

9

Sensemaking I: Analyzing, Coding, and Managing Data

Introduction

The previous chapters introduced the principal methods for studying communication practices and performances. Through the use of these methods, we create *data*: textual, aural, and/or visual records of the objects, events, and processes under study. We are soon ready to shift to a new question in the life of our project: What do the data mean? Or to reframe this question in a way that reflects the true nature of the struggle, and potential triumph, in qualitative studies: What sense can I make of the data?

We have reached the stage of analysis and interpretation—a stage that continues through the remainder of our time in the field and extends far into the writing that eventually brings a research project to fruition. We are now spending more time at home or at the office, interacting with data and making conceptual sense of these layers upon layers of discourse and social action. It is an interesting, even exciting, stage of research. But it also arrives with some challenges.

One of our biggest challenges during analysis and interpretation is just coping with all of the data that must be “processed.” The sheer amount of it can be formidable, taxing the patience, stamina, and capacity of even experienced ethnographers. What parts of the dataset will be utilized, and for

what purposes, are issues that consume quite a bit of our thinking at this stage of a study.

A second challenge we face is that our data analysis seldom points in a single, clear direction. There are always forks in the road that demand choices—from deciding how to code a transcript, to purchasing data analysis software, to choosing and shaping exemplars, to selecting the people who will help you validate the study's findings. Alter these or other aspects of the project, and the whole structure of our sense-making can shift (Heider, 1988).

The third challenge is that a study must speak to (or with) at least two “communities” before reaching its ultimate readership. The interpretations we develop must not only be true to the local, contingent meanings of the scene (the community of participants), they must also partake in the scholarly conversations of one or more subfields of communication (the community of one's peers). We could go it alone, without consulting these key communities, but we would do so at considerable risk to the successful completion of the study.

Finally, there is the challenge of acquiring data analysis and interpretation skills. Until recently, novice researchers were faced with a scarcity of good sources concerning coding, inference, and validation. Even today, the methods sections of journal articles can be of little help in decoding the mystery. Authors sometimes tell us that their themes “emerged” after repeated readings of data. But why those themes emerged, and not others, are matters about which readers are often forced to speculate. Some authors take an individualistic, I-did-it-my-way approach that affords scant guidance to anyone wanting to do a similar study. You may also come across authors who just ignore the need to account for methodology, or blithely tuck a brief paragraph about it into an end note.

To be sure, there is some truth to the idea that qualitative data analysis and interpretation are partly a customized and/or intuitive endeavor. Many of us do operate in ways suited to our sense of what looks right and feels right. And most qualitative researchers do experience “ah-ha” moments when they watch the pieces of a puzzle finally falling into place; but flashes of insight typically happen subconsciously, and for that reason, they are difficult to replay or summon at will.

Fortunately, the last couple of decades have seen advances in making the steps involved in qualitative analysis more explicit, trustworthy, and transparent. This in no way negates the role of serendipity in the process. A major strength of qualitative research will always be its wonderful blend of strategic mindfulness and unexpected discovery. Our goal in these two chapters is to help you achieve this blending.

Chapters 9 and 10 develop an integrated package for making sense of qualitative evidence. In this chapter, you will learn that data analysis is the process of reading, labeling, and breaking down (or decontextualizing) raw data and then reconstituting them into categories, patterns, themes, concepts, and propositions. The work of qualitative data analysis is characterized mainly by *inductive* inference (in which one posits an organizing principle from many particulars of discourse and action), but also to some degree by *deductive* inference (in which one tests the strength of that principle by subjecting it to new instances of discourse and action).

Interpretation, on the other hand, is the process of “[making] a construal” (Spiggle, 1994, p. 492). The work of interpretation is mainly characterized by *abductive* inference—in which one creates a new insight from established facts (Agar, 2006; Jensen, 2002; Richardson & Kramer, 2006). In Chapter 10, you will learn some ways in which theory, your coded data, your field-based experience, and your creative imagination join forces in the abductive process of developing *surprising* findings—findings that cannot be easily or directly inferred from sorting instances of talk or action into categories, as in induction, or from testing the fit of data in categories, as in the logic of deduction. You will learn that those precious “ah-ha” moments can be midwifed into existence with the aid of conceptual devices (metaphor, metonymy, irony, etc.), compelling exemplars, and procedures that help validate the truth value of our interpretations.

In actual practice, analysis and interpretation go hand in hand in a larger project of explicating communication. For the sake of this presentation, we depict them as distinct processes, with data analysis starting soon after we enter the field, and interpretation coming along later and overlapping with the final write-up of the study.

Qualitative Data Analysis: An Overview _____

During data analysis, the qualitative researcher hopes to make progress on three fronts: *data management*, *data reduction*, and *conceptual development*. Gaining some control over data that tend to grow rapidly is the goal of *data management*. Without techniques for coding, sorting, and retrieval, the job of finding our way around a mass of data would be a forbidding (and probably discouraging) prospect. Traditionally, data management tasks have been executed through the manual use of tools like pens and scissors and sticky notes, in tandem with general-purpose computer applications like word processors and database management programs. The development of specialized apps—computer-assisted qualitative data analysis software (CAQDAS)—offer us greater control, more options, and savings of time for

the tasks of organizing our data-text records, retrieving text segments, coding and categorizing, and so on. These techniques for managing data will be described later in the chapter.

The researcher will also recognize at some point that not all of the material gathered in the field will be used. Typically, we find that some data are truly critical for making and validating research claims, while other data are used to fashion exemplars that convey these claims to readers (see Chapter 10), and still other data help us grasp the contextual “big picture” of the cultural scene. The data that remain, which still might be a sizeable proportion of the full dataset, are not used at all. *Data reduction*, then, means that the use value of research evidence is prioritized according to emerging schemes of interpretation. The data are effectively “reduced” by employing a coding system that puts you in touch with just those parts of the empirical materials that enable an analysis and interpretation of the findings.

A few cautionary notes regarding data reduction are in order. First, data reduction is an ongoing process, and it isn’t until coding has reached its natural conclusion that we can finally judge which data are useful in interpreting the study’s results, and which data aren’t. Second, data reduction should not be invoked primarily as a way of saving time spent at research sites—as tempting as that idea may sound. Rather, as we’ll discuss in the section, “Leaving the Field,” your disengagement from field sites is ideally based on a rigorous assessment of the soundness of your analysis for addressing your research questions. Third, data reduction does not mean you are literally discarding data; you never know when the chaff (your unused data) from a project might become wheat in another context. Since a research project might spawn a variety of products for different purposes (journal articles, conference papers, evaluation reports, etc.), it is best to think of data reduction as a purpose-specific activity. Permanently deleting any part of your dataset is an action that you might come to regret.

Finally, data analysis serves the goal of *conceptual development*. While data management and data reduction are necessary and worthwhile goals of data analysis, achieving those two goals alone will fail to move you closer to your ultimate objective: making sense of the ways that human beings make sense of their communication practices and performances. To make progress on that front, researchers employ methods of conceptualization from coded data. Concepts and themes grow profusely early in a project as a result of using a coding system. Later on, the number of concepts and themes may be pruned back, but the linkages between them grow more dense and elaborate. Gradually, and with growing confidence, you begin to envision the outlines of your study’s results—and where you might be able to make a genuine contribution to a subfield of Communication.

In-Process Writing

Data analysis often begins informally from the moment your fieldnotes, interview transcripts, and material culture or document notes are being created. As you are busy making this descriptive record, you are also reflecting on the significance of these past events and discourses. The passage of time since you conducted an interview or finished an observational session, even if it is just a day or two, is often felt as a “cooling off” period, which helps you put those events in a different perspective. It is also a good idea to return periodically to your archive of data and reread it. These readings help you to see the recorded events with “new eyes,” consider tactical changes when you next enter the field, and gain a fresh outlook on how the project as a whole is taking shape.

Much of this reflective thinking can be channeled into “in-process analytic writing,” especially the forms that Emerson, Fretz, and Shaw (2011) call asides, commentaries, and memos. We will deal first with asides and commentaries.

Asides and Commentaries

Asides are “brief, reflective bits of analytic writing that succinctly clarify, explain, interpret, or raise questions about some specific happening or process described in a fieldnote [or an interview transcript]” (Emerson et al., 2011, p. 80). An aside, therefore, keys in on a part of the data record that is of immediate interest or concern. The content of these asides may involve a very wide range of interests and concerns—from speculations about the motives of the persons you’ve just met; to the efficacy of a method you’ve used; to emotions, thoughts, or (mis)understandings about your interactions with others in the field. Asides are bracketed next to the part of the fieldnote or interview transcript to which they refer.

A *commentary* is “a more elaborate reflection [than an aside], either on some specific event or issue, or on the day’s experiences and fieldnotes” (Emerson et al., 2011, p. 81). The former type of commentary, a *focused commentary*, will be “placed just after the fieldnote account [or interview transcript segment] in a separate paragraph set off with parentheses [or brackets]” (p. 81). As with asides, the issues treated in a focused commentary may range widely—for example, a pattern of striking incidents you’ve witnessed, or the importance that your interview subject seemed to assign to an artifact. A commentary that reflects on the whole day’s experiences and fieldnotes (or transcript)—a *summary commentary*—is usually placed at the end of an interview transcript or set of fieldnotes, and considers the broader issues implicated in what happened that day in the field (p. 81).

You need not worry about whether the format and style of your asides and commentaries conform to proper “scholarly” writing. Because these writings are often a jumping-off point for developing concepts, they will be reworked into more refined forms of expression long before they surface publicly in reports or publications. In-process writing is for your benefit alone. The ability to write just about anything that strikes you as potentially important or intriguing is *granted to you, by you*, in order to liberate your thinking at this early stage of analysis. There will be ample opportunity later to sort the truly important incidents from the less important, the brilliant ideas from the pretenders, and the valid interpretations from the untrustworthy ones.

To illustrate how these in-process writings function, Box 9.1 displays one aside and one commentary from a fieldnote written by Stephen Haggerty, a former graduate student at the University of Kentucky. In his project, Stephen was studying the socially situated tactics of persuasion, as practiced by a type of professional persuader: car salespeople. The scene depicted in the fieldnote is a training session for newly hired salespeople led by Skip, a professional trainer.

In this example, Stephen decided to label all of his in-process writings “O.C.,” for “observer’s comments.” The first O.C. is considered an aside because it uses information he obtained later to clarify aspects of the immediate setting. This in-process writing takes the reader briefly “aside” from the action and indicates something of interest about the value of the training session as well as Stephen’s relationship with Skip and the car dealer.

The second O.C. is a commentary, a *focused commentary*, because it alludes to an idea about the duty of salespeople. As commentaries go, this one is very brief, yet it is evident that Stephen’s thinking is starting to move in a conceptual direction; later, when he codes the data, he will in fact create a category for “personal responsibility.”

Box 9.1 Aside and Commentary in a Fieldnote

Again and again Skip made David say the words, “I want you to know that I am here to get the deal you want.” He said that repetition is the key, that we have to know the script when we get on the lot, and without a script, we weren’t going to sell any cars. Skip talked at length about the notion that delivering an effective question will demand an effective response. He made it crystal clear that you can’t give up, that in order to sell as many cars as Steve [the owner of the dealership] wanted them to, they had to qualify the customer.

[O.C.: I found out later that the main thrust of the training session that day was on just qualifying, so they invited me back on the 16th of March for the final training session. It is

(Continued)

(Continued)

interesting to note here that Skip told me later that he gets paid a lot of money for this and that I was “lucky, because you’re gettin’ this for free, son.”]

Skip started to role-play with another person, and as he did, he again reminded the trainees (as well as myself) that repetition is the key. Another very important phrase to learn—Skip called them “word tracks,” or short statements—is “I understand.” Skip said you want to make sure the customer thinks he is the one who is the owner of all of the ideas in the conversation. At this point I noticed a mirror at the back of the break room with a sign above it: “Would you buy a car from this person?”

[O.C.: The concept of personal responsibility to the customer and to the dealership rang clearly here as it would throughout the course of my brief interaction with Skip and the other eight people in the room.]

Memos

The third type of in-process writing, *memos* convert loosely thought-out ideas and insights into an organized written form, and thus “require a more extended time-out from actively composing fieldnotes” (Emerson et al., 2011, p. 123). They are usually created and managed separately from the files you’ve created for fieldnotes, interview transcripts, and other primary data. An in-process memo functions as a “snapshot” of your analytic thinking at a particular point in the research—ranging from issues about field tactics or the quality of data; to reflections about puzzling cases; to a line of thought that integrates previous asides, commentaries, or memos (Gibbs, 2007). You may want to write these memos on a regular basis, resulting in a succession of snapshots that reveal the changes in your ideas over time. Unlike the other kinds of in-process writing, the memo has a more finished quality about it, and the writer “clearly envisions outside audiences and frames his [or her] thoughts and experiences in ways likely to interest them” (Emerson et al., 2011, p. 123).

In summary, in-process writings represent tentative attempts to come to grips with the current state of your research. It is important to preserve these asides, commentaries, and in-process memos, even as the issues you write about undergo change. Some of what you write may appear naïve or irrelevant in retrospect, while other writings will seem surprisingly prescient and well worth revisiting. Additionally, the entire log of these writings may become useful later as an audit trail, documenting in chronological order how your claims, warrants, and evidence came into being, and how you wrestled with the issues of your positionality and field relations. As we’ll discuss later in this chapter, in-process writings can be

computer coded to facilitate their retrieval. They can also be “source tagged” for tracing them back to the original contexts of their production. Importantly, in-process writings must be clearly marked as such, so that they aren’t mistaken for, or mixed in with, the descriptive texts of field-notes and transcripts.

Categories and Codes

At some point—usually after a rich dataset has begun to build up—you will recognize that certain units of data (terms, phrases, references, bits of social action, etc.—*incidents* of varying kinds) appear to fit together. That is, they seem to fit as a group under a common theme or organizing principle. This recognition often signals the need to begin creating categories. The systematic start of qualitative data analysis usually comes with the development of categories and a coding system.

Categories

Category is a covering term for an array of general phenomena: concepts, constructs, themes, types, and other meaningful “bins” in which to put items that are related to each other. *Categorization* is the process of sorting units of data with respect to properties or features they have in common. “The essence of categorization,” writes Spiggle (1994, p. 493), “is identifying a chunk or unit of data (e.g., a passage of text of any length) as belonging to, representing, or being an example of some more general phenomenon.”

Categories are devised in a number of ways. Researchers sometimes draw upon concepts from the extant theory and research literature, and apply them to data in a deductive, or *etic*, fashion. For example, Sass (2000), in his study of emotional labor in a nonprofit nursing home, reported,

[I] search[ed] my field notes for the types of performances that were indicated in the previous literature: task and personal rituals, and socialities of courtesies and privacy. . . . I found that episodes of courtesies and task rituals were prominent in my notes [but] personal rituals and privacy were not as useful for capturing emotional labor at Mercy. (p. 338)

As this example suggests, the research literature can sensitize you to concepts that *could be* pertinent to the incidents you’ve recorded in your field-notes or transcripts; a close inspection of the data—and a mind open to alternative readings—will help you decide whether to apply a concept from the literature.

Data may also be coded for standard demographic categories (e.g., sex, age, occupation); institutional labels (e.g., profit, loss, curriculum); and other descriptive, “precoded” topics in wide use (LeCompte & Schensul, 1999, pp. 58–62). These are called *low-inference* categories because they require a few, relatively simple rules for coders to apply, and because they typically denote concrete, widely agreed-upon indicators. **Race**, for example, is a low-inference category because of the conventional ways it is indicated in official records and in everyday life—“White/Caucasian,” “Black/African American,” “Asian American,” and so forth.

While many low-inference categories derive from institutional sources and visible signs in face-to-face interaction, you can also code for *topics* in a conversation or other kinds of discourse. When coding for topic, a key question to ask yourself is, *What is this about?* Sometimes, you are the one who determines “what this is about,” by asking a specific interview question such as: “So, tell me who is the chief influencer in your group?” Most participants will respond to this question with the name of a person they nominate from their group—a very low-inference category we might label **Influencer**. The question itself establishes the scope of the topical category for the responses it elicits.

However, a transcript will always yield more potential topics than just the answers to an interviewer’s questions. After all, when given the opportunity, people are apt to expand the range of their responses: via asides, tangents, elaborations. Ultimately the decision as to which of these ancillary topics to code for new categories stems from the researcher’s interests.

Another type of category begins to form when the researcher notices that various “chunks” of data have certain implicit properties (or features) in common. In the first move of this inductive process, you search the data—usually in several readings—for units of text (words, phrases, or sentences) that interrelate in ways that suggest an overarching theme, or category. We call them *high-inference* categories, because they call for more complex coding decisions. Coming up with a high-inference category often calls on the researcher to discern an organizing principle, or commonality, that isn’t explicitly stated, but rather captures the meanings that underlie expressions of belief, knowledge, and emotion.

Returning to the question we considered above—“*So, tell me who is the chief influencer in your group?*”—what if we find these responses:

“Well, it could be anyone.”

“Some days, I’d say Erin and Barry, and other days, Lydia and Joanna. Really, everybody pitches in.”

“I don’t think any one person in the group is a ‘chief influencer.’”

“We influence each other, we’re all influencers.”

Clearly, these responses don't fit in the low-inference category of names of individual influencers. Yet they aren't totally random responses either. So, what do we do with them? As we think about what ties them together as a cohesive grouping, we might create a category called **Egalitarian Influencing**. None of the interviewees spoke the word, *egalitarian*, yet we could be justified in saying this construct exists implicitly in the thinking and/or actions of these participants.

Clearly, this type of category, a high-inference category, requires more creative effort than just checking off instances of a fixed label. Organizational studies scholar Barry A. Turner (1988) explains how, in his own work, categorization goes forward:

When I have accumulated several instances of a given theoretical category—from six to twelve, depending on the topic—I try to write very clear, formal theoretical definitions of the working category label which I have been using on that particular [file] card, aiming to produce a definition which would be self-explanatory to a newcomer to the research team. . . . In the process of specifying in abstract terms exactly what are the limits of the particular social category, what social phenomenon it refers to, and what it is *not*, the “sociological imagination” is stretched. (pp. 109–110)

This also happens to be a good description of how the communication imagination is stretched through the activity of categorization. Although it is desirable to write “very clear, formal theoretical definitions” of these abstract terms, we sometimes have to acknowledge that a category isn't entirely clear-cut, stable, or free of ambiguity. In fact, what may seem at first to be a low-inference category may be much less so after boundary-disrupting incidents are encountered during coding. For example, **Race**—as we noted above—is often considered a category with just a few, simple rules for labeling people. But what if you encounter subjects who claim a mixed-race or a no-race identity? What if a person who *looks* Asian claims to be of a different racial type or racial heritage? In short, it can be a challenging task to create an abstract category that is grounded in, and responsive to, the first-order meanings of the native inhabitants of a scene, especially when these meanings are contested or in flux. We often have to create (and justify) high-inference categories if we want to push beyond a simple, purely descriptive level.

Codes

Codes are the linkages between data and categories. As Charmaz (1983) puts it, “Codes . . . serve as shorthand devices to *label, separate, compile, and organize* data. Codes range from simple, concrete, and topical categories to

more general, abstract conceptual categories for an emerging theory” (p. 111). The first sentence of Charmaz’s statement describes quite accurately what codes are used for. In Charmaz’s second sentence, however, we see a common conflation of category and code. A code is not the same thing as a category. A code is the “shorthand device” that identifies specific data as an element of a category. A category, on the other hand, expresses the theoretical definition for a grouping of elements that have been identified by the use of one or more codes.

For an example of the distinction between code and category, let’s return to the fieldnote in Box 9.1. The sentence beginning, “Skip said you want to make sure the customer thinks . . .,” is coded as an instance of the category, **Personal Responsibility**. The core meaning or organizing principle, of this category is the salesperson’s sense of responsibility to customers, management, etc.

Coding, as we’ll explore in more detail in the section on grounded theory, is the act of deciding “exactly what are the limits of the particular social category, what social phenomenon it refers to, and what it is *not*” (Turner, 1988, p. 110). Usually the code comes before the category. We scan the texts of our data for phenomena that are interesting to us, or that stand out in the scene we’re studying. After we’ve coded the incident, we may be ready to describe the category it represents. Alternately we may wait on describing the category until our coding has “filled it up” with more incidents of the same kind. Box 9.2, adapted from Gibbs (2007), displays some of the phenomena that are typically coded from texts.

Box 9.2. What can be Coded?

1. Specific acts, behaviors—What people do or say.
2. Events—These are usually brief, one-off events or things someone has done. It is not uncommon for the respondent to tell them as a story.
3. Activities—These are of longer duration than acts and often take place in a particular setting and may have several people involved.
4. Strategies, practices or tactics—Activities aimed towards some goal.
5. States—General conditions experienced by people or found in organizations.
6. Meanings—A wide range of phenomena at the core of much qualitative analysis. Meanings and interpretations are important parts of what directs participants’ actions.
 - a. What concepts do participants use to understand their world? What norms, values, rules and mores guide their actions?
 - b. What meaning or significance does it have for participants, how do they construe events, what are their feelings?
 - c. What symbols do people use to understand their situation? What names do they use for objects, events, persons, roles, settings and equipment?

7. Participation—People's involvement or adaptation to a setting.
8. Relationships or interaction—Between people, considered simultaneously.
9. Conditions or constraints—The precursor to or cause of events or actions, things that restrict behavior or actions.
10. Consequences—What happens if . . .
11. Settings—The entire context of the events under study.
12. Reflexive—The researcher's role in the process, how intervention generated the data.

Adapted from Gibbs (2007)

Codes also have a more prosaic purpose. They are indispensable tools for handling the “office” function of data analysis (Gibbs, 2007, pp. 1–2). Codes mark the units of text in fieldnotes, transcripts, documents, and audio-visual materials, which permit researchers to sort, retrieve, link, and display data. Since qualitative projects often generate dozens of pages of material, comprising hundreds of lines of text, it is inevitable that not all of this text will be coded. Uncoded data are not quite the same as data that no longer exist (out of code, out of mind?); however, it is true that once a project has swung into full analysis mode, researchers typically utilize only the data they have coded. As noted earlier, data reduction is a necessary goal of data analysis, and coding serves the purpose of marking the islands, archipelagos, and other land masses of meaningful data from the surrounding sea of raw, uncoded data.

Coding Narratives

Today researchers have many options for choosing a coding method. The list is long and includes such specialized types as emotion coding, values coding, process coding, dramaturgical coding, causation coding, motif coding, hypothesis coding, and protocol coding (Saldana, 2013). There are also coding methods that work well for visual imagery like photographs, video, and graphics (Clarke, 2005; Figueroa, 2008), and group-interaction data such as the results of focus groups (Clayman et al., 2009; Duggleby, 2005). Accordingly, your choice of a code should be based on the type of data you've collected and your objectives for analyzing and interpreting it. Communication scholars are particularly interested in strategies that help them make sense of talk, stories, and other narrative texts. Let's look briefly at a strategy for coding verbal exchanges—developed by the late Communication scholar H. Lloyd Goodall, Jr. (2000, pp. 102–108)—that comes from the cultural hermeneutics tradition.

The communication researcher begins the coding process by identifying a sample of talk with reference to a continuum of conversational forms, from the most rule-governed speech (e.g., phatic communication) to speech that enacts multiple constructions of meaning and identity (e.g., skilled conversation, personal narratives, and dialogue). The next move in this strategy invokes the overarching question of *what is going on* in the verbal exchange. The researcher then deploys a series of questions, using all available knowledge of the scene of study. Each of these questions—*What is the frame or context? What is being said? How is it being spoken? Where are you in this scene?*—includes a number of more specific probes for coding the speech. For example, here are some of the probes used in coding for the question, *How is it being spoken?*:

- What are the *rhythms*, the *vocal tones*, and the *silences* contributing to the overall meanings? Where does the *storyline* come from? From personal history? From cultural myth? How is it gendered?
- What are the *life scripts* being invoked?
- What does it all add up to? What does it *mean*? (Goodall, 2000, p. 107)

Once the results have been coded, the analyst goes on to the next move: Rendering the “*personal meanings* [of the verbal exchange] . . . which is a formal way of suggesting that *you* are isolating the key moments in the exchange and attributing special meaning to them” (p. 108). Among the conceptual devices that Goodall recommends for finding these moments are *rich points* (speech acts—such as jargon, slang, or ironic turns of phrase—in which cultural knowledge is expressed) and *turning points* (talk about critical decisions in the life of an individual, a group, or an organization). The act of explicating these personal meanings goes well beyond coding and categorization and takes us into the territory of interpretation, of which more will be said in Chapter 10.

Grounded Theory

One of the most influential models for analyzing qualitative data is the *grounded theory* approach. More than five decades after its introduction by Barney Glaser and Anselm Strauss (1967), grounded theory “has spread from its original home discipline of sociology to a multitude of disciplines” (Goulding, 2017, p. 61). Among the reasons for the model’s enduring popularity is that it articulates a compelling “logic of discovery”—a way of building theory inductively that “favours analysis over description, fresh categories

over preconceived ideas and extant theories, and systematically focused sequential data collection over large initial samples” Charmaz (2006, p. 187). Not every qualitative researcher subscribes to this privileging of analysis over description. However, grounded theory embodies an outlook on theory development “from the ground up” that is broadly attractive to qualitative researchers, including many in Communication.

Grounded theory also brought a sense of clarity and order to specific components of qualitative research. In particular it opened up what was once considered an impenetrable “black box”—data analysis—and unpacked a formal, iterative process of coding and conceptualization, or what is known as the *constant-comparative* method. This method is probably grounded theory’s greatest impact across the social sciences. Turner (1988) even contends that “the qualitative researcher has no real alternative to pursuing something very close to grounded theory” (p. 112).

The merits of grounded theory have been debated intensively, with its major proponents advocating different versions of the model (Charmaz, Thornberg, & Keane 2018; Goulding, 2017). Grounded theory also has its “sworn enemies . . . [and] has also been used as a piecemeal, pick-and-mix, catchall overarching term to describe and label any form of qualitative research, from in-depth interviews to projective techniques” (Goulding, 2017, p. 61). Our tour of grounded theory will not venture into these controversies and intramural disputes. Rather, we focus here on the general model of grounded theory as it is employed widely in Communication, and how it helps us accomplish important tasks with data. Following this exposition, we will examine two examples that illustrate how grounded theory’s guidelines for coding and categorization are put into practice.

We begin by citing three tenets of grounded theory that are particularly vital for analysts of qualitative data:

1. Emergent theory is “grounded in” the relationships between data and the categories into which they are coded.
2. Categories develop through an ongoing process of comparing units of data with each other (a process known as the *constant-comparative method*).
3. Codes, categories, and category definitions continue to change dynamically while the researcher is still in the field, with new data altering the scope and terms of the analytic framework.

It is possible to adhere to one or two of these tenets and forego the others. However, grounded theory tends to operate optimally as a data analysis model when all three tenets work together. Coding usually starts early, when the researcher’s memories of the events depicted in the data are still fresh.

In this first stage of the model, we code our data into as many categories as seems feasible. As discussed earlier, some of the categories may originate in the research literature. However, the inductive thrust of grounded theory places emphasis on creating categories from firsthand contact with evidence, which in turn is informed by our physical, emotional, and cognitive experiences in the field. This close relationship between data and categories, as expressed in the first tenet, serves to keep theory “grounded” in the social realities and cultural understandings of the scene.

Two kinds of coding—open coding and *in vivo* coding—figure prominently in this early phase. *Open coding* is the initial, unrestricted coding of data (Strauss, 1987, pp. 28–32). In open coding, we go through the texts (fieldnotes, transcripts, documents) line by line and “categorize a chunk of data on the basis of its coherent meaning—its standing on its own—not by an arbitrary designation of grammar” (Spiggle, 1994, p. 493). In fact, it is through the process of open coding that categories are built, named, and have attributes ascribed to them. The major goal of open coding, as the name implies, “is to *open up* the inquiry. Every interpretation at this point is tentative. . . . Whatever is wrong in interpreting those lines and words will eventually be cancelled out through later steps of the inquiry” (Strauss, 1987, p. 29; emphasis in original).

Another major type of coding, *in vivo coding*—the “folk or indigenous terms” (Saldana, 2013, p. 91) used by social actors to characterize their scene, their identities and actions—is conducted at the same time as open coding. This is a valued type of code in grounded theory research because it helps keep categories closely “grounded” in the discourses and localized meanings of a scene. For example, Steven Haggerty’s commentary in Box 9.1 reveals that “word tracks” is a term used in car sales circles, and “I understand” is one instance of these word tracks he heard in the sales training session. So, *word tracks* is an *in vivo* code for finding and making sense of certain kinds of salesperson talk, such as “I understand.” Later in your analysis, *in vivo* coding can point you to particularly interesting parts of the data for constructing exemplars.

Open and *in vivo* coding could potentially go on indefinitely. Indeed, the number of codes and categories often grows very rapidly during the “first cycle” of coding, and it isn’t unusual for a given incident (e.g., a fieldnote observation, a phrase from an interview) to be placed into many different categories. Keeping this process from getting out of control is the researcher’s concern for *comparing each incident of a code to other incidents* to decide which bins (categories) they go into. As these comparisons are being made, you are also busy writing the definitions of the categories.

By this time, you’ll need to create a *codebook*. A codebook is “a tool for the development and evolution of a coding system and is an important

means for documenting the codes and the procedures for applying them” (Weston et al., 2001, p. 395). Its main purposes are to list and define the categories, list the codes used for identifying instances of each category, display examples of text (from fieldnotes and/or interview transcripts) that exemplify the category, and include the location of incidents in the data of the codes for each category. (See LeCompte & Schensul, 1999, for variations on the design of codebooks.)

Importantly, the codebook is used as a teaching and reference tool when multiple people are involved in coding. Through many rounds of independent coding of the data and subsequent discussion of coding decisions, all supervised by the researcher, coders gradually learn to coordinate their activity around a shared sense of how to utilize the codes and categories. Since reliability isn’t typically a benchmark of data quality in qualitative research (see Chapter 10), the degree of concordance between coders, especially as measured by intercoder reliability coefficients, isn’t enforced as strictly as in quantitative content analysis. Indeed, the training of coders in qualitative studies is as much about the practical aspects of how to talk about differences of coding as it is about how to apply the theoretical definition of codes and categories (Hak & Bernst, 1996). According to Barbour (2001),

what is ultimately of value is the content of disagreements and the insights that discussion can provide for refining coding frames. The greatest potential of multiple coding lies in its capacity to furnish alternative interpretations and thereby to act as the “devil’s advocate” . . . in alerting researchers to all competing explanations.

Around the midpoint of the data analysis, we often turn to writing *theoretical memos* that serve to flesh out the thematic meanings of the categories (Charmaz et al., 2018). These are of the same genre as the memos we discussed earlier. However, they differ from their in-process cousins by being focused on the process of categorization and the relationships among categories. You may also want to use theoretical memos to write notes about the social actors, the influence of your own identity and behaviors on the action you’re observing, puzzling or ambiguous findings you’ve encountered, and emerging ideas about the theories that you may use during the interpretation phase of sense-making. Some of these theoretical memos may be so intricately related to the data-texts that you will weave them into the final write-up of your project.

Gradually, you begin to settle on an understanding of each category—what it means, what it refers to, what its properties are, and how it is different from other categories. In other words, the *constant-comparative method* serves to define a category with greater precision, as well as to help differentiate

it from the other categories. As Glaser and Strauss (1967, p. 107) note about this stage of the inquiry, the analyst “starts thinking in terms of the full range of types or continua of the category, its dimensions, the conditions under which it is pronounced or minimized, its major consequences, its relations to other categories, and its other properties.”

You may also notice how full (or empty) of evidence the categories are. If a category is coded for just one incident, it may be possible to merge it with another similar category. (However, it isn't necessarily a problem if a category is represented with just a single incident.) On the other hand, categories with relatively larger amounts of evidence may prompt you to rethink the use of your codes, and whether certain codes are a better fit with other categories. Eventually, by the time you've coded and recoded all of the data, the number of categories will begin to level out, or even decline somewhat.

The next two stages of the grounded theory approach are called *integration* and *dimensionalization*. These processes are concerned with defining and reconfiguring the categories you have created thus far and developing deeper meanings of them. The process of integration starts with axial coding—developing a set of *axial codes* whose purpose is to make connections between categories. An axial code tends to act on a category in several specific ways: “The [*causal*] conditions that give rise to it; the *context* (its specific set of properties) in which it is embedded; the action/ interactional *strategies* by which it is handled, managed, carried out; and the *consequences* of those strategies” (Strauss & Corbin, 1990, p. 97). Axial coding brings previously separate categories (or concepts) together under a new thematic category. Axial codes help to spell out a new organizing principle for the thematic category—that is, what it is that justifies bringing two or more categories together “under one roof.”

Dimensionalization is one of the final steps in the grounded theory process. According to Spiggle (1994), “dimensionalization involves identifying properties of categories and constructs. . . . Once a category has been defined, the analyst may explore its attributes or characteristics along continua or dimensions” (p. 494). When we do a dimensional analysis, we examine each category—again by reference to the coded elements that make up the category—and try to tease out its key variations (dimensions).

Ultimately, the category set becomes “theoretically saturated” (Glaser & Strauss, 1967, p. 110). Newly collected data add little new value to the concepts, and “later modifications are mainly on the order of clarifying the logic, taking out non-relevant properties, integrating details of properties into the major outline of interrelated categories and—most importantly—reduction” (p. 110). This is about as far as we can go in coding, organizing, and “explaining” the data, before we go on to create, evaluate, and enhance interpretations. This next phase of sense-making is explored in Chapter 10.

Coding Examples

This is a good point to pause and consider two examples of the coding ideas we have been discussing. The first one comes from Tom's study of the practices of political advance teams.

Attractions of Advance

Box 9.3 shows a transcribed excerpt from Tom's interview with Nicholas (a pseudonym), a former appointee in the George W. Bush administration with more than 20 years of experience in advance. Here, Nicholas responds to a question Tom has asked about what attracts him so much to advance work that he, a professional man with a family, keeps coming back to this life on the road for one election cycle after another.

There is a lot going on here—much more, it turned out, than just responding to Tom's question. It is not unusual for an interviewee to bring many more topics and ideas into a conversation than the ones the interviewer asked about. But let's just focus on the open codes created by Tom that relate to his question. The category **Attractions of Advance** was created to “house” these codes. In reading the transcript, Tom decided to code the segments “witness to the inner workings of government,” “witness to what happens behind

Box 9.3 Excerpt From an Interview With Nicholas

Nicholas: The thing about advance that attracts most people to do it is that you are witness to the inner workings of government. You are witness to what happens behind the scenes. You have a close interaction with a candidate, or the president for that matter, or the secretary of state. You interact with them. You know, for people who are goal-oriented, when you do advance, you basically start with a blank slate or a blank canvas. And advance is given a lot of leeway to go in and paint that picture. You're given general guidelines to follow. And you can take those guidelines, but you can kind of connect the dots how you want to. You can kind of paint the picture how you want to. So you're given certain rules to follow. You know, you're given this much time, he's going to get here then, and he needs to leave here at this time. And within that he needs to accomplish A, B, and C. But you can kind of paint the picture on how he gets to A, B, and C. So for people who are creative, it allows you to express your creative abilities. And so that's very attractive to a lot of people. You know, advance never pays well. Nobody gets in it for the money. You get in it for the experience. You get in it to be able to manage yourself under very stressful circumstances. In advance, there's no room for failure.

the scenes,” and “a close interaction with a candidate” with the code name, *Close to the Action*. All of these phrases had the quality of being “close in” to politics. (Later, as he coded other transcripts, Tom decided to split this code into two codes named *Close to Candidates* and *Close to Political Backstage*.)

Then, Tom found a cluster of segments that had to do with a code he named *Creativity in Constraint*:

“You basically start with a blank slate or a blank canvas. And advance is given a lot of leeway to go in and paint that picture.”

“You can take those guidelines, but you can kind of connect the dots how you want.”

“Paint the picture how you want to. So you’re given guidelines to follow.”

“He needs to accomplish A, B, and C. But you can kind of paint the picture on how he gets to A, B, and C.”

All of these are colloquial statements, even clichés, but as Tom discovered, they have special meaning for the advance world. Advance staffers are often told to go to a city and build out an event to be held at a specific date in the future. Headquarters may give them certain instructions about how to do this, but otherwise advance people are given a fair amount of freedom to devise an overall plan for the event. Thus the phrases, “blank slate/canvas,” “paint the picture,” and “connect the dots,” convey the notion of a space of creativity functioning within or between a set of constraints. Nicholas’s statements also seemed to suggest—without explicitly saying so—that the advance person is allowed to be creative, without necessarily being encouraged or pushed by others to do so. Is that really the case? What, indeed, is the actual attraction being expressed by Nicholas? Creativity itself, or being given the chance to be creative? Or is it a bit of both? These questions suggest an opportune moment to write a commentary for this passage of the transcript.

Tom also created separate codes, again under the category **Attractions of Advance**, for the phrases, “You get in it for the experience” (code name: *Political Experience*), and “You get in it to be able to manage yourself under very stressful circumstances” (code name: *Testing Oneself*). The two phrases seem to express closely related ideas, but it became clear as the interview went on that what Nicholas meant by saying, “You get in it for the experience,” was different from the experience of testing himself under pressure. Later, Tom wrote about the close kinship between these two codes in a memo:

Part of why they like politics, and keep going back to it, is the chance to prove themselves in the high-stakes atmosphere (and get the accompanying adrenaline rush) of campaigns.

In a memo you can write provisional claims or hypotheses about what is going on “in” a category, or the relationship between categories. This conceptual thinking can provide grist for further work in the field, such as looking for incidents that strengthen or disprove the claim, or in the analytic process, such as creating axial codes that express the nature of thematic categories.

Invisible Disability in Families

A good published example of the grounded theory approach to coding is Canary’s (2008) study, published in the *Journal of Applied Communication Research*, about constructions of identity in families with children who have invisible disabilities. (A person with an invisible disability does not display signs of his or her disability that others can readily recognize. Common invisible disabilities include autism, learning disabilities, speech and learning impairments, mental incapacity, and emotional disturbances.) Canary was mainly interested in exploring “the importance of interactions within families as well as interactions between family members and others outside of the family as family members construct what ability and disability mean for them” (p. 438).

Her research strategy involved interviews with, and observations of, four ethnically diverse families, each with at least one child having an invisible disability. Although she sought to interview all of the families’ members, she was unable to obtain interviews with a small number of individuals due to their age or unavailability at the time of the study. Her resultant sample of interviewees consisted of nine adult and twelve child participants. The adult interviews centered on “descriptions of target children, discussion of disability in the family, accommodations for disability in the home, sibling interactions, and parent interactions with professionals” (Canary, 2008, p. 443). The areas emphasized in the child interviews included “descriptions of siblings, sibling interactions and relationships, preferential treatment, and family dynamics” (p. 443).

Following transcription of the twenty-one interviews (and, in some cases, translation from Spanish to English), Canary engaged in open coding of the transcripts. This coding produced forty-five individual codes focusing on how the participants described themselves, their interactions with other family members (e.g., siblings, parents, children, extended family members), disability and ability, characterization of their families, and interactions with people outside of the family. Canary reanalyzed these coded data “to generate broader axial codes that represented connections between the first-order codes” (Canary, 2008, p. 443). The process of axial coding resulted in six thematic categories. Box 9.4 displays three of these thematic categories

(*identity situated in sibling relationships*, *identity situated in family unit*, and *disability is perplexing*), the codes used to constitute each category, and representative samples of interview discourse. We see revealed in this analysis the major categories that family members rely upon in creating stories about dis/ability. This coding and categorization activity is a prelude to the act of interpreting discourse and social action—the subject of Chapter 10.

Box 9.4 Categories and Codes

Category Name	Example	Codes Included in Category
Identity situated in sibling relationships.	"They'd find the easiest way possible to give her the answer . . . they just put her right there."	Age and birth order Help each other/stick together Conflict inevitable Sibling activities together Sibling conflict behaviors Sex differences/gender roles Little sibling interaction
Identity situated in family unit.	"main thing we probably talk about is respecting and loving each other . . ."	Family activities Family conversation topics Family descriptions Family routines Homework help Family important Respect and love important Parental responses to conflict Learn from parents Parental messages about conflict
Disability is perplexing.	"nothing you can really do other than figure out what the problem is, and nobody can figure that out."	Different ability levels Locus of problem Noticeable difference External influences Trying to figure out Getting/will get better Frustrating Takes time Like everybody else Doesn't affect anybody Not an issue Don't understand problem Intimidating Adults talk privately

Source: Canary, H. E. (2008). Negotiating dis/ability in families: Constructions and contradictions. *Journal of Applied Communication Research*, 36, 437–458.

Leaving the Field

The progress of the data analysis also plays a role in deciding when you should finish collecting data and leave the field site. However, these decisions are often driven by practical factors. For example, major changes in the scene—such as a turnover in leadership, or the departure of the group itself—may signal that it is time to pack up the proverbial tent and move on. Other, more personal reasons for exiting research sites include fatigue, physical and psychological stresses, the depletion of funding, the pressure to publish, or the need to finish the assignment, thesis, or dissertation. Participants often empathize with the researcher in these situations because the need to cope with work-induced stress or the deadlines of other people and organizations is nearly universal.

If you are not forced by practical or personal issues to exit the field prematurely, then criteria of research quality should be the controlling factors. Snow (1980) described three tests of “information sufficiency” by which researchers can gauge when it is time to finish data collection. The first test, *taken-for-grantedness*, means that you have achieved a high degree of *emic* competency within a cultural membership; at this point, you understand the participants’ world so well that you are seldom surprised or baffled by what they say or do. Ask yourself these questions: Could I “pass” as a member of this group (i.e., perform the role of the other convincingly)? Can I predict with a fairly strong degree of certainty how a member would behave in situations like the ones I’ve studied so far? Am I still curious about the people I’ve studied? Am I beginning to take for granted what I initially found fascinating, odd, or paradoxical? If you answer yes to one or more of these questions, then you are probably quite far along in gaining the “insider” knowledge that you sought in embarking upon the study.

We have already referred to the second test, *theoretical saturation*. As Snow (1980) notes, “saturation is signaled by the continued observation of what is already known, and by repetitive field notes” (p. 103). Whereas *taken-for-grantedness* is a test of your ability to grasp (or even perform) native meanings of communicative action, saturation is a test of the analytic categories and explanations you’ve constructed from the data. In evaluating saturation, ask yourself these questions: What would the addition of one more incident, artifact, or interview respondent do to the codes, categories, and memos I’ve already assembled? Would it alter, or add anything new to, an existing category? Would the linkage I’ve posited between two or more concepts become stronger (or weaker) if I continued to observe or interview at this field site? In Chapter 10, we will introduce more formalized procedures, such as analytic induction, that help validate your research claims and thus indicate when it is time to curtail your data collection.

We see an example of both criteria—taken-for-grantedness and theoretical saturation—in Wellman’s (1994) account of the winding down of his 3-year study of a longshoremen’s union:

Impatience followed fatigue. I found myself less sympathetic to stories and exchanges repeated endlessly, and my ability to record them faithfully was decreasing. I knew my job in the field was completed when I discovered the source of my impatience. I had heard the stories, witnessed the exchanges, and observed the events so many times that I knew how they would end when they began. I could predict the process as well as the outcome. . . . My research categories were saturated. To convince myself that saturation was not simply an expedient excuse for fatigue, I tried to actually predict how the process would unfold. When I succeeded, I knew the time to leave had come. (p. 582)

The third test, *heightened confidence*, tells you that “the observations and findings are faithful to the empirical world under study and shed light on preexisting or emergent questions and propositions” (Snow, 1980, p. 104). Heightened confidence essentially refers to the *credibility* of claims or concepts. As you will learn in Chapter 10, member validation (or member check) is a common way of establishing the credibility of our claims, from the standpoint of the participants in the scene you’re studying.

When disengaging from the scene, you should reserve time for activities that can only be done “on the ground,” including touching bases with gatekeepers and other stakeholders, collecting copies of key documents, and making sure you’ve delivered on any promises you made upon entering. For some researchers, a “relational contract” drawn up during the entry phase can make the process of exiting go more smoothly for all concerned.

[A] relational contract covering research relationships, allocation of tasks, and moral obligations with regard to research disengagement can be valuable in charting courses through which field exit may take place. . . . Such contracts often outline who has ownership of data and how the researcher may use the material for scholarly writing after exit. (Michailova et al., 2014, p. 144)

Friendships are often formed that transcend the roles of researcher and human subject, and strong emotions that have been buried or managed during fieldwork—running the gamut from love and affection to sorrow and mistrust—may rise to the surface when you leave. It is certainly possible, and usually not unethical, to maintain personal relationships with one or more of your subjects after their participation in the study is over. You may also want to “debrief” the study participants about what you’ve found, in whatever format you think would be most effective and well-received: a group meeting,

a series of one-on-one meetings, a written memo, a summary on your website. Doing this may help you avoid misunderstandings or hurt feelings after you've published articles from the study, particularly if the findings are unflattering to the community (e.g., Ellis, 1995). Nevertheless, it is vital that you try to leave the field site on good terms, if for no other reason than you—and other researchers—may want to use this site again. Indeed, the practice of revisiting field sites can be a highly informative way of accounting for cultural change in the spirit of reflexive ethnography (Burawoy, 2003).

Tools for Analysis

Today, most qualitative researchers in communication rely on computers and related software to assist with at least some of their data analysis needs. Although general-purpose programs like word processors and database managers serve these needs quite adequately for many investigators, for others it is the software designed specifically for qualitative research that has the most appeal. By incorporating the functions most common to qualitative data analysis—for example, text coding and retrieval and category building—these software packages can save a great deal of time and effort and create a smooth digital transition from the field to the report. Even more intriguing is the idea that the programs afford new ways of thinking about—and representing—the relationships among data.

Most of this section provides a basic introduction to computer-assisted qualitative data analysis software (CAQDAS). First, however, we discuss the analysis of data by hand. Before the current era, all qualitative researchers used physical tools—paper, scissors, index cards, sticky notes, file folders, binders—in physical spaces. Many people still prefer to use these tools and their associated techniques, and knowledge of how they work may help you decide whether to adopt a more digitalized approach to analysis.

Manual Methods

We begin by glimpsing the data-handling methods of a Communication scholar. At the time she wrote this, Alyssa Eckman was a doctoral candidate at the University of Kentucky and facing the challenge of making sense of ethnographic data for her dissertation study of newspaper “advertorial” practices:

My field notes were kept in Microsoft Word files on disk and backed up on two different computer hard drives (home office and work office) because my biggest fear was losing all that I had accumulated. Fortunately I kept my notes in a day-to-day journal, and that linear, organized approach saved me several

months later because they helped bring my memories back to life, back into focus. I also collected documents from meetings, print-outs of relevant e-mails, and even a few napkins-turned-notepads from lunch gatherings with co-workers. This accumulation of materials all landed in my home office, which is a small 10-by-12 foot loft overlooking my living room. . . . I read my notes several times and began to note recurrent terms, actions, comments, and experiences. I will confess to using the pile method of organizing the data around the six themes that I eventually developed. My small loft office proved too small, so one weekend I moved to the living room and surrounded myself with the field notes, interview transcripts, and the rest of the documents collected during the field study. Armed with scissors, a set of multi-colored markers, and post-it notes, I began to cut my notes into pieces. Each piece landed in a color-coded pile, a pile that either represented a recurrent theme or a “question mark” to be addressed later. A couple of the “question mark” collections were eventually re-designated as recurrent themes as supporting data began to grow in those piles.

Several aspects of handling data manually are notable in this passage. First, Alyssa did avail herself of a word-processing program to store, index, retrieve, and print data texts. Very few qualitative researchers, if any, shun computers altogether. Second, the use of physical space opens up interesting opportunities for arranging and viewing the data. The living room floor became a large canvas for Alyssa’s materials, enabling her to look at them from different angles and move back and forth across the piles of notes like an “action painter” à la Jackson Pollock. Third, Alyssa was literally in touch with her data as physical objects, including napkins-turned-notepads. Many researchers find this close, palpable contact with data more satisfying (and easier on the eyes) than gazing for hours at a screen.

The pile-sorting just depicted is one way to go about the manual method of data analysis, but there are others. For example, Atkinson (2009) culled quotations from his transcribed interviews with social movement activists that illustrated

worldviews, uses of alternative media, and performances of resistance. As quotations were identified they were copied and pasted into a Word file so that all of the relevant quotes could be printed and cut out. I then proceeded to sort through the quotations using the constant comparative method and pasted them onto the wall of my office so as to develop distinct and separate categorical themes. (p. 53)

Yet another variation is to write codes in the margins of fieldnote or transcript pages and put the marked-up pages in file folders or a ring binder. The tabs in the ring binder (or the file folders) divide the project’s materials into

sections, such as (1) instruments used in the study (e.g., interview guides); (2) summary information about the sample of cases (e.g., demographic information for the participants); (3) codebook; (4) in-process memos; (5) theoretical memos; and (6) coded data-texts, which may be subdivided by type of data (fieldnotes, interview transcriptions, etc.) or by types of participant.

No matter how you go about it, a key objective is the ability to manage and retrieve all of the data relating to a topic—let’s say, all instances of talking about the chief influencer in their group. Consulting your codebook, you find the category, **Influencer**, which shows the codes for finding all of the distinctive ways in which your participants referred to the influencers in their groups. Next to the code is information about the location of the data-texts—listed by case number, record number, and transcript page number. A given unit of data can be coded into several categories, so it is helpful to cross-reference the codes in the codebook.

The arrival of new data is often the impetus for revising a coding scheme. You should keep earlier iterations of your codebook so that you can go back and follow the progress of the study as it unfolded over time. When you are ready to engage in conceptual and interpretational work, the pages relevant to the themes you’ve categorized may be pulled out of the binder. Some researchers use colored pens, which enables them to see at a glance the code “families” to which a page or index card belongs. An archive of the complete data is always kept for reference and as a source for more cuttings. However, you should be wary of the data becoming decontextualized from the evidence surrounding it. As Miles and Huberman (1984) remarked about this issue, “meaninglessness of isolated chunks is the potential problem [with putting chunks of data onto cards]” (p. 66).

Conducting an analysis by hand is still preferred by many people, particularly when the amount of data to be processed is not substantial. According to LeCompte and Schensul (1999), it is “not worthwhile to use a computer to code fewer than 100 pages of text data because of the time required to do it” (p. 91). However, a study of even moderate length and complexity can produce data-texts in excess of a hundred pages. At this point, even the ardent traditionalist may wonder if there are better ways of using his or her time. Thus it is to computer-assisted methods we now turn.

Computer-Assisted Qualitative Data Analysis Software (CAQDAS)

In the early 1980s, qualitative researchers began using word-processing programs to type, edit, and store fieldnotes, transcripts, and other texts. Looking back on it, this was the Paleolithic Age of computerized qualitative

research. Tom's experience was typical for that period. He purchased an Apple IIe with 64K of memory and two external "floppy disk" drives (no internal hard drive). Within a couple of years, he got two field projects done and kicked several publications out the door. By then, however, the Apple IIe was sadly obsolete. It had been superseded by more powerful computers that could handle word processors bundled with spreadsheets and database managers in "suites" (e.g., Microsoft Office), which provided a more robust functionality for manipulating and displaying data.

Meanwhile, qualitative researchers in search of a "killer app" didn't have long to wait. The first programs designed for qualitative data analysis, such as The Ethnograph, QUALPRO, QUALOG, and Notebook, appeared in the mid-1980s. By today's standards, they were clunky, hard to use, and lacked many of the features essential to the research trade. For example, many programs required users to "[type] in line numbers and code names at a command prompt, and there was little or no facility for memoing or other annotation or markup of text" (Weitzman, 2000, p. 804). In the 1990s and 2000s, more sophisticated software packages came along—e.g., NUD*IST, HyperResearch, ATLAS.ti, NVivo—spurred by a potent combination of factors: huge increases in memory and processor speed, the growing demand for integrated user applications, and the popularity of qualitative methods in the academy and industry. Commercial sales rose sharply as the perceived "need" for computerized data analysis spread among qualitative researchers. A whole ecosystem also sprang up to support the software: publishers, textbooks, workshops, conferences, consultants, online forums, and so forth. As the 2010s rolled around, revolutionary advances in mobile technology, SNS platforms, and distributed access to data (e.g., cloud computing) gave researchers the ability to share, code, and analyze data on the move, wherever they happened to be (Murthy, 2013).

This chapter is not the place to learn how to use the software. For that, you are advised to try out the demos that publishers release, go to workshops, take university courses, read the scholarly literature on the subject, read product reviews, talk to your colleagues, and of course, comparison-shop. As we will discuss at the end of this section, the usage of a software product can tilt the way you look at—and think about—qualitative data in subtle ways. So, it makes sense to carefully consider what you want out of it before making a major investment of time and money.

Although often billed as theory-building tools, the programs themselves do not literally "theorize." Nor do they operate on data in any way that has not been initiated by you, the user. The key thing these programs do is give you a menu of choices for almost any analytic task you're facing; once you've made your choice, they carry it out quickly, accurately, and comprehensively. In

other words, the software only *assists* in the process of analysis—thus, the acronym, CAQDAS, for computer-*assisted* qualitative data analysis software. But it may be the best assistant, or rather, machine assistant, a data analyst could wish for.

The functionality of CAQDAS is categorized into three broad types of tools: text search, text code-and-retrieve, and code-based theory building. *Text search tools* find words and phrases in one or several databases. They show each instance of a word or phrase and its surrounding context—that is, key word in context (KWIC)—and they index the results. Word lists and concordances (lists of the words or phrases in their contexts) can be created, and the program will do frequency counts of words and phrases instances. Somewhat more arcane counting tasks can be performed, such as “counting the characters coded in each category, which can then be used to obtain the percentage of transcripts coded” (Odena, 2013, p. 358). In most programs, the results of these searches can be sorted into output files, which can be used in further analysis or report writing. Text searching is very helpful if you already have a good idea of what to look for—for example, if you’re taking a more deductive approach. The tools are also very fast. A keystroke or a click instantaneously gives you all of the hits for your word or phrase. However, to perform interesting actions with the text you’ve retrieved, you have to take the next step up to text code-and-retrieve tools.

Text code-and-retrieve tools apply codes to units of text of varying length and retrieve and display these units by clicking on the codes. The initial assignment of codes to text may take about as much time as the manual coding of paper text. (However, recent editions of NVivo have an “auto coding” feature designed to recognize and automatically code patterns of text, such as paragraph styles.) The real savings of time and effort come when you’re ready to retrieve blocks of text linked to a code or a combination of codes. Once you see the linked text pop up in a window, with coding stripes aligned vertically in the margin, you may gain an instant appreciation for the value of this tool.

The codes are also infinitely revisable. Slight changes in how a code is defined can, again with a click or two, immediately change all of the affected texts. CAQDAS conveniently allows you to insert asides, commentaries, and memos in the text or in the margins—an absolutely “must-have” option. In some of the higher-end programs, multimedia (audio/video) materials can be imported, segmented, and coded. Most programs allow multiple users to log on and work on the same project. This latter capability of CAQDAS “[gives] unprecedented levels of transparency. . . . [I]t is now possible for supervisors and examiners to view not only the data, but also what a student has done with that data and to track the processes in the whole research journey” (Johnston, 2006, p. 385).

Text code-and-retrieve tools are often modeled after the grounded theory approach. For example, NVivo is “specifically designed to allow data to be coded and analyzed as they are being collected. The benefits of this program are that it allows for open and axial coding. It can also act as an audit trail and, important for the grounded theory process, memos can be attached to documents and coding categories” (Goulding, 2017, p. 67). Some programs include a “code sequence” feature that permits you to follow one code after another in a text. Some also allow you to sort by precoded data from a “face sheet”—that is, a document that contains summary information about the respondent or the site—rather than by the contents of the fieldnote or interview transcript. This is very helpful when, for example, you want to see all instances of men involved in a certain activity. By permitting a variety of ways of viewing codes, categories, and data, code-and-retrieve tools put grounded theory into motion to an unprecedented degree. They also take a lot of the drudgery out of coding and give you the ability to roam, play, and produce insights with data.

Code-based theory building tools incorporate the functions of the prior two types but go beyond them in exploring the relationships between categories. For example, theory builders allow you to develop a hierarchical model out of links between “nodes” (the text points represented by codes). The links may be defined by one or more different relationships, such as causal (“leads to”), associational (“is a type of”), and Boolean (“and/or/not” relations). The program often uses a variety of graphic objects of different sizes and shapes, colors, and fonts that show the concepts and their linkages. In this way, you can “see” ideas modeled in a window on your screen. The structure of links represents one or more theoretical propositions that can then be tested against samples of text and subsequently revised to obtain a better fit.

Hyperlinks embedded in any file enable the user to make “compound” categories consisting of pictures, texts, audio or video recordings, websites, memos, or annotations. The programs do not actually “touch” or otherwise alter one’s data. Rather, they act as indexes that sit “on top of data,” completely separate from it (unlike Alyssa with her scissors). Finally, many of these programs interface with statistical packages like the Statistical Package for the Social Sciences (SPSS), and have built-in content analysis features such as intercoder reliability.

Does the use of CAQDAS make a study better? Does it improve the performance of qualitative data analysis—compared to manual methods? No one really knows. A lot of excellent qualitative research was done long before the era of personal computing, and a lot of mediocre research utilizing CAQDAS has been published since. Again, we need to be reminded that the software only facilitates the analysis we would otherwise do manually.

If the data are of poor quality, or the researcher lacks the ability to find meaningful associations, even top-of-the-line software won't help. However, CAQDAS certainly does some things better than humans, such as "making sure that no stone was left unturned" (Odena, 2013, p. 364). It is also better at eliminating unconscious bias from the data analysis process; in other words, the programs just don't care if you like or dislike a respondent.

Not surprisingly, for such a powerful technology introduced into a community of scholars with varying work styles, contentious issues surround the use and value of CAQDAS. Some worry about its impact on the creative spirit of inquiry. Particularly among students, CAQDAS can promote a sort of "code and retrieve cycle," in which individuals "have found themselves coding in a somewhat mechanistic manner, often for excessively long periods of time, without using some of the in-built tools to help them to see the proverbial wood from the trees" (Johnston, 2006, p. 383). This "incessant desire to code every part of a document" (p. 383) seems to be at work in Schiellerup's (2008) vivid (and somewhat disconcerting) tale of her dissertation experience:

In total there were more than 1000 pages of transcripts and field notes. I could not code much more than two interviews a day if I also wanted to be writing about the interpretations that I was making by recording them in my research diary or by using the memoing function in ATLAS/ti. The interview frame had been constructed to gather more data on the conceptual framework. Therefore, the codes not surprisingly became very full as the coding progressed. I began to despair of the purpose of all this coding. I assumed I had to go back to reading them to develop the analysis further from there (constant comparison). This began to look increasingly unrealistic the fuller the codes became. I developed additional codes to help me build theory and to decrease the size of individual codes. At final count there were close to 900 codes. (p. 167)

Others are concerned that the software shapes the way we think about data. The settings and options of CAQDAS programs may contain implicit "theories" of analysis that guide users along certain tracks and away from other possibilities. Some kinds of software, for example, rely more on hierarchical structures than other types of modeling. The unwary user might therefore create a top-down model for the data when a looser, richly networked one would be a better fit. (We should note, however, that open-source programs exist for qualitative researchers that allow more flexible, transparent uses; Greenberg, 2011.) The more implicit the bias in the software, the harder it is to recognize what it may be doing to your project. Schiellerup (2008) offers this advice: "It is important . . . to have a handle on the data analytical process before embarking on using CAQDAS, otherwise

one may find oneself inappropriately socialized by the ‘agenda’ inscribed in the software” (p. 168).

Finally, some express concern about the loss of “feel” with the data. More seriously, CAQDAS may promote a degree of emotional and intellectual alienation from the cultural scenes we study. This is not a neo-Luddite call to return to cards, scissors, and tape. Computing has infiltrated every part of the research process, and it is probably impossible to return to “the garden”—the mythic imaginary of a nondigitized world. Rather, it is just to suggest that those of us who study lived experience should remain vigilant to how information technology affects our sensitivity to voice, touch, emotion, and all manner of sensuous detail, as well as our sensitivity to alternative ways of thinking about communication. If we lose those sensitivities, we certainly have lost too much.

Conclusion

We began this chapter by calling attention to some of the challenges of qualitative analysis. Some of these challenges, such as the need to be open to multiple (possibly competing) interpretations, are central to the enterprise of qualitative research. Challenges such as that will always be with us. On the other hand, progress has been made in explicating how data analysis works, so that novices no longer have to feel like they’re groping in the dark for a light switch. And the widespread use of computer-assisted qualitative data analysis software (CAQDAS) has brought a new level of systemization to the use of codes, categories, and theory-building models. These developments stand in sharp contrast to the days, not all that long ago, when qualitative data analysis was a “mysterious, half-formulated art” (Miles, 1979, p. 593).

At the same time, researchers still value the artful side of data analysis. Many of them still find ways to improvise “work-around” solutions to problems, and to figure out clever yet justified ways of coding, categorizing, and conceptualizing. These user-generated practices—which are sometimes reported in journals like the *Journal of Contemporary Ethnography* and *Qualitative Inquiry*—are often aimed at getting more “performance” out of the “factory-produced” models (i.e., textbooks like this one). As long as you take care not to manipulate data in ways that distort the local contexts of meaning, there is nothing wrong with doing this. In fact, the satisfaction you feel at the end of a study can be even greater when you’ve found your own path to a set of compelling, useful findings. In the next chapter, you will move to the next level of qualitative sense-making and learn strategies for creating, evaluating, and enhancing interpretations of data.

Exercises

1. Take a discursive text, such as an interview you have transcribed, or a public text like a newsmaker interview or messages from an online forum, and code it with a partner.

- Did you and your partner tend to generate the same codes?
- What were the areas of disagreement or uncertainty?
- How many times did you revise the codes until you were satisfied with their meanings?
- Did the coding help you think about how to develop categories?

2. Published qualitative studies in communication report their data analysis procedures with varying specificity. Whether this is a “problem” is itself a subject of some debate. Choose two communication journal articles that use qualitative field methods. Compare and contrast the articles’ descriptions of their data analysis and interpretation procedures.

- In what respects do these articles differ?
- Is the way that the data analysis is reported for each article appropriate to its goals, findings, and overall style?
- Is important information about the study’s data analysis left out or described inadequately?
- Would you advocate a standard for reporting data analysis in journal articles? If so, what elements would constitute that standard? If not, why not?