„Empirische Bewertung in der Informatik“

Grounded Theory: Empirisch begründete Theoriebildung

Prof. Dr. Lutz Prechelt
Freie Universität Berlin, Institut für Informatik
http://www.inf.fu-berlin.de/inst/ag-se/

- Qualitative Forschung und Grounded Theory (GT)
- Kodieren (offen, axial, selektiv)
- Theorie-Sensitivität
- Theoriegetriebene Datenauswahl
- Reflektion
- Qualitätsmaßstäbe für GT-Studien
- Unterstützende Software: ATLAS.ti

Course „Empirical Evaluation in Informatics“

Grounded Theory

Prof. Dr. Lutz Prechelt
Freie Universität Berlin, Institut für Informatik
http://www.inf.fu-berlin.de/inst/ag-se/

- Qualitative research and GT
- Coding (open, axial, selective)
- Theoretical sensitivity
- Theoretical sampling
- Reflection
- Quality criteria for GT studies
- GT software: ATLAS.ti

Qualitative research

- Qualitative research aims at providing either of
  - rich and precise **descriptions** of phenomena
    - Ideally with little or no interpretation of the data
  - rich **interpretations** of what phenomena mean
  - rich **explanations (theories)** of how/why phenomena come to be
    - With no more interpretation than necessary
- It is used in domains where at least one of the following is true:
  - non-measurable aspects of human behavior are relevant (or even central)
    - such as emotions, judgment, etc.
  - the feasible quantitative theories are too poor
    - as their scope is too narrow
    - usually also in the realm of human behavior

Purposes of qualitative research

- Qualitative research (like quantitative research) can have many different kinds of purpose:
  - develop basic knowledge
  - guide practitioners
  - evaluate programs
  - develop policy
  - complement quantitative findings
  - build research instruments
  - serve commercial or political ends

Qualitative vs. quantitative research

- Quantitative research
  - Input data: Quantitative or qualitative
    - Qualitative input data is strictly categorized to make it countable
  - Analysis methods: **Quantitative**
    - This is where the name comes from!
  - Results: Quantitative or qualitative
    - Often qualitative, such as *"A is significantly faster than B"*
- Qualitative research
  - Input data: Quantitative or qualitative
  - Analysis methods: **Qualitative**
    - This is where the name comes from!
  - Results: Qualitative

"Grounded Theory"

- The Grounded Theory approach is a qualitative *research method*
  - that uses a systematic set of procedures…
  - to develop inductively…
  - a theory that is grounded in observations (data)…
  - about some phenomenon.
- The *results of using this approach* are also called "grounded theories"

Research process and result intermingle

- "It is no linguistic accident
  that *'building', 'construction', 'work'*
  designate both a process and its finished product.
  Without the meaning of the verb
  that of the noun remains blank.

  John Dewey, *Art as Experience*, 1934

Running example: TIPP

- TIPP stands for *"Task Influence in Pair Programming"*
- It is a study trying to understand
  - how the nature of the programming task
  - influences the behavior and success of programmers working in pairs
- TIPP is based on observing a modest number of pairs.
  - Live observation in a semi-realistic working environment.
  - Audio, video, and screen video are captured.
- Each pair solves two small programming tasks:
  - *Phonecode* has an algorithmic nature
  - *Taglib* primarily involves understanding and using libraries
- TIPP will repeatedly be used for illustration in this presentation

Research questions in GT studies

- As all research, GT work is guided by a research question
- GT-style questions tend to be concerned with **action** (or interaction) and **process**
- Examples:
  - What actions and interactions make a pair-programming pair successful?
  - How do designers identify important abstractions?
  - How do programmers go about understanding complex logic someone else has written?

Character of research results

- Grounded Theory research brings about much richer kinds of results than quantitative research
- Example: For this research question
  - "What actions and interactions make a pair-programming pair successful?"
- A quantitative result (from a controlled experiment) may be
  - "Pairs that switch the roles of driver and observer regularly produce software with significantly fewer defects compared to pairs that do not switch or switch only rarely."
- A Grounded Theory for the question will actually describe many different actions and interactions used, their conditions and interplay, and how they influence the pair's success in all its aspects.

Elements of the Grounded Theory method

- Coding:
  - open coding (What?)
  - constant comparison (Grounding)
  - axial coding (When, where, why, how, what follows?)
  - selective coding (Telling the story)
- Theoretical sensitivity
- Data gathering:
  - everything is data
  - theoretical sampling
- Reflection:
  - asking questions to enhance theorectical sensitivity
  - writing memos and diagrams

"Coding"

- Coding is the process of analyzing data
  - GT is constantly involved with data, not with a theory
  - This is what provides the grounding
- The two mean ingredients of coding are
  - making comparisons (among data)
    - GT is often called "constant comparative method of analysis"
  - asking questions
- The outcome of coding is elements of a theory
  - concepts and their relationships

"Open coding": Identifying the What

- Consider your phenomena
  - Phenomenon: Anything, small or large, that is (a) reflected in the data and (b) possibly relevant
- Assign a concept to each small phenomenon
  - Concept: An abstract(!) descriptive label
    - Beware of standard technical terms: They may be loaded with unwarranted assumptions
  - Anything can be a concept: thing, abstraction, action/event, attribute, relationship
- Group concepts into categories and subcategories
  - Category: A concept representing a set of related concepts
    - Note that categorization is not an *instance-of* relationship; it is a *pertains-to* (*has-to-do-with*) relationship
  - Subcategory: A class of similar concepts
    - This is usually(???) *instance-of*

Open coding: Example

- Sitting in an expensive restaurant, we notice a lady in red standing in the look-in kitchen, apparently doing nothing.
- As Grounded Theorists, we decide to find out what her job is based on observation and induction.
- We observe phenomena [and code them by concepts]:
  - She is intently looking around in the kitchen. [a work site]. She is [attentively] [watching] [kitchen work].
  - Someone walks to her and asks a question. She answers. She is [passing information].
  - She is standing in the middle of the activity without disrupting it. [unintrusiveness].
  - She walks quickly [efficiency] into the dining area and proceeds to [watch] here also. [monitoring]

Open coding: Example (2)

- (But what is she monitoring?) She is monitoring:
- How the waiter interacts with the customer [quality of service];
- the time between seating, ordering, delivery of food [timing of service];
- [customer response] and [satisfaction with the service].
- As a waiter receives the orders for a large party, the moves in to help him. [providing assistance].
- She talks with the maitre d', they look around for empty tables and the eating status of the guests. [conferring]

- The conceptual labels are chosen such that they might be useful again in other contexts than that where they were first seen.

Open coding: Example (3)

- Now we group our concepts into **categories**:
  - [Monitoring], [conferring], [watching] group as [[assessing and maintaining the flow of work]].
    - Note that this categorization is quite preliminary, as are all conceptualizations in this phase of GT work.
  - [Attentiveness], [efficiency], and [unintrusiveness] are [[attributes]]. But attributes of what?
    Of a [[person good at assessing and maintaining the flow of work]]. A better name might be [[food orchestrator]]

So we end up with 1 category and 2 subcategories:

- Category: [[food orchestrator]]
  - Subcategory: [[types of work for assessing and maintaining the flow of work]]
  - Subcategory: [[Attributes of a good food orchestrator]]

Open coding: Naming a category

- Purpose of category names:
  - Allow to remember something, refer to and speak about it
- Sources of category names:
  - the researcher's own invention
  - common terms from the field (literature)
  - taken directly from the data (e.g. interview response)
- Beware of unavoidable connotations of common terms!

- Names need not be perfect
  - They can (and sometimes will) be changed anytime later

Open coding: Properties and dimensions

- For each concept, we can identify a number of *properties* that each instance of the concept will (or may) exhibit
  - e.g. [watching] has a [duration], a [target], an [intensity], and perhaps a [manner] and [frequency].
  - e.g. (in TIPP) [browsing documentation] has a [speed], a [repetitiveness], a [specificity], a [thoroughness], etc.
- Each property is a concept itself.
- Each property has *dimension*: its range of possible values (quantitative or, most often, qualitative)
- Identifying the properties of a concept or the property values of an instance of the concept is called *dimensionalizing*.

Open coding: Granularity

- How large is the unit of observation for which we will assign a concept?
- This may vary widely
  - it depends on the kind of material (text, optical observation etc.)
  - and the current focus and goal of the coding
- Examples:
  - Individual words, phrases
  - Individual sentences; paragraphs
  - Individual turns in a dialog or discussion
  - Individual movements
  - Episodes of action
- In later phases of GT work, granularity tends to go up
  - and the frequency of open coding tends to go down

Constant comparison (Grounding)

- The way a GT achieves its grounding in the data is by continual comparison during the theory building:
  - Compare phenomena to similar phenomena; compare concepts to similar concepts from your data
    - What is equal? What is similar? What is different?
    - Find additional properties and dimensions
    - Sharpen the concepts
    - Detect unwarranted assumptions and lack of grounding
    - Enrich the categories with more concepts
  - Compare concepts to different concepts from your data
    - What do they still have in common?
  - Ask questions
    - Who? When? Where? What? How? How much? Why?
- Purpose: Enhance your theoretical sensitivity

Theoretical sensitivity

- An important personal quality of the researcher.
- Awareness of the subtleties of the meaning of data
  - insight, understanding, judging pertinence
- Requires a balance of creativity and rigorousness
  - Periodically step back and ask what's going on
  - Maintain a skeptical attitude
  - Follow the research procedures
- Sources of theoretical sensitivity:
  - Pertinent literature (technical, non-technical)
  - Professional or personal experience
  - Research experience
  - Analytic process (continual interactions with the data)

Theoretical sensitivity: techniques

- Constant comparison
  - as described above
  - Joking example
    (regarding "Detecting unwarranted assumptions"):
    - At a job interview:
    - "Where did you receive your training?"
    - *"In Yale"*
    - "Great. And what is your name?"
    - *"Yim Yohnson"*

Theoretical sensitivity: techniques (2)

- Analysis of a word, phrase, or sentence
  - Pick an item of which you expect significance
  - Extensively list all of its possible meanings (likely or not)
  - Try to validate these meanings from the data
- Flip-flop technique
  - When analyzing some situation X and out of ideas
    - The data seems not to provide relevant information
  - Imagine the opposite of X.
  - What characteristics would it have?
    What is important? Is any of this in the data?
    - e.g. when analyzing the behavior of a market leader, imagine the characteristics of a market that has no leader
    - e.g. when analyzing a confused TIPP team, how might a really sharp team act? How not?

Theoretical sensitivity: techniques (3)

- Close-in comparison
  - Imagine to compare your phenomenon to something that is related, but different
    - Similar in many, but not all respects
  - What would be the relevant similarities and differences?
  - Is any of this in the data?
  - Example: To understand the aspects of the body image of overweight women, imagine everyday situations of an overweight woman vs. a woman of normal weight.
  - TIPP example: To understand the characteristics of clueless, impatient documentation browsing, imagine the behavior during clueless, but patient, browsing.

Theoretical sensitivity: techniques (4)

- Far-out comparison
  - Like before, but compare to something very different
    - Similar only in few respects
  - Example: "How is a priest like a prostitute?"
    - Both provide a public service; both treat information confidentially; both perform counseling
  - TIPP example: How would a pair act that has solved the exact same task just an hour ago?

Theoretical sensitivity: techniques (5)

- Waving the red flag
  - Whenever you think you are confronted with something absolute
    - Like "all", "always", "never", "everyone knows" etc.
  - Wave an imaginary red flag in your mind and question if the absolute is really true
    - What exactly does that mean?
    - Is it always so?
    - Why? How?
    - What are the consequences?
    - Is there any way to get around this?
  - With respect to the data, this is most relevant when analyzing secondary sources (such as interviews)
    - rather than direct observation
  - With respect to your own inferences, it is always relevant

Elements of the Grounded Theory method

- Coding:
  - open coding (What?)
  - constant comparison (Grounding)
  - axial coding (When, where, why, how, what follows?)
  - selective coding (Telling the story)
- Theoretical sensitivity
- Data gathering:
  - everything is data
  - theoretical sampling
- Reflection:
  - asking questions to enhance theorectical sensitivity
  - writing memos and diagrams

Axial coding: Identifying relationships

- Axial coding identifies connections between categories
  - When, where, why, how, what follows?
- After open coding has broken up the data for analysis, axial coding puts it back together in new ways
  - by making connections between a category and its subcategories
    - specifically conditions, context, strategies, and consequences
  - This is still about developing a category, but goes beyond properties and dimensions.

Axial coding: The *coding paradigm*

- For each category X of primary interest, try to describe
  - the **causal conditions** that lead to such a phenomenon
  - the **context conditions** which the phenomenon creates
    - a set of properties
  - the **actions or interactional strategies** used by the participants
  - the **intervening conditions** that influence the actions and choice of strategies
  - the **consequences** that arise from the phenomenon.
- Each of these forms another category
  - which is a subcategory of X.
  - X is called the **phenomenon** (often a situation or event)
    - We typically investigate only about half a dozen such categories (everything else is subcategories)

TIPP example

- Regarding [documentation browsing] investigate
  - **causal conditions:** What leads the pair to attempt d.b.? When exactly do they start?
  - **context conditions:** What technical means are available? What manner of browsing is used? etc.
  - **actions or interactional strategies:** What recipe for success is behind the browsing? How do the partners communicate about it? Change it?
  - **intervening conditions:** What intermediate successes (or lack thereof, or other events) prompt changes in strategy, manner, means?
  - **consequences:** What is the result of successful browsing? Of unsuccessful browsing? Of browsus interruptus?
- and others that appear potentially relevant.

Axial coding: Two notes

- 1. A note on *context conditions*:
  - They are <u>not</u> only the environment that is already present when the phenomenon comes into existence
  - Rather, they are also the properties of the phenomenon itself
    - which form the context of the strategies chosen for dealing with the phenomenon
  - Environmental conditions will often be classified as *intervening conditions*
    - as they influence the strategies and consequences from the outside
- 2. As usual, all findings need to be grounded in data.
  - In fact, a researcher will quickly alternate between open coding and axial coding much of the time.

Axial coding: What happens

During axial coding, 4 strands of action occur in parallel:

- 1. The (hypothetical) relating of subcategories
  - by describing the relationships between them and the phenomenon (the category, not its instances)
- 2. The verification of these hypotheses against data
  - by constant comparison, including recognizing variations and counter-examples
- 3. The continual search for properties and their values
  - by asking questions
- 4. The beginning exploration of variation in phenomena
  - by comparing different instances of phenomena and noting patterns in the data
- There is constant interplay between (A) proposing statements and (B) checking them

Selective coding: Telling the story

Selective coding aims at putting together the worked-out categories into an actual Grounded Theory

- It is done by selecting one core category
  - often already present, sometimes formed anew
- and working this into a *narrative* about the central phenomenon of the study
  - arranging all other categories around the core category
    - much like in axial coding: context, conditions, strategies, consequences
  - often roughly in the form of (nested) if-then rules
    - that hold at least by-and-large for at least most of the cases
- always grounding each relationship you propose directly in the data (validation)

TIPP example

Conceivable candidate core categories for TIPP:

- documentation browsing
- documentation reading
- decision-making
- clarifying discussion
- leadership switch
- trial-and-error
- recognizing error
- ...?

- (Be aware that this is pure, wild speculation!)

Selective coding: Emotional requirements

Successfully tackling selective coding requires the following insights:

- 1. Faith that it **can and will be achieved**
  - although it may seem intimidating
- 2. Recognizing that it **has to be worked at** and is not based on just romantic inspiration
  - the process is actually similar to axial coding
- 3. Recognizing that it is nothing to be found (like the solution to an equation), but **something to be created**
  - and hence there is more than one way to do it
- 4. Accepting that you **cannot pack everything into one theory**
  - so you may get two or more GTs almost for the price of one

Selective coding: If a case does not fit in

If a case appears not to fit with the rest of the theory, consider the following reasons:

- **Lurking conditions:** Some relevant conditions have not yet been integrated into the theory
  - e.g. something like a person's behavioral type is often a helpful concept
- **Process:** the case is not homogeneous over time, but rather has switched categories
  - see below

- Unless these occured by oversight, theoretical sampling may now be the appropriate action
  - Gathering more data to fill in recognized gaps; see below

Selective coding: Process

- Any longitudinal study involves process:
  Change over time
- A GT must capture as much process as is required for a good understanding of the phenomena:
  - The rules that describe an important entity (e.g. person) are not static
  - Rather, they change in the course of action or interaction
  - This change is moderated by intervening conditions
- Process description captures:
  - The rule change,
  - the action or interaction sequence leading to it,
  - the intervening conditions moderating it
- The change as such is also a category with dimensions

Selective coding: Process (2)

Example:

- Rule change:
  - A manager changes management style towards a harder, more restrictive one
- Action sequence leading to it:
  - A decline in productivity caused by a loss of discipline of the employees represented by many small actions/interactions
- Intervening conditions moderating it:
  - When discipline and productivity go back up, the style change is gradually reversed

Selective coding: The conditional matrix

- A device for increasing theoretical sensitivity w.r.t.
  - the range of conditions relevant to an action
  - the range of possible consequences of an action
  - the relationships between all three
- Consists of a **sequence of nested scopes** (specificity layers, onion rings) in each of which conditions, actions, and consequences can be considered
  - For most actions, conditions from more than one layer are relevant
    - Without the matrix, the researcher would often concentrate on one layer only
  - For many actions, consequences occur on more than one layer
  - (And yes, the conditional matrix is not a math matrix at all)

Selective coding: The conditional matrix (2)

- Example layer sequence
  (and relevant dimensions or features)
  - Action; Interaction
    - (strategies, motivation, goals, ...)
  - Individual; Group
    - (biography, personality, knowledge, ...)
  - Sub-organization; Organization; Institution
    - (structure, rules, problems, history, ...)
  - Community; National; Internat'l (law, culture, history, ...)

- Other sequences of layers may be more appropriate
  - depending on the given study
  - ideally so that many actions can be traced several levels up in their conditions and consequences

Elements of the Grounded Theory method

- Coding:
  - open coding (What?)
  - constant comparison (Grounding)
  - axial coding (When, where, why, how, what follows?)
  - selective coding (Telling the story)
- Theoretical sensitivity
- Data gathering:
  - everything is data
  - theoretical sampling
- Reflection:
  - asking questions to enhance theorectical sensitivity
  - writing memos and diagrams

Data gathering

- Data gathering refers to making the observations to be analyzed

- It involves:
  - Deciding when, what and how much to observe
    - see Theoretical Sampling
  - Finding or creating the context in which to observe
    - e.g. field observation, interview, or others
  - Making and recording the actual observations

Everything is data

- GT has some typical (preferred) modes of observation:
  - field observation
    - In principle the richest and most reliable form
  - interview (unstructured or loosely structured)
    - Often the easiest way of obtaining rich data
- However, GT accepts any kind of data that is available as potentially valid data for the study
  - documents of all sorts
    - in particular episodic kinds (announcements, protocols, etc.)
  - third-party literature such as other studies
  - previously existing audio and video material
  - etc.
- No data needs to be ruled out as biased, subjective etc.
  - Instead, it is put into perspective during analysis

Theoretical Sampling (TS)

- TS marks the most pronounced difference of GT to most other methods (qualitative or quantitative)
- It refers to the merging of data gathering and analysis
  - Initially, some data is gathered ("open sampling")
  - Then analysis starts very broadly
  - When areas of interest have been identified,
    - more data is gathered ("open sampling")
    - and analyzed more focussedly
  - During axial coding: further data gathering is done to maximize the differences at the dimensional level
    - "relational and variational sampling"
  - During selective coding: further data gathering is done to
    - verify the story line and fill in poorly developed categories
    - "discriminative sampling"

TS: Some notes

- TS means most data gathering is directly driven by theory development
  - namely by the relevance of certain concepts
    - which is recognized gradually.
  - Initial sampling is thus purely provisional
    - as the relevance of the data cannot yet be known.
  - During a GT study, the focus of TS generally gets narrower and deeper
- TS mostly samples incidents, not persons
  - TS implies non-text data need not always be fully transcribed
- Consistency: Data should be gathered on each category systematically
  - but chances presenting themselves should be exploited

TIPP example

Incremental theory-building will drive the decisions what further data to obtain in TIPP, e.g.:

- Pairs solving the same (vs. other) tasks
- Strong, capable pairs vs. average vs. underaverage
  - or gelled pairs vs. unacquainted pairs
- Pairs working in the lab vs. real-life offices
- Pairs working on artificial tasks vs. real-life tasks
- Pairs whose members have certain personality types, work styles, etc.
  - (as far as this can be known in advance)
- etc.

TS: When to finish

- Theoretical sampling ends when each category is *theoretically saturated*.
  - no new or relevant data seems to emerge
  - all paradigm elements are accounted for
    - along with variation and process
    - conditions (causal, context, intervening), action/interaction, consequences
  - category relationships are well established and validated
- GTs without theoretical saturation are inadequate

Elements of the Grounded Theory method

- Coding:
  - open coding (What?)
  - constant comparison (Grounding)
  - axial coding (When, where, why, how, what follows?)
  - selective coding (Telling the story)
- Theoretical sensitivity
- Data gathering:
  - everything is data
  - theoretical sampling
- Reflection:
  - asking questions to enhance theorectical sensitivity
  - writing memos and diagrams

Reflection

- In contrast to many other methods, GT does not pretend that the person of the researcher might be unimportant
- Rather, GT understands that the researcher must reflect about the work and must bring in his/her creativity
  - Subjectivity is unavoidable in research of social proceses
    - and should hence be managed (rather than considered an unsolvable problem)
  - Maximizing creativity is a key to a successful GT study.
  - Note: All creativity is disciplined by the strict framework of the GT method
- Reflection and creativity happens mainly via two kinds of activities:
  - asking questions
  - writing memos

Asking questions

The most basic and constant activity during GT work is asking questions:

- During constant comparison:
  - How is this incident like that one?
  - How is it different?
  - Do these two concepts in fact belong to the same category?
- Regarding theoretical sensitivity:
  - What other properties may this category have?
  - Can I know the dimension of this property?
  - What other conditions may this action have?
  - For any action: Why? How? When? What follows?
  - How could I validate this hypothesis?
- etc. etc.

Writing memos and diagrams

- Memos are short essays about the work
- Diagrams are schematic visual representations of relationships

- Roles of memos:
  - relieving memory of details
  - summing up intermediate results
    - increasing theoretical sensitivity
  - keeping track of how the work proceeded
  - sketch how the work should proceed further
  - providing overview
- Roles of diagrams:
  - providing overview

Types of memos

There are basically three kinds of memo:

- Code notes:
  - contain the results of coding
  - primarily describe conceptual labels and dimensions
- Theoretical notes:
  - contain the results of inductive or deductive thinking
  - regard categories, properties/dimensions, relationships, variations, processes, and the conditional matrix.
- Operational notes:
  - contain directions to yourself and team members
  - regard sampling, questions, possible comparisons, and potential dimensions/relationships/variations/processes, etc.
- All memos should talk on the conceptual level

Types of diagrams

- Diagrams are usually semantic networks
  - Nodes: concepts, categories, or individual phenomena
  - Edges: relationships, typically according to the coding paradigm
    - is-causal-condition-of
    - intervenes-in (possibly different kinds of intervention)
    - is-consequence-of
    - is-strategy-for
    - is-context-for
    - other kinds of relationship may also be helpful

Elements of the Grounded Theory method

- Coding:
  - open coding (What?)
  - constant comparison (Grounding)
  - axial coding (When, where, why, how, what follows?)
  - selective coding (Telling the story)
- Theoretical sensitivity
- Data gathering:
  - everything is data
  - theoretical sampling
- Reflection:
  - asking questions to enhance theorectical sensitivity
  - writing memos and diagrams

GT work as teamwork

A GT study can be performed by a team rather than a single researcher

- in fact that may help defuse subjective bias
- and imcrease the overall theoretical sensitivity

- Data gathering can be done by individual members
- Analysis can, in principle, also be done by individual members
- The big problem is maintaining a shared conceptual understanding:
  - Analysis results have to be shared fully and thoroughly
    - Ideally, all results are tested/validated by the other members
  - Decisions on theoretical sampling must be made together

Quality criteria for GT studies

- Canons of "good science" apply to qualitative studies just like they do to quantitative studies
- However, they are not the same canons
  - They are structurally similar, but most individual criteria need to be reformulated.
- Basically, criteria of good science are:
  - significance, relevance
  - theory/observation compatibility
  - **generalizability**
  - consistency
  - **reproducibility**
  - precision
  - verification

Criteria: Generalizability

- By design, GTs are (and want to be) very specific
  - This is required since they attempt to characterize extremely complex phenomena
- Generalizability, thus, cannot mean a good GT has to apply to a wide variety of situations

- Rather, the GT should apply whenever the conditions described by the GT hold
  - Even then differences are possible, because the situation may involve additional conditions not considered in the GT

Criteria: Reproducibility

- By the nature of social phenomena, it is impossible to reproduce exactly equivalent data gathering
- By the subjective and creative nature of the GT analysis process, it is also impossible to exactly reproduce the analysis, given the data
- Still, a GT study should be reproducible in these senses:
  - Given the same study focus but gathering new data, a new researcher should arrive at **similar findings** insofar as the same conditions are present in the new data.
  - Given the same data and study focus, a second researcher should arrive at **findings that are compatible** with those described in the GT
  - Given the same data and a description of the analysis process, a second researcher should **agree with the analysis**.

Criteria: Describing the research process

To be credible, a report on a GT study should answer the following questions:

- How was the original sample selected?
  - Why?
- What major categories emerged?
- Which events pointed to some of these categories?
- How did theory development drive further sampling?
- Which hypotheses on category relationships were tested?
  - How? How were they formed?
- Did discrepancies occur?
  - How were they handled?
- How and why was the core category selected?

Criteria for judging empirical grounding

Many studies that claim to use GT, don't.

The following questions can be used to recognize them:

- Are concepts generated?
  - And are they technical, rather than common sense?
- Are the concepts systematically related?
  - Are the linkages grounded in the data?
  - Systematically carried out?
- Do the categories have conceptual density?
  - Is much variation built into the theory?
  - Do the categories have good explanatory power?
- Are the broader conditions (→conditional matrix) affecting the phenomenon considered in the explanation?

Criteria for judging empirical grounding (2)

- Has process been taken into account?
  - And is change linked to conditions giving rise to it?

And furthermore:

- Do the findings seem significant?

Historical note

- Note that GT was invented together by Barney Glaser and Anselm Strauss (1967)
- These two, however, come from rather different background in terms of the philosophy of science they subscribe to
  - In short: Glaser comes from an objectivist background; Strauss from a pragmatist background.
- This has led to two versions of GT, one propagated by each author.
  - They have many commonalities, but also significant differences.
  - In short: Glaser focuses on creativity; Strauss on rigorosity
- The Strauss variant of GT is generally considered the more scientifically sound one.

Summary: GT in a nutshell

1. Data collection and analysis are interrelated processes
2. Concepts are the basic units of analysis
3. Categories must be developed and related
4. Sampling proceeds on theoretical grounds
5. Analysis makes use of constant comparisons
6. Patterns and variations must be accounted for
7. Process must be built into the theory
8. Writing theoretical memos is integral part of doing GT
9. Hypotheses about relationships among categories should be developed and verified as much as possible
10. A grounded theorist need not work alone

(according to *"Grounded theory research: Procedures, canons, and evaluative criteria"*, Strauss/Corbin, 1990)

References

- Anselm Strauss, Juliet Corbin: *"Basics of qualitative research: Grounded Theory procedures and techniques"*, Sage 1990.
  - German: *"Grundlagen qualitativer Forschung"*, Beltz

- http://en.wikipedia.org/wiki/Grounded_Theory_(Strauss)

Software supporting GT work

- A number of software packages exist that support qualitative data analysis, e.g.
  - ATLAS.ti 5 www.atlasti.de
  - Kwalitan 5.09 www.kwalitan.net
  - MaxQDA 2 www.maxqda.de
  - N6 (NUD*IST 6) www.qsr.com.au
  - NVivo 2 www.qsr.com.au
  - QDA Miner 1.0.15 www.simstat.com
- They have different strengths and weaknesses
- The most suitable for our purposes is clearly **ATLAS.ti**
  - supports quoting multimedia data (pictures, audio, video)
  - can represent arbitrary semantic networks
  - supports teamwork (file merge)

ATLAS.ti 5 basic concepts

- Raw material resides in native files
  - text (RTF and everything convertible to RTF)
  - pictures, audio, video (all file formats supported by Windows MCI)
- An analysis project resides in a **hermeneutic unit** (HU)
- A **primary document (PD)** has a name and references a raw data file (called a **data source**)
- A **quotation** is a contiguous block in a PD
- A **code** is a non-unique label assigned to a quotation
- Arbitrary text can be written into a **memo**
  - and memos can themselves be used as PDs if needed

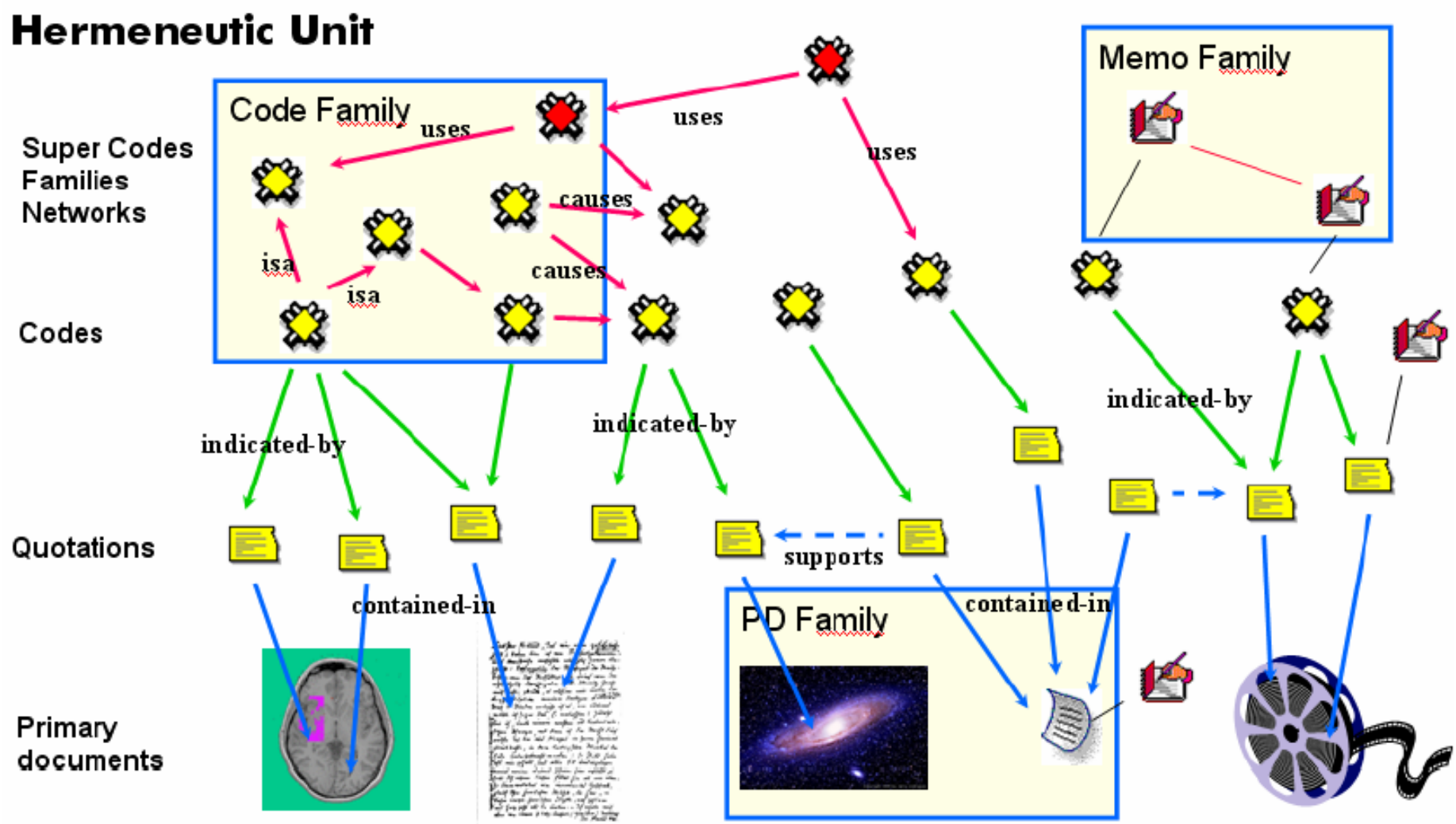
ATLAS.ti 5 basic concepts (2)

- **Links** are hyperlinks from one object to another
  - Start/Endpoint: Any PD, quotation, code, memo, family, superfamily, or network view
  - *Strong link*: typed, named, pre-declared intra-type relation that can be marked as symmetric, asymmetric, transitive.
  - *Weak link*: "pertains-to"; used between objects of different type (extra-type, e.g. code and quotation)
  - Hyperlinks are available for navigation everywhere in GUI
- **Network views** are directed graphs and represent semantic networks based on objects and hyperlinks

ATLAS.ti 5 basic concepts (3)

- Codes can be arranged arbitrarily into **families**
  - as can PDs or memos
  - any one can be a member of many families
- **Super families** are boolean combinations of families
  - Families and super families make good queries for searching and retrieving data
- How to attach properties to quotations:
  - Direct method: create a separate code for each possible value, attach those directly
    - Form one family for each property to collect the values
  - Family method: create one auxiliary code for each quotation, collect them into one family for each value
- All of these object types have attributes
  - Generic attributes: created, modified, author, comment

An ATLAS.ti network view



ATLAS.ti: Working in teams

- The same data source (raw file) can be attached to one or more PDs, in one or more HUs
- This produces two possible problems:
  - If a HU or a data source is moved, they lose connection
    - Repair: use PD paths relative to HU (variable HUPATH) or relative to a fixed data directory (variable TBPATH, "textbase")
  - If a data source is modified from within ATLAS.ti (which is possible for RTF data sources), all its quotations must be updated – even in different HUs
    - ATLAS.ti does this automatically by means of a log file (one per data source) if data sources are never changed "by hand"
    - or semi-automatically ("copy bundle", "install bundle") if copies of data source files are involved
- This way, a team may produce several independent analyses pertaining to the same raw data files

ATLAS.ti: Working in teams (2)

A HU (the *source*) can be merged into another (*target*):

- For each object type…
  - PD, code, memo, network view
- …select a merge mode:
  - Add: add each object to target (modifying the name by appending "_1" etc. on clashes)
    - There can only be one link between any two codes
  - Unify: Like add, but ignore objects with the same name
    - PDs: same number; Quotations: same coordinates.
    - Data source modifications are synchronized for PDs
  - Ignore: Do not transfer these objects
    - Members of families "!MERGEIGNORE" are always ignored
- This way, a team may cooperatively work on a common analysis

ATLAS.ti: A warning about merge

Note that this merging model is different from that used by, say, CVS:

- It is a two-way merge, rather than three-way
  - no common base revision
  - this is intentional: merging may occur unplanned
- Consequence: Deletions or modifications of objects cannot be detected; merging is purely additive!
  - "unify" does not really do anything
  - This requires thought regarding the team cooperation style. Preferable modes are (one of):
    - Each member works on different PDs
    - Each member works with different families of codes
    - Member B reviews work of A, but changes are then made by A
    - Any mode that carefully avoids overlap (e.g. time-slice 1 HU)

ATLAS.ti user and rights management

- Each ATLAS.ti installation maintains its own user database (name/password pairs)
  - By default there is only one user *"Super"* and login is not required (implicit, no password)
- Each HU can specify which users have which rights:
  - read/write, read only, none
- Logging in as a specific user provides useful values for the 'author' attribute of all objects created.
  - Can also be used for querying/filtering
  - Useful for teamwork!
- The HU can further be protected by a HU password

Summary

- Grounded Theory is the *constant comparative method* of qualitative research
- It leads to theories that are fully grounded in data
- Its main prerequisite is *theoretical sensitivity*
- Its main processes are
  - *open coding, axial coding, selective coding*
  - writing *memos*
  - *theoretical sampling*
- It can be supported by appropriate software
  - such as *ATLAS.ti*

Thank you!