

Grounded Theory and Thematic Analysis

Or: How to Analyze Qualitative Data

Sources today:

- Virginia Braun & Victoria Clarke. Using Thematic Analysis in Psychology. Qualitative Research in Psychology. 2006. 3:2, 77-101.
- Kathy Charmaz. Constructing Grounded Theory, 2nd Ed. Sage.

Goal: Identify Patterns in Data

- Suppose you have a bunch of free-form, qualitative data
 - Interview transcripts
 - Web forum posts
 - Survey responses
- You want to identify and report on patterns.
 - "70% of the participants wanted a Mexican restaurant at Price Center."

Problem I: Data Are Messy

- "I'd love a place that served empanadas."
- Are empanadas Mexican cuisine? Or Argentine? Or both? Do most Argentine places serve empanadas?
- Do these data support a proposal to add an Argentine restaurant?

Problem 2: Subjective Interpretation

- "You said that students want healthier food options. How do you know?"
 - "We counted 47 different responses that asked for healthier food."
 - "Like what?"
 - "Here's one: 'Food at Price gives me a stomachache.'"
 - "Is that about health, or do \$5 tacos induce financial stress?"

Problem 3: Bias

- "You started the survey because you wanted an Argentinian restaurant, so it's not surprising that you found students wanted one."
- "How many wanted an Italian restaurant?"
- "Uh, I don't know."
- "Aha. Your analysis was biased."

Structured Analytic Methods

- Want: *structured analytic method to reduce and reveal bias*
- Time for some epistemology...

Epistemology

- Wikipedia: **Epistemology** (from Ancient Greek ἐπιστήμη (epistémē) 'knowledge', and *-logy*), or the **theory of knowledge**, is the branch of philosophy concerned with knowledge.
- What does it mean to know something?
 - Positivism: facts derived from data with logic
 - Interpretivism [1]:
 - "the important reality is what people imagine it to be" [2]
 - "systems exhibit emergent behaviors not reducible to their component parts"

[1] K.-J. Stol, P. Ralph, and B. Fitzgerald, “Grounded theory in software engineering research: a critical review and guidelines,” in *Proceedings of the 38th International Conference on Software Engineering*, New York, NY, USA, May 2016, pp. 120–131. doi: [10.1145/2884781.2884833](https://doi.org/10.1145/2884781.2884833).

[2] Bogdan, R., Taylor, S. (1975). *Introduction to qualitative research methods: A phenomenological approach to the social sciences*. New York: John Wiley.

Positivism Vs. Interpretivism

- Positivism: typical of natural sciences
 - My experience: most computer scientists are positivists
- Interpretivism: typical of social sciences
 - Qualitative methods come from *this* community

A Core Technique: "Coding"

- Sorry: this term came from sociologists (not computer scientists)
- "Code" == "Category"
- Coding is about categorizing data elements.

Coding Example

Data extract	Coded for
<p>it's too much like hard work I mean how much paper have you got to sign to change a flippin' name no I I mean no I no we we have thought about it ((inaudible)) half heartedly and thought no no I jus- I can't be bothered, it's too much like hard work. (Kate F07a)</p>	<ol style="list-style-type: none">1. Talked about with partner2. Too much hassle to change name

Coding Guidelines

- Add as many codes as possible
 - But can keep research questions in mind
 - Items can have multiple codes
- Can keep it simple (spreadsheet) or use fancy software (MaxQDA, ATLAS.ti)

Coding Exercise

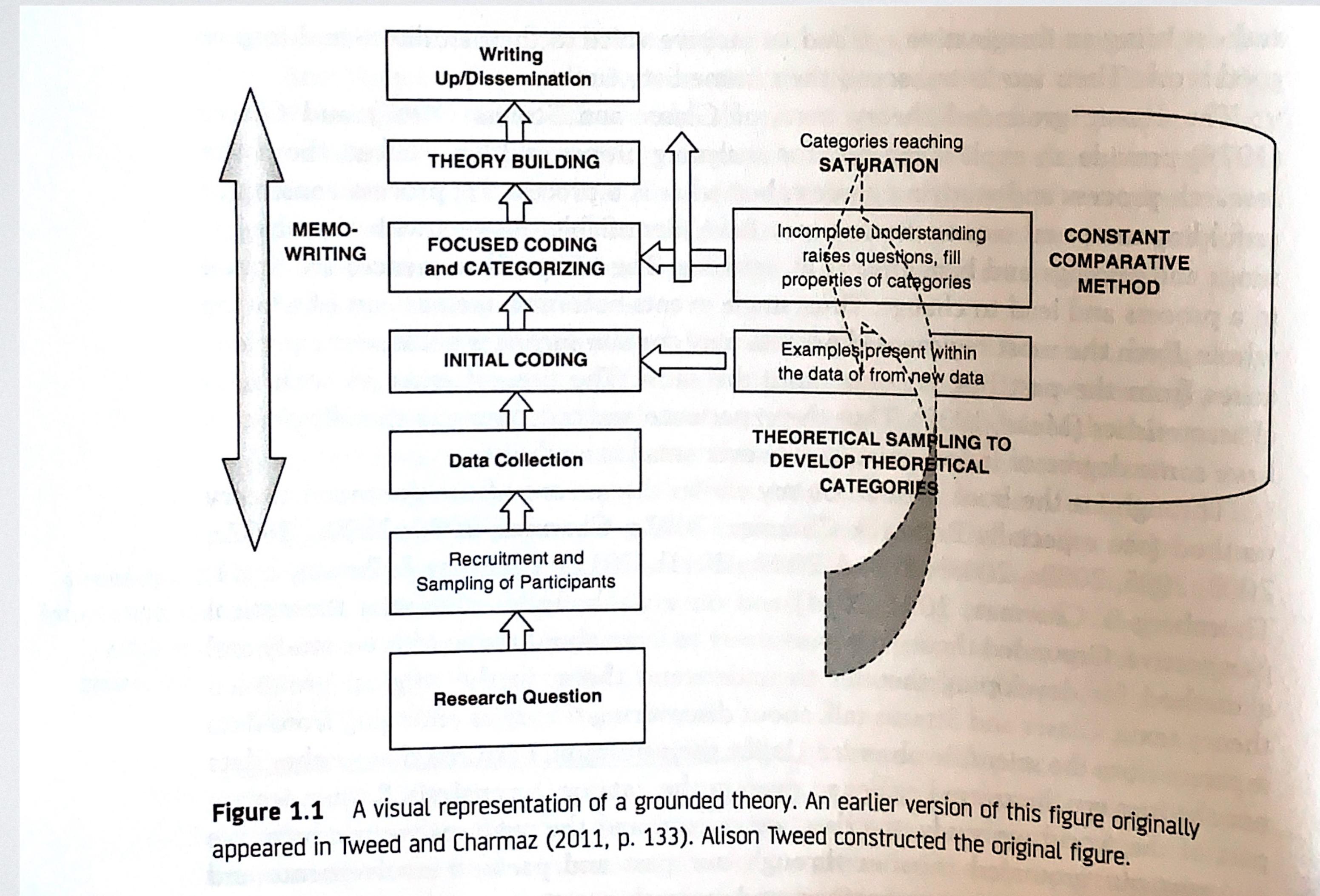
What To Do With the Codes?

- Grounded Theory
 - Goal: build a *theory* that explains the data
- Thematic Analysis
 - Goal: identify *themes* that occur in the data

Grounded Theory

- Pick one:
 - Classic / Glaserian
 - Straussian
 - Constructivist
- Today: focus on constructivist

Grounded Theory



Tenets (per Charmaz)

- Collect & analyze data simultaneously, iteratively
- Analyze actions & processes rather than themes and structure
- Use comparative methods
- Draw on data to develop new conceptual categories
- Develop inductive abstract analytic categories through systematic data analysis
- Emphasize theory construction rather than description or application of current theories
- Engage in theoretical sampling
- Search for variation in studied categories or process
- Pursue developing a category rather than covering a specific empirical topic

Focused Coding

- Focused coding: identify codes that are particularly interesting (worthy of further study)
- Focused codes appear more frequently or have more significance

Memos

- What is going on?
- What are people doing/saying/trying to say/remaining silent about?
- What do people take for granted?
- What connections/comparisons can you make?
- Examine "in vivo" codes (codes taken from data): "living one day at a time," "jumping through hoops"

Memo Example (Context: Brain Injury)

BOX 7.6

Early Memo: Explaining All-Encompassing Loss

Explaining all-encompassing loss means making explicit unknown or forgotten meanings of the magnitude of patients' loss. Explaining here means pointing out types of loss patients have experienced, delineating their extent, and making these losses known and understood. Explaining all-encompassing loss means taking the *patient's* perspective and looking at what is lost. Loss resides in the chasm between the life once lived and current institutional existence. Participant D points out, 'I think, I don't think enough emphasis is placed on that fact that these people have lost their life. And basically through that they have perhaps experienced a huge range, huge range of losses and because of their behaviour they may have alienated people so that there is just this massive aspect of loss that they all have.' Thus, loss can result in spiraling consequences. Awareness of losing one's life, one's way of being in the world causes patients enormous suffering that they may express through frustration, anger, and aggression, which leads to being rejected and further suffering, and subsequently more acting up. Misbehaving cloaks loss, and then, staff misread the patient's behavior. Conditions that exacerbate this process include the nature of the patient's impairment and its relative visibility, the extent to which it complicates daily life, and the institutional situation itself. To what extent does being subject to this situation impart messages that lapses in self-control are routine events and therefore desensitize patients as well as staff to troublesome behavior?

Explaining all-encompassing loss not only asks the listener to envision losses, but also to envision who a patient was before experiencing brain impairment. Thus, staff would gain a different image of the patient than that of the person they encounter in their daily work. Explaining links the past with the present and accounts for the present. (In contrast to Participant D's accounting for the present by looking at the past, Participant B considers impairment but concentrates on the present, not the magnitude of loss, nor the suffering it may cause.) Participant D adopts the role of the teacher who elucidates for the interviewer why patients act as they do. To what extent can or does she make her views known and heard? How does she deal with co-workers who fail to grasp these meanings of loss?

By explaining all-encompassing loss, Participant D presents herself as empathetic, insightful, and different – separate? – from co-workers. How does her empathy alter relationships with patients? How and when do her insights affect her work with patients? What, if any, are the implications of setting oneself apart from co-workers in this setting?

From Tweed and Charmaz (2011, pp. 140–141).

Theory Development

- Leverage memos and constant comparison
 - Compare codes to codes, cases to cases
 - e.g. Why are these two cases different?

My Opinion

- Grounded theory is great if you do not have preconceived notions
- But apply carefully...

Example

- Barke, James, and Polikarpova:
 - "Through our grounded theory analysis, we identified two main modes of developer interactions with Copilot: **acceleration** and **exploration**. In acceleration mode, a programmer uses Copilot to execute their planned code actions, by completing a logical unit of code or a comment. **Acceleration works within user's sense of flow.** For example, recall how in Sec. 2.1 Axel accepted Copilot's suggestion of `rule.split(" => ")`, knowing it was what he wanted to type anyways. This is a characteristic example of acceleration, where Copilot was helping him program faster.
 - "**In exploration mode, a programmer relies on Copilot to help them plan their code actions.** A programmer may use Copilot to assist with unfamiliar syntax, to look up the appropriate API, or to discover the right algorithm. In Sec. 2.2, when Emily was searching for the right set of matplotlib calls, she was considering alternatives, gaining confidence in the API, and simply trying to learn how to finish her task. All of these intentions are part of the exploration mode when using Copilot. **We found that programmers alternate between these two modes** as they complete their task, fluidly switching from one mode to the other."

Thematic Analysis

- Phase 1: familiarizing yourself with your data
- Phase 2: generating initial codes
- Phase 3: searching for themes
- Phase 4: reviewing themes
- Phase 5: defining and naming themes
- Phase 6: producing the report

Phase I: Familiarizing Yourself With the Data

- Transcribe data
 - "The time spent in transcription is not wasted, as it informs the early stages of analysis, and you will develop a far more thorough understanding of your data through having transcribed it."
- Read it all
- Write initial ideas

Phase 2: Generate Initial Codes

- "Codes identify a feature of the data (semantic content or latent) that appears interesting to the analyst, and refer to 'the most basic segment, or element, of the raw data or information that can be assessed in a meaningful way regarding the phenomenon'" (Braun and Clarke)

Phase 3: Searching for Themes

- Group codes into categories
- Merge duplicates
- Build hierarchies

Phase 4: Reviewing Themes

1. Read collated extracts for each theme and ensure they all fit the theme
2. Consider each theme in context of the whole data set: re-read data set, and add missing codes

Phase 5: Defining and Naming Themes

- Can you describe the scope and content of each theme in a few sentences?
- If not, perhaps this is not a theme.

Phase 6: Writing a Report

- Explain each theme
 - Justify with examples (quotes)
- Give quantities when helpful ("28% of participants felt food from their cultures of origin was not represented at Price")

Conclusion

- Grounded theory is very useful for generating deep theories
 - But expensive
- Thematic analysis is more lightweight, but may not lead to theories