GROUNDED THEORY REVIEW
An international journal

Theoretical Coding in Grounded Theory Methodology

Cheri Ann Hernandez, RN, Ph.D., CDE

November 2009

Grounded Theory Review, Vol 8 (Issue #3), 51-66

The online version of this article can be found at:

https://groundedtheoryreview.org

Originally published by Sociology Press

https://sociologypress.com/

Archived by the Institute for Research and Theory Methodologies

https://www.mentoringresearchers.org/

The Grounded Theory Review (2009), vol.8, no.3

Theoretical Coding in Grounded Theory Methodology

Cheri Ann Hernandez, RN, Ph.D., CDE

Abstract

When doing classic grounded theory research, one of the most problematic areas, particularly for novice researchers, is the theoretical coding process. The identification of theoretical codes is essential to development of an integrated and explanatory substantive theory when a researcher is using classic grounded theory research methodology, but it is not a part of Straussian qualitative data analysis as described by Strauss and Corbin. A theoretical code is the relational model through which all substantive codes/categories are related to the core category. Like substantive codes, theoretical codes emerge through the data analysis process, rather than being overlaid on the data through the use of conjecture or 'pet' codes. The purpose of this article is to provide an overview of the theoretical coding process and to review the theoretical coding families and individual theoretical codes that have been identified previously by Glaser.

Introduction

Grounded theory (GT) is a research methodology for discovering theory in a substantive area. In many of his publications, Glaser (1978, 1992, 1998, 2001, 2003, 2005) has carefully delineated the various aspects of GT research methodology, and has consistently elucidated areas that have been difficult for published GT researchers, often illustrating the erroneous assumptions or methodological errors found in such research (Hernandez, 2008). One of the most problematic areas, particularly for novice researchers, is the theoretical coding process which includes finding the theoretical code that will integrate the emerging substantive theory. Perhaps one of the reasons for this confusion is that many researchers have not understood that classic (also known as Glaserian) GT and Straussian GT are two very different methods (Hernandez, p. 44) and, as a result, many research articles list references from both Glaser and Strauss as the methodological underpinning of their

studies. However, theoretical coding as described by Glaser (1978) is not a part of Strauss' approach to grounded theory data analysis (Strauss & Corbin, 1998).

The purpose of classic GT research is to uncover the main problem in a substantive area, as well as the resolution to this problem. The resolution is known as the core category. The final theoretical code is the one that emerges, through the coding process, and serves to integrate all of the substantive categories with the core category. The approach to data in classic GT methodology consists of two main processes. First, during the *open coding process*, the data are broken down into *substantive codes* (either *in vivo codes* or *sociological constructs*) as interview, field notes and/or other written data are coded in a line by line manner and incidents are compared with one another, for similarities and differences (Glaser, 1978) until the core category is found. Then, as *selective coding* results in the saturation of all of the categories through theoretical sampling, these substantive codes are built up into a substantive theory as they are integrated into a cohesive structure by the emergent *theoretical code.* The purpose of this article is to provide an overview of the theoretical coding process and review the theoretical coding families and individual theoretical codes that have been identified previously by Glaser (1978, 1998, 2005) as being relevant for grounded theory research.

Understanding Theoretical Codes in Classic GT

In any GT study, several theoretical codes may emerge but eventually, through ongoing coding and memoing, one theoretical code is chosen as the theoretical code for the study. A GT study's theoretical code is the relational model through which all substantive codes/categories are related to the core category. In GT methodology, "Substantive codes conceptualize the empirical substance of the area of research. Theoretical codes conceptualize how the substantive codes may relate to each other as hypotheses to be integrated into the theory" (Glaser, 1978, p. 55). Substantive codes break down (fracture the data) while theoretical codes "weave the fractured story back together again" (Glaser, 1978, p. 72) into "an organized whole theory (Glaser, 1998, p. 163). The relationship, therefore, between substantive and theoretical codes is that theoretical codes "theoretically render an empirical pattern" (Glaser, 1978 p. 74). Another way of saying this is that "Theoretical codes implicitly conceptualize how the substantive

The Grounded Theory Review (2009), vol.8, no.3

codes will relate to each other as interrelated multivariate hypotheses in accounting for resolving the main concern" (Glaser, 1998, p. 163). Theoretical codes must not be preconceived, rather they are emergent in the data, and therefore, "earn their way into the theory as much as substantive codes" (Glaser, 1998, p. 164).

Coding processes for substantive codes and theoretical codes are not two isolated or disconnected processes. Both types of coding occur simultaneously, to a certain extent, but the researcher "will focus relatively more on substantive coding when discovering codes within the data, and more on theoretical coding when theoretically sorting and integrating his memos" (Glaser, 1978, p. 56). Without substantive codes, theoretical codes are empty abstractions (Glaser, p. 72). The importance of the substantive codes cannot be over-emphasized. If the substantive codes do not fit the data, then the theoretical codes that relate these substantive codes are probably irrelevant to the substantive area: The researcher has only a contrived theory that is not grounded in the data.

Theoretical codes are either implicit or explicit but, whether implicit or explicit, their purpose is to integrate the substantive theory (Glaser, 2005, p. 11). Theoretical codes from the Process Family are often explicit and easily identified by researchers when study participants talk about changing over time or about going through stages, phases or transitions. However, other theoretical codes are more implicit. These more implicit theoretical codes can be uncovered as a theoretically sensitive researcher continues coding and memoing, or through observing participants act in ways that are contrary to what they have espoused in interviews. This latter example would imply that vaguing or properlining (from the Cultural Representation Family) is occurring.

Theoretical codes are flexible – "they are not mutually exclusive, they overlap considerably... [and] one family can spawn another" (Glaser, 1978, p. 73). The overlap in theoretical codes can be seen in Table 1 by comparing the individual theoretical codes within the coding families that have been placed next to each other. For example, there is overlap between the Process and Basics coding families, with the basic processes frequently having stages, phases, transitions, sequencing and so on, all of which are theoretical codes found under the Process Family.

Over the past three decades, Glaser has identified many theoretical codes and theoretical coding families that can emerge in grounded theory: 18 in *Theoretical Sensitivity* (Glaser, 1978), 9 in *Doing Grounded Theory* (Glaser, 1998), and 23 in *Theoretical Coding* (Glaser, 2005). See Table 1 for a summary of these theoretical codes. This table has been organized so that the theoretical coding families and codes, identified by Glaser in three of his books, have been positioned next to the coding families to which they are closely related or a part of. However, Glaser has been adamant that there are potentially many more theoretical codes that might emerge in GT research; therefore, the theoretical codes found in Table 1 do not comprise an exhaustive list.

Researchers learning to do grounded theory need to be aware that seasoned GT researchers may speak about theoretical coding (a *verb* denoting the process of finding theoretical codes through emergence) as the process they use to find a theoretical code (a *noun* denoting the actual type of relationship between two or more substantive codes or between the core category and all other substantive codes). Theoretical coding can occur throughout the GT process, whether it is during open coding or selective coding (the two major phases of the GT methodology) because theoretical coding is simply detecting the relationships between two or more categories. Several theoretical codes can be discovered as coding proceeds during one GT study. However, discovery of the ultimate theoretical code that integrates the substantive theory will probably occur during the selective coding phase, that is, after the core category has emerged.

As previously stated, in any GT study there can be several emergent theoretical codes because a theoretical code simply specifies the relationship between two or more substantive codes. Theoretical codes from several theoretical coding families may emerge as being relevant in specifying the emergent relationship between categories (known as major categories, codes, or variables) and subcategories (known as smaller categories, codes, or variables), and even between the core category and the subcore (major) categories and their properties. However, the theoretical code that ultimately emerges as the one that most fully integrates the substantive theory is one that specifies the overall relationship between the core category and all other categories. When more than one theoretical code can fit the data, then the

researcher must make a choice but this decision will be "grounded in one of the many useful fits" (Glaser, 1978, p. 72). The following example will illustrate this point. Hernandez (1991, 1996) discovered the substantive theory of integration in her research with adults with Type 1 diabetes. Integration was the core category to which all other substantive codes were related through a *basic social process* (a theoretical code from the Basics Family). However, the first phase of the theory of integration was named "having diabetes" (major category) and the smaller categories related to "having diabetes" as *strategies* (theoretical code from the Strategy Family) which helped to prevent the person who had diabetes from moving into the second phase, "the turning point" (major category). In addition, it was observed that as participants with diabetes moved through the three phases of integration (having diabetes, turning point, science of one) there was an increase in the level (theoretical code from the Degree Family) of integration. In the end, a basic social process emerged as the final (overall) theoretical code for the substantive theory of integration because of its fit (i.e., it was able to show the relationship of all of the categories to the core category of integration) and thus provided the best overall fit for the data. For example, it was discovered that an individual with diabetes could remain in the turning point phase (second phase) for a period of time but later revert back to the having diabetes phase and this represented the best fit with the *basic social process* theoretical code rather than the *degree* theoretical code.

A major characteristic of the theoretical code for a GT study is that it must be emergent through the data, not preconceived (or overlaid on the data) by the researcher. Unfortunately, many researchers have a 'pet' theoretical code that they apply to all data, rather than remaining open and waiting for emergence. When viewing research data through the blinders of a pet category, there is a danger of systematically ignoring important data that are relevant to the substantive theory but do not fit with this pet code. Emergence is always better than conjecture (Glaser, 2005, p. 42), therefore theory generated through 'pet code overlay' may not be one that adequately explains the resolution of the problem experienced by participants in the substantive area.

Theoretical codes are important to grounded theory because they potentiate its explanatory power and increase its completeness and relevance, resulting in a grounded theory with

greater scope and parsimony (Glaser, 2005, p. 70). Without theoretical codes, the substantive codes become mere themes to describe (rather than explain) a substantive area; the descriptive thematic approach is characteristic of qualitative research methods such as phenomenology or ethnography but not Classic GT.

Ways to Enhance Researcher Ability to 'See' the Emergence of Theoretical Codes

Some researchers mistakenly believe that core categories generate theoretical codes (Glaser, 2001, p. 210). They do not. Theoretical codes emerge from the data as a theoretically sensitive researcher analyzes the data, through coding, memoing and sorting the memos, or possibly through developing a schematic model (conceptual map) of the substantive codes. Several strategies for eliciting theoretical codes are described in the section below.

1. Theoretical Sensitivity. The researcher's theoretical sensitivity enhances his or her ability to recognize the theoretical codes as they emerge during coding and memoing. Knowledge of the various theoretical coding families will help to sensitize researchers (Glaser, 1998, p. 175), making the researcher "sensitive to rendering explicitly the subtleties of the relationships in his data…It sensitizes him to the myriad of implicit integrative possibilities in the data" (Glaser, 1978, pp. 72 & 73). Therefore, "the goal of a GT researcher is to develop a repertoire of as many theoretical codes as possible…the more theoretical codes the researcher learns the more he has the variability of seeing them emerge and fitting them to the theory. They empower his ability to generate theory and keep its conceptual level" (Glaser, 2005, p. 11). Researchers are encouraged to read literature in any field to learn about other theoretical codes (Glaser, 2005, p. 42). In this way, researchers build an understanding and repertoire of many potential theoretical codes; this will allow emergence of the theoretical codes rather than always reverting to a cherished 'pet' code that a researcher forces or overlays on the data. Researchers are advised to be familiar with the theoretical codes in Table 1 so that they can recognize them when they see them in the data they are coding.

2. In Vivo Codes. An in vivo code is one of the two types of

substantive codes that emerge as data are coded during the open coding process, and these in vivo codes can point to possible theoretical codes. In vivo codes "tend to be the behaviors or processes which explain how the basic problem is resolved or processed" (Glaser, 1978, p. 70) and, therefore, "can imply theoretical codes; for example, cultivating implies looking into consequences since anticipating consequences [a theoretical code] is why people cultivate" (Glaser, 1978, p. 70).

3. Memoing and Sorting Memos. Writing memos will force researchers to theoretically code (Glaser, 1978, p. 85) to determine how a particular category is related to other categories that have been discovered already. Researchers' ideas that are developed through memoing include "hypotheses about connections between categories and/or their properties" (Glaser, 1978, p. 84) and thus begin "to integrate these connections with clusters of other categories to generate the theory" (Glaser, 1978, p 84). In other words, memos bring out the relationships (i.e., the theoretical codes) among the various categories and their properties. "Memos serve as a means of revealing and relating by theoretically coding the properties of the substantive codes" (Glaser, 1978, p. 84). The memoing process helps the researcher determine which of the theoretical codes provides the best relational model to integrate the substantive theory because it is during memoing that different emerging theoretical codes are discussed and tried out as possible ways of organizing the grounded theory (Glaser, 2003, p. 31).

The major process through which a grounded theory is written up, is through sorting of the memos that have been written throughout the study process. During sorting, the researcher places each memo onto the pile to which it belongs, based on the substantive code (s) to which it refers. According to Glaser (2005), about 90% of the theoretical codes found in a study are identified through the sorting of mature memos (p. 42).

4. Models. Glaser (1978) identified the development of a model as one way to theoretically code; using this method, the researcher models the "theory pictorially by either a linear model or a property space" (p. 81). The researcher writes the substantive concepts (codes) on a piece of paper in circles or squares and draws solid or broken lines between them to demonstrate the relationships between and among all of the concepts. However, Glaser recommended that these models be

used with constraint and caution: researchers might be tempted to *deduce* relationships through logical elaboration, rather than eliciting them from the data by emergence (*induction*). This error may derail the emergence of a good substantive theory because deduced relationships may not be relevant (Glaser, p. 82).

Researcher Uses of Theoretical Codes

Glaser (1978) identified four general uses of theoretical codes. The two major uses will help researchers integrate and write-up their substantive theories. The last two purposes are for critiquing GT studies and for grant writing. These four uses specified by Glaser are: 1) helping the researcher maintain a conceptual level when writing about concepts and the relationships among them; 2) preventing researchers from getting bogged down in the data through endless illustrations; 3) critiquing other researchers' grounded theory reports; and 4) when writing a grant proposal that forces the researcher to preconceive possibilities prior to the start of the research and, therefore, before the researcher knows anything about the data to be collected (Glaser, p. 73). An important dictum when talking about a GT or writing it up, is to talk or write substantive codes but think theoretical codes (Glaser, 1998, p. 164). The theory of integration (Hernandez, 1991, 1996) can be used to illustrate this dictum. Whenever the author writes about the theory of integration, she writes about the substantive codes within each of the three phases. Therefore, she acknowledges that there are three phases (theoretical code of basic social process forms the Basics coding family) but the focus of the write-up is on the explanation of the substantive codes within these phases.

Conclusion

The identification of theoretical codes is essential to development of an integrated and explanatory substantive GT. The theoretical code that emerges to integrate the substantive theory is not, itself, the core category; rather it is the conceptual model of the relationship of the core category to its properties and to the other (non-core) categories. It is this relational model that integrates the substantive categories into a theory. Preconception, through conjecture or overlay of pet theoretical codes, will derail the emergence of a credible substantive grounded theory. Just as theoretically sensitive GT researchers are able to recognize sociological constructs in the data, so to will

these researchers be able to detect the emergent theoretical codes as they follow GT methodology and when they have built up a repertoire of relevant theoretical codes. Although, several theoretical codes may emerge in any one GT study, the theoretical code that is most relevant will be the one that captures the relationships between all essential categories and the core category (i.e., provides the best fit for the data).

Author

Cheri Ann Hernandez, RN, Ph.D., CDE
Associate Professor
Faculty of Nursing
University of Windsor, ON
Canada
Email: **cherih@uwindsor.ca**

References

Glaser, B. G. (1978). *Theoretical sensitivity*. Mill Valley, CA: Sociology Press.

Glaser, B. G. (1992). *Emerging vs. forcing: Basics of Grounded Theory analysis.* Mill Valley, CA: Sociology Press.

Glaser, B. G. (1998). *Doing grounded theory: Issues and discussions.* Mill Valley, CA: Sociology Press.

Glaser, B. G. (2001). *The grounded theory perspective: Conceptualization contrasted with description*. Mill Valley, CA: Sociology Press.

Glaser, B. G. (2003). *The grounded theory perspective II: Description's remodeling of Grounded Theory methodology.* Mill Valley, CA: Sociology Press.

Glaser, B. G. (2005). *The grounded theory perspective III: Theoretical coding*. Mill Valley, CA: Sociology Press.

Hernandez, C. A. (1991). The *lived experience of Type 1 diabetes: Implications for diabetes education.* Unpublished dissertation, University of Toronto, Toronto, Ontario.

Hernandez, C. A. (1996). Integration: The experience of living with insulin dependent (Type 1) diabetes mellitus. *Canadian Journal of Nursing Research, 28*(4), 37-56.

Hernandez, C. A. (2008). Are there two methods of grounded theory? Demystifying the methodological debate. *The Grounded Theory Review, 7*(2), 39-66.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory (*2nd Ed.). Thousand Oaks, CA: Sage.

The Grounded Theory Review (2009), vol.8, no.3

Theoretical Coding in Grounded Theory Methodology

Theoretical Coding Tables

Table 1 Theoretical Coding: Mapping of Coding Families & Theoretical Codes (Glaser, 1978, 1998, 2005)

1978	1998	2005
The 6 Cs Causes Contingencies Consequences Covariance Contexts Conditions		**Causal Family** Conjunctural causation TC Amplifying causal looping Perpetual compounding Bias random walk Equifinality Reciprocal Causation Social Arena TC, (Social Worlds TC, Social Constraints TC Social constraints TC (boundary maintaining conditions)
Process Family Stages, Staging, Phases, Phasing Progressions, Passages, Transitions, Trajectories Gradations, Steps, Shaping Ranks, Ordering, Chains Sequencing, Temporaling, Cycling	**The Basics Family** Basic social process Basic social structural process Basic social structural conditions Basic social psychological process Basic psychological process	**Temporal Family** Timelines Pacing of action (self, others) Pacing growing conditions Cycling TCs: Temporal ordering of work, etc. Organizational cycling of productions Actions/interaction spiraling up or down
Degree Family Ranks, Grades, Continuum, Levels Limit, Range, Intensity, Extent, Amount	**Average Family** Mathematical actions (mean, median, mode)	Levels TC (e.g., social stratification, authority structures, spirituality)

1978	1998	2005
Cutting points, Critical juncture Statistical averages, Standard deviation Probability, Possibility, Polarity	Cutting points, Confidence limits Structural norms Probabilities of Occurrence	
Dimension Family Dimensions, Elements, Divisions, Piece of, Properties of, Facet Sector, Portion, Segment, Part, Aspect, Section		
Type Family Type, Form, Kinds Styles, Classes, Genre		Ideal Type TC (Constructed Types, Paradigms, Pattern Variables)
Strategy Family Strategies, Tactics, Mechanisms, Ways, Techniques Manipulation, Manoeuvrings, Ploys, Dominating Dealing with, Handling, Arrangements, Managed Goals, Means, Positioning		Cross Pressures TC (External Conflict) Moment capture TC (essential opportunistic action)
Interactive Family Mutual efforts, Reciprocity, Mutual trajectory Mutual dependency, Interdependence Interaction of effects, Covariance	**Interactive Family** Traffic interaction Asymmetrical interaction (off-balance power equilibriums or power relations)	Systems Parts TC (e.g., functional reciprocity)
Identity-Self Family Self-image, Self-concept, Self-worth, Self-evaluation Identity, Social worth Self-realization, Transformations of self, Conversions of identity	**Unit Identity Family** Work, family, recreational life	**Unit Identify Family** Binary code TC (binary retreat, binary deconstruction)

1978	1998	2005
Cutting Point Family Boundary, Benchmark, Division Critical juncture, Cutting point, Turning point, Breaking point Cleavage, Tolerance levels, Deviance Dichotomy, Trichotomy, Psychotomy In-out, Intra-extra, Point of no return	**Boundary Family** Limits, Outer limits, Confidence Limits, Front line, Deviance Boundary maintaining mechanisms Tolerance zones, Transition zones	Outer limits TC (e.g. legal, moral, security, safety, breaking or moving boundaries)
Means-Goal Family End, Purpose, Goal Anticipated consequences, Product		Asymptote TC (getting as close as possible) Fractals TC Autopoiesis TC (e.g. structural coupling)
Cultural Family Social norms, Social values, Social beliefs Social sentiment	**Representation Family** (Cultural Representation) Descriptive (e.g. conceptualizing, baselining, vaguing) Proscriptive (e.g., properlining) Prescriptive (e.g., interpreting) Sentimental, Evaluative	Frames TC (Political, Religious, Ideological, Cultural)
Consensus Family Clusters, Agreement, Contracts, Cooperation Definitions of the situation Uniformities, Opinion, Homogeneity-Heterogeneity, Conformity, Conflict, Dicensus, Differential Perception, Non- Conformity Mutual Expectation		

1978	1998	2005
Mainline Family Social control, Socialization, Social organization Social order, Social Mobility, Status Passage Recruitment, Stratification Social institutions, Social Interaction, Social World		
Theoretical Family Parsimony, Scope, Integration, Density, Clarity Fit, Relevance, Modifiability, Utility, Condensability Conceptual level, Inductive-Deductive Balance Multivariate Structure, Degree of Interpretive, Explanatory & Predictive Power		
Ordering or Elaboration Family Structural ordering Temporal ordering Generality (Conceptual ordering)		
Unit Family (Structural Units) Collective, Group, Organization, Aggregate Nation, Social world, Territorial units, Society Situation, Context, Arena Behavioral pattern, Family Positional units (status, role, role relationship, status set, role-set, person-set, role partners)	**Structural-Functional Family** Reference groups Structural change Authority structure Functional requirements of structure	Action TC (variation in types of action for each unit) Social constraints TC
Reading Family Concepts Problems Hypotheses		

1978	1998	2005
Models Linear model Property space		Isomorphism TC (theory model)
	Paired Opposite Family Dichotomies, Polar opposites Paired Alternatives	**Paired Opposite Family** Symmetry-asymmetry TC Binary TC Micro macro TC Balancing TC
	Scales Family Likert Guttman Cumulative Funneling down, Delimiting Random walk	

TC= Theoretical Code