminology, a dashed arrow is drawn from the proposition bin to the assertion bin. This is because all propositions are also assertions, but not all assertions are propositions.

DISPLAY 4.8 ⬡ The Hypothesis Process
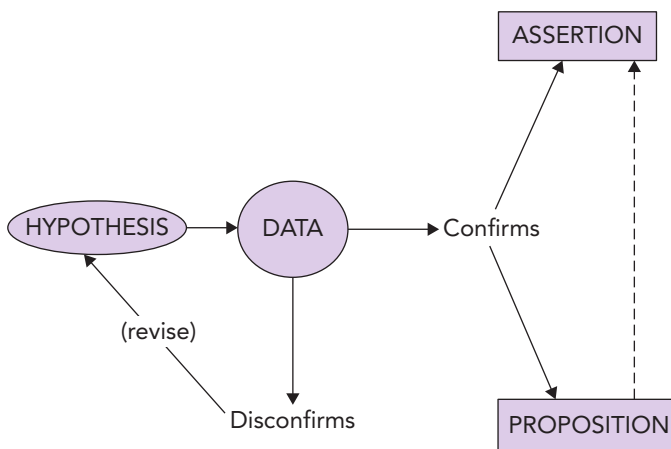
WITHIN-CASE AND CROSS-CASE ANALYSIS

A primary goal of within-case analysis is to describe, understand, and explain what has happened in a single, bounded context—the "case" or site. That is the task of the traditional ethnographic researcher, whose effort is to emerge with a well-grounded sense of local reality, whether the focus is on an individual, a family, a classroom, a school, a tribe, a formal organization, a community, or even a culture as a whole.

Many researchers have leaned toward multiple individual cases (e.g., teachers, alcoholics, middle managers, battered women, taxi drivers). And during the past few decades, there's been a substantial growth in studies of complex settings using multi-case designs, often with mixed methods approaches and multiple research team members (Creswell & Plano-Clark, 2018).

One advantage of studying cross-case or multiple cases is to increase generalizability, reassuring yourself that the events and processes in one well-described setting are not wholly idiosyncratic. At a deeper level, the purpose is to see processes and outcomes across many cases, to understand how they are qualified by local conditions, and thus to develop more sophisticated descriptions and more powerful explanations.

But developing a good cross-case analysis or synthesis is not a simple matter. Alcoholic A turns out to be quite different in personality dynamics from Alcoholic B and can't be easily compared, as Denzin (1993) eloquently shows us. Or suppose, for example, you have developed a good causal network explaining processes in a particular case. If you have a dozen such cases, just adding up separate variables, as in a quantitative survey approach, will destroy the local web of causality and result only in a smoothed-down set of generalizations that may not apply to any specific case in the set—let alone others. Each case must be understood on its own terms, yet we hunger for the understanding that comparative analysis can bring.

Purposes of Cross-Case Analysis

One reason to conduct a cross-case analysis is to enhance *generalizability or transferability to other contexts*. Although it's argued that this goal is sometimes inappropriate for qualitative studies, the question does not go away. We would like to know something about the relevance or applicability of our findings to other similar settings, to transcend the particular in order to understand the general. Multiple cases, adequately sampled (Are they typical? Are they diverse? Are they unusually effective or ineffective?) and analyzed carefully, can help us answer the question, Do these findings apply beyond one specific case?

A second, more fundamental reason for cross-case analysis is to deepen *understanding and explanation*. Multiple cases help the researcher find negative cases to strengthen a theory, built through examination of similarities and differences across cases. That process is much quicker and easier with multiple cases than with a single case. Multiple cases not only pin down the specific conditions under which a finding will occur but also help us form the more general categories of how those conditions may be related.

A Key Distinction: Variables Versus Cases

Our search for helpful cross-case analysis methods will be aided if we clarify two basically different approaches to inquiry.

Ragin (1987) emphasizes that a *case-oriented approach* considers the case as a whole entity, looking at configurations, associations, causes, and effects *within* the case—and only then turns to

comparative analysis of a (usually limited) number of cases. We would look for underlying similarities and constant associations, compare cases with different outcomes, and begin to form more general explanations.

The *variable-oriented approach* is conceptual and theory-centered from the start, casting a wide net over a (usually large) number of cases. The "building blocks" are variables and their interrelationships, rather than cases. So the details of any specific case recede behind the broad patterns found across a wide variety of cases, and little explicit case-to-case comparison is done.

As an example, *a case-oriented approach* would consist of looking at about five different families to observe how particular couples and single parents raise their children. Each parent would be interviewed to get his or her own family background, education, and so on, in addition to particular circumstances, such as ages of all family members, income, work and child care schedules, and so on. These various and richly detailed family biographical profiles would then be compared for analysis.

A *variable-oriented approach* would consist of looking at 50 families representing a diverse sample of structures (two parent, single parent, gay couple, one stepparent and one biological parent, foster parents, etc.) to observe and interview them about a predetermined set of variables included under the main category of "parent–child communication" (e.g., informal dinner conversations, directions and instructions, discipline matters, problem solving, mentorship for "growing up," bedtime stories, tone of voice, etc.).

Each approach has pluses and minuses. Variable-oriented analysis with 50 families is good for finding probabilistic relationships among variables in a large population, but it is poor at handling the complexities of dealing with multiple subsamples; its findings are often very general, even vacuous. Case-oriented analysis with five families is good at finding specific, concrete, historically grounded patterns common to small sets of cases, but its findings often remain particularistic and ill-suited to generalizability.

The implication is not that one or the other approach is better for qualitative data analysis. Rather, the issue is one of making deliberate choices, alternating and/or combining or integrating methods as a study proceeds. The forthcoming methods and display chapters will show how we can focus on either variables or cases or both simultaneously as analytic needs arise.

Strategies for Cross-Case Analysis

How do qualitative researchers proceed when it comes to analyzing data from multiple cases? Here we outline several approaches (and consult the methods profiles in later chapters for more information). The goal here is to show what choices can be made as you approach the question of cross-case analysis.

Displays can help you summarize and compare findings within (and across) cases, but they also can be straitjackets. They may force the data into shapes that are superficially comparable across cases, but you actually may be comparing intrinsically different things on dimensions that turn out to be trivial. As a general rule of thumb, if the formats of within-case displays for a cross-case study are comparable, the work of the cross-case analyst is much easier.

Case-Oriented Strategies

Yin (2018) advocates a *replication* strategy: One case is studied in depth, then successive cases are examined to see whether the pattern found matches that in previous cases. It's also useful to examine cases where the pattern is expected on a theoretical basis to be weaker or absent.

Denzin (2001) approaches the problem through *multiple exemplars*. The issue is not so much "analysis" as interpretive *synthesis*. After deconstructing prior conceptions of a particular phenomenon (e.g., the alcoholic self), you collect multiple instances (cases) and then bracket or isolate the relevant data passages, inspecting them carefully for essential elements or components. The elements are then rebuilt into an ordered whole and put back into the natural social context.

Many researchers approach cross-case comparison by forming *types* or *families*. You inspect cases in a set to see whether they fall into clusters or groups that share certain patterns or configurations. Sometimes, the clusters can be ordered or sorted along some dimensions. For example, Morse and Bottorff (1992) found that 61 lactating mothers fell into four groups: those who could express milk, those who could not, those who perceived it as easy, and those who perceived it as a hassle. The meaning of the experience was fundamentally different for each type of mother.

Researchers usually assume that the cases at hand are more or less comparable, structured in similar ways. *Metasummary*, *metasynthesis*, and *meta-ethnography* (Major & Savin-Baden, 2010; Noblit & Hare, 1988; Sandelowski & Barroso, 2007) make no such assumptions. These approaches systematically *synthesize* interpretations across two or more cases, even if they were conducted by different researchers with different assumptions and different participant types.

Variable-Oriented Strategies

Researchers often look for *themes* that cut across cases. Hlava and Elfers (2014) examined "The Lived Experience of Gratitude" across a diverse range of 51 individuals as a phenomenological study. The overarching concept of gratitude consists of several variables that the co-researchers discerned from interviews with their participants. Hlava and Elfers first explain their data-analytic process of how they progressed from codes to themes:

> The coding cycles used an In Vivo method to preserve verbatim statements of participants, thus allowing the lived experience of gratitude to be expressed in its own terms. The First Cycle of coding employed descriptive methods to identify all statements relevant to sensory and somatic experience. The second iteration of coding used a structural method to label and organize statements by content. . . . Finally, a thematic approach was used to frame the presentation of the lived experience of gratitude. . . . Emphasis was placed on identifying experiences common to all participants, as well as features unique to individuals. (p. 438)

Just some of the themes related to gratitude developed by the co-researchers include

- Gratitude as self–other relatedness
- Sensations in the heart and chest/warmth
- Thankfulness and appreciation
- Comfort, security, acceptance

In phenomenological studies, researchers interview participants in order to compile cross-case meanings. This is not to neglect the primacy or dignity of each individual's story, but to consolidate their experiences to discern their common or most frequent constituent elements (i.e., variables).

Mixed Strategies

It's possible, and usually desirable, to combine or integrate case-oriented and variable-oriented approaches. At a number of points in the forthcoming methods and display chapters, we suggest a

strategy that might be called *stacking comparable cases*. You write up each of a series of cases, using a more or less standard set of variables (with leeway for uniqueness as it emerges). Then, you use matrices and other displays to analyze each case in depth. After each case is well understood (the cross-cutting variables may evolve and change during this process), you stack the case-level displays in a "meta-matrix" (which has columns and subcolumns, rows and subrows), which is then further condensed, permitting systematic comparison.

An example of a mixed strategy is Rosen et al. (2016), who examined legal records from highly adjudicated court cases to find and describe the acute precipitants for physical elder abuse. The *cases* were legal records from 87 successfully prosecuted physical elder abuse cases from 2003 to 2015 in an urban district attorney's files. The *variable* examined was the triggers that led to physical violence (see Display 4.9). The co-researchers employed a cross-case content analysis where each research team member

made determinations of categories for the acute precipitants by organizing and corroborating them with the rest of the team. These categories were discussed by the entire team at several meetings, and we combined categories that were similar and eliminated those without significant support in the transcripts. We developed consensus around 10 categories of acute precipitants using this process. After finalizing categories, the narratives of each case were reviewed to identify whether each of the acute precipitant categories was present. This was also completed with consensus. (Rosen et al., 2016, p. 5)

DISPLAY 4.9 ⬢ Categories of Acute Precipitants of Physical Elder Abuse	
Acute Precipitant	**Description/Examples**
Threat or concern that the victim would involve the authorities	• Victim threatening to call the police and/or reminding abuser about order of protection
Disputes over minor household issues	• Arguments about moving another person's possessions, closing a door loudly, or leaving the television on • Abuser unwilling to contribute to housework despite financial dependence on victim
Confrontation about financial exploitation/demanding money	• Victim refusing or unable to provide support or continue long-time support
Dispute over theft/destruction of property	• Confrontation about victim's missing property • Destruction of victim's property as part of escalating confrontation
Victim attempting to prevent the perpetrator from entering or demanding that he or she leave	• Abuser uses force to attempt to enter the victim's home • Abuser uses keys to enter without the victim's permission and then refuses to leave
Issues with multigenerational child rearing	• Conflicts between grandparents and parents about appropriate child care strategies • Issues of custody and parental access to children that involve grandparents • Arguments between teenage children and a custodial grandparent about household rules and the child's behavior that escalates to assault by the teenager
Conflict about the abuser's substance abuse	• Victim confronting the abuser about ongoing substance use • Victim trying to stop substance use or prevent access to alcohol and/or drugs
Victim threatening or attempting to leave/escape	• Victim trying to leave during an argument, often attempting to de-escalate
Presence during/intervention in ongoing family violence	• Victim present during intimate partner abuse or child abuse and injured by accident • Victim attempting to intervene
Conflict about romantic relationship	• Abuser's suspicion of infidelity • Threat by victim to end the relationship

Source: Rosen et al. (2016), p. 7.

CLOSURE AND TRANSITION

We resonate with qualitative methodologist Robert E. Stake (1995), who muses, "Good research is not about good methods as much as it is about good thinking" (p. 19). Good thinking means to look for and find patterns in the data. Good thinking means to construct substantial categories or themes from an array of codes. Good thinking means to transcend the localness of a particular case to find its generalizability and transferability to other contexts. Research methods are excellent tools, but they are only as good as the craftsperson who uses them.

This chapter provided an overview of analysis fundamentals for and with qualitative data. Yet it does not presume to be the definitive guide; see the appendix for a list of additional resources. These are foundation methods for the more display-oriented strategies that now follow in Part Two.