The Development of Categories: Different Approaches in Grounded Theory

Udo Kelle

INTRODUCTION

One of most crucial and fascinating ideas in The Discovery of Grounded Theory was that it would provide a methodological groundwork for directly deriving categories from data of social research. Thereby, grounded theory was meant to represent an alternative to the classical hypothetico-deductive approach which requires the construction of clear-cut categories and hypotheses before data are collected. However, the development of methodological guidelines for empirically grounded category building turned out to be much more challenging and difficult than initially thought. In The Discovery of Grounded Theory, the metaphor of ‘emergence’ was invented which had a far-reaching impact on the methodological debate but, at the same time, was difficult to be translated into tangible methodological rules. Glaser and Strauss’s initial idea that categories would emerge from the data if researchers with sufficient theoretical sensitivity would apply a technique of constant comparison was difficult to realize in practice. Consequently, this idea was modified and refined several times in the ongoing development of grounded theory leading to a variety of different, new, and complex concepts like theoretical coding, coding families, axial coding,
coding paradigm, and many others, which supplemented and sometimes displaced the concepts of constant comparison and theoretical sampling from the early days.

The basic concept of category building through theoretically sensitive constant comparisons will be outlined in the first section of this chapter. It will be shown that the major problem of those concepts lies in the failure to explicitly conceptualize the role of previous theoretical knowledge in developing grounded categories. In the second section, I will discuss the progress which was made in this respect through the different approaches which Glaser and Strauss developed after they had finished their methodological cooperation in the late 1970s. I will focus on the most prominent differences between the Glaserian and the Straussian approach, comparing Glaser’s notions of theoretical coding and coding families with the concept of coding paradigm put forward by Strauss and Corbin. The pros and cons of both modes of category building will be treated, and under which conditions and for which research questions these different approaches are best suited will be discussed.

The most basic challenge in grounded category building is to reconcile the need of letting categories emerge from the material of research (instead of forcing preconceived theoretical terms on the data) with the impossibility of abandoning previous theoretical knowledge. In the third section of the chapter, I will show how classical methodological concepts (especially the concept of empirical content) can be employed to distinguish between theoretical notions that force the data and concepts that support the emergence of new categories.

**CATEGORY BUILDING THROUGH CONSTANT COMPARISON: THE BASIC RULES FROM THE DISCOVERY OF GROUNDED THEORY**

How can theoretical categories be developed in the ongoing process of empirical research? The main purpose of Glaser and Strauss’s first methodological book *The Discovery of Grounded Theory* was to show that empirical data can play a crucial role in that process: the book was written in order to give examples and rules for category building with the help of empirical data. Both authors wanted to provide an alternative to the hypothetico-deductive approach in sociology which demands that precise hypotheses are developed before data are collected. Consequently, Glaser and Strauss started *The Discovery of Grounded Theory* by criticizing the ‘overemphasis in current sociology on the verification of theory, and a resultant de-emphasis on the prior step of discovering what concepts and hypotheses are relevant for the area that one wishes to research’ (Glaser & Strauss, 1967: 1f). Contrary to the idea that the main purpose of empirical research is the testing of explicit theoretical assumptions, they proposed a method for the ‘initial, systematic discovery of the theory from the data of social research’ (p. 3) that would lead to the development of categories grounded in the data. But how can the grounding of categories be assured?
In *The Discovery of Grounded Theory*, the two most basic rules of category building are given which still form the backbone of category building in grounded theory:

- Categories must not be *forced* on the data, they should *emerge* instead in the ongoing process of data analysis.
- In developing categories, the sociologist should employ *theoretical sensitivity*, which means the ability to see relevant data and to reflect upon empirical data material with the help of theoretical terms.

### The emergence of categories and their properties from the data

The most basic operations which provide the basis for category building are *coding* and the *constant comparison* of data, codes, and the emerging categories. Most interestingly, the term coding stems from the quantitative tradition of social research: there it means that predefined codes are used to qualify certain bits of data. For this purpose, units of analysis (e.g. paragraphs of a certain newspaper article in a research project about mass media, or an answer to an open ended question in a survey) have to be determined, and a precise coding scheme has to be constructed before the analysis. In this case, each code represents a value of a certain variable (for example, the value affirmation or disapproval of a variable called ‘Evaluation of political events’). With the help of such a coding scheme, every unit of analysis can be investigated in order to find out whether a certain value of a variable applies to it; one may, for example, analyze paragraphs of newspaper articles in order to find out whether the authors express affirmation or disapproval for certain actions taken by the government. The purpose of this process is to count numbers of codes once the coding process is finished for all relevant data. Coding of that kind is always part of a hypothetico-deductive strategy and requires that the full coding scheme be developed before data are coded.

In a publication in 1960, Howard Becker and Blanche Geer adopted the term coding for the qualitative research tradition where it meant relating text segments in field protocols to certain predefined codes. The main purpose of qualitative coding *sensu* Becker and Geer was not to count codes but to be able to find all the different text passages which (in the researcher’s opinion) refer to a certain topic:

We have tentatively identified, through sequential analysis during the field work, the major perspectives we want to present and the areas ... to which these perspectives apply. We now go through the summarized incidents, marking each incident with a number or numbers that stand for the various areas to which it appears to be relevant. This is essentially a coding operation ... its object is to make sure that all relevant data can be brought to bear on a point (Becker & Geer, 1960: 280f).

Since an important aim of this process was to ‘constitute proof for a given proposition,’ this procedure still showed a certain proximity to a classical hypothetico-deductive approach. In their monograph *The Discovery of Grounded Theory*,...
Glaser and Strauss distanced themselves from that approach by maintaining that their method was not meant as a technique for the (provisional) testing of hypotheses, but ‘is concerned with generating and plausibly suggesting many categories, properties and hypotheses about general problems’ (Glaser & Strauss, 1967: 104). For that reason, coding in grounded theory had to be conducted without a predefined coding scheme. Categories should ‘emerge’ from the data if the analyst ‘starts by coding each incident in his data into as many categories as possible’ (p. 105). This emergence of categories should be supported by the ‘constant comparative method’: while coding, the analyst constantly compares the already coded incidents (which usually means the text segments which relate to the incidents) with each other and with incidents not yet coded. ‘This constant comparison of the incidents very soon starts to generate theoretical properties of the category’ (p. 106).

From the early days of grounded theory, many users of the method found it difficult to understand the notions ‘category’ and ‘property’ and to utilize them in research practice, since these terms were only vaguely defined in The Discovery of Grounded Theory. Glaser and Strauss gave the following example: In their own research project about the interaction between nurses and moribund patients in hospitals they established the category ‘social loss’; nurses tend to think about their dying patients in terms of the loss which their death would mean for their social environment (e.g. ‘What will the children and the husband do without her,’ p. 106). By constantly comparing incidents relevant for (and coded with) that category, they found various ‘theoretical properties’ of the category, for instance ‘... we realized that some patients are perceived as a high social loss and others as a low social loss, and that patient care tends to vary positively with degree of social loss’ (p. 106). Whereas the notion ‘category’ can refer to any noun which the researcher found relevant for their research area, the notion ‘property’ is more difficult to grasp: are the concepts ‘high social loss’ and ‘patient care’ theoretical properties of the category ‘social loss,’ and what does that mean?

It is helpful here to draw on basic mathematical set or type theory to gain a better understanding of the relation between incidents, categories, and their properties: *objects* (i.e. *incidents* like text passages describing utterances of nurses about their patients) can be assigned to a certain class or type or category (e.g. ‘utterances about social loss’). *Classes* or *types* or *categories* (these three notions can be treated as equivalent) can be divided into subclasses, subtypes, or subcategories. Subclasses or subcategories can themselves be divided into further subclasses. A basic idea of grounded theory is that the whole structure or system of categories should not be exclusively developed in a top-down manner by deriving subclasses from major categories. Instead, researchers are encouraged to find major categories by carefully comparing the initially found categories (which may later become subcategories) and by integrating them into a larger structure. A hierarchically ordered structure of subcategories can develop (see Figure 9.1). It becomes possible to differentiate incidents in the
Figure 9.1 Hierarchical structure of (sub)categories.

Data or text passages classified as dealing with social loss according to the subcategory ‘degree of social loss’ by forming further subcategories ‘high social loss,’ ‘medium social loss,’ and ‘low social loss.’ The whole range or set of these three subcategories then represents a theoretical property of the category social loss.

It is important to note that classes can always be divided into subclasses according to different criteria. Nurses use social attributes to calculate social loss, some of which are perceived immediately (like age, gender), while others are learned after some time (like social class, educational status). Thus, significance of social attributes could form a further subcategory of social loss with the two subcategories apparent attributes and learned attributes.

There are sets of (sub)categories which are mutually exclusive, while others are not. The subcategories high social loss, medium social loss, and low social loss form a range of mutually exclusive (non-overlapping) classes: there should be no incident which can be assigned to more than one of the three subcategories (which means that a certain utterance about a social loss cannot express simultaneously high social loss and low social loss). There are other categories or classes to which objects can be assigned simultaneously. Figure 9.2 demonstrates a very simple geometrical example for that. The 10 objects can be classified according to the category ‘size’ (with the subcategories small and big) and according to the category ‘shape’ (with the subcategories rectangular and circular).

In grounded theory, incidents and text segments can also be assigned to several categories in a similar (albeit more complicated) fashion: by carefully comparing text segments dealing with social loss, one may find that many of them can be also attributed to the category ‘patient care’ (since they contain nurses’ utterances about how to care for dying patients) and to the subcategories...
intense patient care’ or ‘cursory patient care.’ One may even find that text segments relating to high social loss can often be assigned to intense patient care which shows a possible relation between the two categories. This relation can likewise be regarded as a theoretical property of the category social loss.

By looking for commonalities and differences between incidents, the constant comparative method can thus reveal two different kinds of theoretical properties: possible sets of subcategories of a given category, and relations to other categories. The decisive question of constant comparison is thus: according to which criteria do the incidents vary? These criteria of variation form the categories or subcategories the analyst looks for, or they are at least suggestive of such categories. Thereby the analyst has to obtain the most basic rule of grounded theory: do not force preconceived categories on the data, but let the categories emerge from the data.

**Theoretical sensitivity as a prerequisite for category building**

Similar to the concepts ‘category’ and ‘properties of categories,’ many researchers who start their first grounded theory project will also find the idea of emergence difficult to apply in practice. In particular, the request ‘literally to ignore the literature of theory and fact on the area under study, in order to assure that the emergence of categories will not be contaminated’ (Glaser & Strauss, 1967: 37) can lead inexperienced users of grounded theory procedures to adopt an unrealistic idea about their work. Novices who wish to firmly observe the principle of ‘emergence’ often experience the search for categories as extremely tedious and a subject of sometimes numerous and endless team sessions, leading to a proliferation of categories which makes the whole process insurmountable. Often these researchers translate the instruction to let categories emerge from the data into a demand to transform every idea or concept which comes into their minds when reading the textual data into a category. In a methodological self-reflection, a group of junior researchers who had asked me for advice described this problem as follows:

Especially the application of an open coding strategy recommended by Glaser and Strauss—the text is read line by line and coded *ad hoc*—proved to be unexpectedly awkward and time consuming. (...) Our attempts to analyse the data were governed by the idea that we should address the text *tabula rasa* and by the fear to structure data too much on the basis of our previous knowledge. Consequently *every* word in the data was credited with high significance. These uncertainties were not eased by advice from the corresponding literature that open coding means a ‘preliminary breaking down of data’ and that the emerging concepts will prove their usefulness in the ongoing analysis. Furthermore, in the beginning we had the understanding that ‘everything counts’ and ‘everything is important’—even every marginal incident and phenomenon was coded, recorded in numerous memos, and extensively discussed. This led to an insurmountable mass of data ... (cf. Kelle et al., 2002, translated from German to English by Udo Kelle).

These researchers did not use line-by-line coding as a device for the initial breaking down of the data (as proposed by Glaser and Strauss), but as a tedious task of tracking down a ‘complete and true meaning’ in the data. Although such
a procedure certainly goes against the intentions of the founders of grounded theory, it is interesting to note that the experience of these researchers underlines an important epistemological insight which is nowadays generally accepted: scientific observations are always ‘theory laden’ (cf. Hanson, 1965).

The idea that theoretical categories and propositions could be derived by simple (‘inductive’) generalization from observable data by researchers who have freed their minds from any theoretical preconceptions whatsoever before collecting empirical data manifests a rather outmoded view of scientific inquiry (nowadays often called ‘naive empiricism’ or ‘naive inductivism,’ cf. Chalmers, 1999). It is one of the most widely accepted insights of the philosophy of science and cognitive psychology that ‘there are and can be no sensations unimpregnated by expectations’ (Lakatos, 1978: 15) and that the construction of theoretical categories, whether empirically grounded or not, cannot start ab ovo, but have to draw on already existing stocks of knowledge. ‘Both historical examples and recent philosophical analysis have made it clear that the world is always perceived through the “lenses” of some conceptual network or other and that such networks and the languages in which they are embedded may, for all we know, provide an ineliminable “tint” to what we perceive’ (Laudan, 1977: 15). Empirical researchers (whether in the natural sciences or in the humanities) would need such ‘lenses’ or conceptual networks otherwise they would not be able to observe and describe meaningful events.

Certainly Glaser and Strauss did not overlook this problem, as one can easily see if one reads The Discovery of Grounded Theory with care. On page 3 they emphasize: ‘Of course, the researcher does not approach reality as a tabula rasa. He must have a perspective that will help him see relevant data and abstract significant categories from his scrutiny of the data’ (Glaser & Strauss, 1967: 3). Later they give a more detailed account of what they mean by the researcher’s ability to ‘see relevant data.’ To discover grounded theories one needs ‘theoretical sensitivity,’ an ‘ability to have theoretical insight into [one’s] area of research, combined with an ability to make something of [one’s] insights’ (p. 46). But how can this ability be achieved? The Discovery of Grounded Theory contains only a few pages that address this question, comprising two advisory statements about how theoretical sensitivity can be enhanced:

1. The sociologist should harbor ‘an armamentarium of categories and hypotheses on substantive and formal levels. This theory that exists within a sociologist can be used in generating his specific theory’ (Glaser & Strauss, 1967: 46). Obviously existing theories can be clearly helpful in category building, and fruitful insights shall be drawn from them. Glaser and Strauss maintain that an empirically grounded theory usually combines categories and hypotheses which have emerged from the data with concepts arising from the researcher’s previous theoretical knowledge.

2. A strong commitment to ‘one specific preconceived theory’ (Glaser & Strauss, 1967: 46), especially to ‘grand theory,’ must be avoided to get around the danger that categories of a ‘pet theory’ are forced on the data.

‘Indeed the trick is to line up what one takes as theoretically possible or probable with what one is finding in the field’ (Glaser & Strauss, 1967: 253).
This trick, however, is difficult to learn for many researchers. Given the concept of ‘emergence,’ so highly esteemed in *The Discovery of Grounded Theory*, and given the advice to abstain from reading literature about the field under investigation, some readers may be drawn towards the idea that theoretical concepts emerge from the data only if the empirical field is approached with no preconceived theories or hypotheses whatsoever. Others who wish to demonstrate ‘theoretical sensitivity’ may be worried that they will force inappropriate categories on the data when applying a specific sociological theory. It is obviously a challenging task to approach the empirical field without a predefined set of categories (which are applied on each empirical observation), and to hold various different theories in abeyance in order to use theoretical concepts only if they fit the data. Unfortunately, *The Discovery of Grounded Theory* neither gives clear-cut methodological rules nor practical examples about how previous theoretical knowledge can be fruitfully introduced in the process of category building. Thus the two basic rules presented previously, to abstain from forcing preconceived concepts, and to utilize theoretical sensibility in this process, are obviously difficult to reconcile.

**THE GLASERIAN VS. THE STRAUSSIAN APPROACH: TWO WAYS OF THEORETICALLY INFORMED CATEGORY BUILDING**

*The Discovery of Grounded Theory* invites empirical researchers to develop their own ideas instead of applying ungrounded theories in their empirical field and restricting the empirical work on the testing of hypotheses, which has a liberating and stimulating effect, especially on junior researchers or doctoral students. However, in the years following *The Discovery of Grounded Theory*, the apparent antagonism between emergence and theoretical sensitivity remained a major problem for teaching the methodology of grounded theory.

In their later writings, Glaser and Strauss undertook a variety of efforts to clarify the idea of theoretical sensitivity and to account for the necessary ‘theory-ladenness’ of empirical observation.Regrettably, these efforts led both authors in different directions and contributed to a major split between them. In 1978 (more than one decade after the publication of *The Discovery of Grounded Theory*), Barney Glaser tried to clarify the concept of theoretical sensitivity in his own monograph of that title. In doing so, he coined the terms ‘theoretical codes,’ ‘theoretical coding,’ and ‘coding families’ to describe a process whereby analysts have a great variety of theoretical concepts at their disposal to structure the developing categories and the emerging theory. Strauss developed a different conception: the most important categories shall be developed and related to each other by a so-called ‘paradigm model,’ a straightforward model of human action and interaction rooted in pragmatist social theory. (In his later writings Strauss gave up this strong inclination towards action.) These two models for grounding categories and their differences will be
described and their advantages and disadvantages will be discussed in the following section.

**Barney Glaser’s model of theoretical coding**

According to Glaser, one of the main reasons to write *Theoretical Sensitivity* was that the discussion of theoretical sensitivity turned out to be ‘a major gap in *The Discovery of Grounded Theory*’ (Glaser, 1978: 1). Glaser here attempts to unfold the technical aspects of theoretical sensitivity by inventing a new differentiation in the coding process. He distinguishes between ‘theoretical coding’ and ‘substantive coding’ while linking two different types of codes to these forms of coding: substantive codes and theoretical codes. (Codes or conceptual codes is thereby used as synonymous for ‘categories and their properties.’)

*Substantive codes*, which shall relate to the empirical substance of the research domain are developed during open coding, the first stage of the coding process. It could be either words informed by the language use of actors in the field (so called ‘in vivo codes’) or notions drawn from sociological terminology (which Glaser calls ‘sociological constructs’). In order to establish relations between such substantive codes, the analyst needs theoretical codes which ‘conceptualize how the substantive codes may relate to each other as hypotheses to be integrated into a theory’ (Glaser, 1978: 72). Theoretical codes are terms which describe possible relations between substantive codes and thereby help to form theoretical models. The examples Glaser uses for such theoretical codes are formal and highly abstract concepts from epistemology and sociological ‘grand theory’ which make basic claims about the ordering of the (social) world. Terms like causes, contexts, consequences, and conditions, for instance may help to develop links between codes: by calling certain events (which were coded with the help of substantive codes) ‘causes’ and others ‘effects,’ the previously developed substantive codes can be integrated into a causal model.

Glaser then presents a list of so-called theoretical ‘coding families,’ merging concepts which come from various (sociological, philosophical, or everyday) contexts, for example:

- terms, which relate to the degree of an attribute or property (‘degree family’), like ‘limit,’ ‘range,’ ‘extent,’ ‘amount,’ etc.
- terms, which refer to the relation between a whole and its elements (‘dimension family’), like ‘element,’ ‘part,’ ‘facet,’ ‘slice,’ ‘sector,’ ‘aspect,’ ‘segment,’ etc.
- terms, which refer to cultural phenomena (‘cultural family’) like ‘social norms,’ ‘social values,’ ‘social beliefs,’ etc.

and fourteen further coding families containing notions from diverse theoretical backgrounds, debates, and schools of philosophy or the social sciences. Thereby, many of these terms can be subsumed under different coding families: the term *goal*, for instance, belongs to a coding family referring to action strategies
(strategies family) as well as to a coding family referring to the relation between means and ends (means-goal family).

The diverse coding families can obviously serve as a fund of concepts which may guide researchers in developing their ability to think about empirical observations in theoretical terms. However, their utility for the development of theoretical relations between the ‘substantive codes’ is limited. The reason for that is that one can make sense of coding families (which Glaser presents as an unsorted list of terms) only if one is clear about their inner relations and their embeddedness into greater conceptual networks. This can be easily demonstrated with regard to Glaser’s first coding family (the six C’s, which he obviously considers as the most important) referring (among other things) to causal relations. Terms denoting causal relations (like cause, condition, consequence, etc.), however, are in themselves not sufficient for the development of causal models. To construct a causal model about the relation between specific events it would be necessary to use substantial (i.e. sociological, psychological) categories which define those types of events which can be generally considered as causes and those which are usually to be seen as effects. By using merely a formal coding family denoting causal relations without reference to substantial categories, one could treat arbitrarily all kinds of events which can be found in the research field as causes and effects. To simply use the notion of causality while investigating youth delinquency could mean that one regards deviant behavior either as a cause or as an effect of negative sanctions. To choose between these two possibilities one would not primarily need formal terms (like cause and effect) but a theoretical code based on a sociological perspective. This could either (like classical learning theory or role theory) explain sanctions as a result of behavior or it could describe deviant behavior as a result of stigmatizing sanctions (like labeling theory). A crucial problem with Glaser’s list of coding families is that it lacks a differentiation between formal or logical categories (like causality) and substantial sociological concepts (like social roles, identity, culture); both types of categories would have to be linked to each other in order to develop empirically grounded categories. Although Glaser’s list of coding families certainly does not exclude such a sophisticated use of theoretical codes, the whole problem is not even mentioned in Theoretical Sensitivity. This causes problems particularly for novice researchers trying to make adequate use of the whole concept of theoretical coding: in a similar fashion to the notion of theoretical sensitivity, the concept does not entail a set of methodological rules applicable in research practice concerning how to structure the emerging categories with the help of theoretical knowledge. The problem is not so much that Glaser’s list of coding families would not be sufficient to stimulate the discovery of possible theoretical relations between incidents in the data or between newly developed categories. It is rather that the employment of such an unordered list for the construction of grounded theories is very difficult if the researcher does not have a very broad theoretical background knowledge at hand concerning the different theoretical perspectives entailed in the list.
**Anselm Strauss’s concept of a ‘coding paradigm’**

In his book *Qualitative Analysis for Social Scientists*, published in 1987, Anselm Strauss goes one step further in explicating the concrete steps a researcher can take to develop categories from empirical data with a theoretical perspective in mind. The book had evolved from a research seminar set up to train students in grounded theory procedures (Legewie & Schervier-Legewie, 2004). Like Glaser before him (who used experiences from the seminar which he had run between 1968 and 1979 to author *Theoretical Sensitivity*), Strauss was aware of the difficulties novices ‘have in generating genuine categories. The common tendency is simply to take a bit of the data (a phrase or sentence or paragraph) and translate that into a précis of it’ (Strauss, 1987: 29). Many users of grounded theory procedures, novices in particular, often did not come to terms with developing true theoretical categories. Frequently, categories plucked from the data were only summaries or descriptions.

When Strauss wrote his own new methodological monograph, he and Barney Glaser had not worked together on joint research projects for some time. Although pages of the introduction are filled with extensive quotes from *Theoretical Sensitivity*, and some of the terms Glaser had invented were mentioned (for instance, *open coding*, *in vivo codes*, or *sociological constructs*), Strauss paid absolutely no heed to two concepts pivotal to Glaser’s view on theoretical sensitivity: *theoretical coding* and *coding families*. Furthermore, Strauss invented the new term (coding paradigm) which he used to structure data and to clarify relations between categories:

> It is central to the coding procedures. Although especially helpful to beginning analysts, in a short time this paradigm quite literally becomes part and parcel of the analyst’s thought processes. Whether explicit or implicit, it functions as a reminder to code data for relevance to whatever phenomena are referenced by a given category (Strauss, 1987: 27).

The coding paradigm comes into play during ‘axial coding,’ an advanced stage of open coding. Whereas open coding starts by ‘scrutinizing the fieldnote, interview, or other document very closely; line by line, or even word by word. The aim is to produce concepts that seem to fit the data’ (Strauss, 1987: 28), axial coding ‘consists of intense analysis done around one category at a time in terms of the paradigm items’ (Strauss, 1987: 32). This category forms the ‘axis’ around which further coding and category building is done and may eventually become the core category of the emerging theory.

Strauss elaborates further on the concepts ‘coding paradigm’ and ‘axial coding’ in *Basics of Qualitative Research* published together with a new co-author, Juliet Corbin (Strauss & Corbin, 1990). This book attempted to describe grounded theory procedures in a didactic step-by-step mode. The coding paradigm fulfils the same function as a Glaserian coding family; it represents a group of abstract theoretical terms which are used to develop categories from the data and to find relations between them. Similar to Glaser’s coding families, the coding paradigm takes into account that the development
of categories requires either a previously defined theoretical framework or at least the possibility to draw on a selection of such frameworks if one wants to avoid being flooded by the data.

However, the coding paradigm turned out much more instructive for many grounded theory users than the coding family conception. While Glaser had proposed a long and only loosely ordered list of more or less related groups of sociological and formal terms, Strauss and Corbin advise the researcher to use one general model of action rooted in pragmatist and interactionist social theory (cf. Corbin, 1991: 36; Strauss, 1990: 7) to build a skeleton or 'axis' for the developing categories and their relations. Thereby, a general ‘paradigm model’ (Strauss & Corbin, 1990: 99) is established which determines the analysis of action and interaction strategies of the actors as the main purpose of grounded theory. Special emphasis is laid on the intentions and goals of the actors in this process. Categories developed during open coding shall be investigated whether they relate to: (1) phenomena at which the actions and interactions in the domain under study are directed; (2) causal conditions which lead to the occurrence of these phenomena; (3) additional intervening conditions by which the investigated phenomena are influenced; (5) action and interactional strategies the actors use to handle the phenomena; and (6) the consequences of their actions and interactions. Thus the analyst is advised to identify types of phenomena, contexts, causal, intervening conditions, and consequences which are relevant for the most important category or categories in order to develop an 'axis' for the grounded theory. If, for instance, social aspects of chronic pain are investigated, the researcher shall try to determine action contexts in the data which are typical for patients with chronic pain as well as characteristic patterns of pain management strategies. Thereafter, it can be examined which pain management strategies are used by persons with chronic pain under certain conditions and in varying action contexts. This may then lead to the construction of models of action which provide the basis for a theory about action strategies generally pursued under conditions of chronic pain.

**Pros and cons of the two approaches**

Regarding the role of previous theoretical knowledge in the research process, one can now draw on two different versions of grounded theory which vary to a considerable extent. In 1992, Glaser attacked Strauss and Corbin in a monograph published in his private publication venture titled *Emergence vs. Forcing*. In this, Glaser accuses Strauss of having betrayed the common cause of grounded theory. By applying the concepts of axial coding and coding paradigms, researchers would force categories on the data instead of allowing the categories to emerge. Glaser’s charges, which were written in an exceptionally polemical and even personally hostile manner, were never answered publicly by Strauss and Corbin. Nevertheless, several questions remain: which of the two approaches would better reflect the original intentions of grounded theory?
Which one would be better suited for developing categories from the data and which is more easily applicable in research practice?

All of the differing concepts ‘theoretical sensitivity,’ ‘theoretical coding,’ ‘axial coding,’ ‘coding paradigm,’ and ‘coding families’ represent attempts to solve a fundamental methodological problem which arises with the claim to let categories ‘emerge’ from the data: a strategy of investigation which approaches an empirical domain without any theoretical preconceptions is simply not feasible—such a method would yield a plethora of incoherent observations and descriptions rather than empirically grounded categories or hypotheses. The emergence of theoretical categories which can adequately describe phenomena in the empirical field is always dependent on the researcher’s theoretical sensitivity, their ability to grasp empirical phenomena in theoretical terms. This competence demands an extended training and a broad background in sociological theory (cf. Glaser, 1992: 28). This is especially the case if the Glaserian approach is used: one would need an advanced understanding of different thoughts of school, their terminology, and their possible relations to make use of Glaser’s list of coding families, to choose the coding families most adequate for the data and to combine different coding families in a meaningful way. Nevertheless, the methodological usefulness of this list is limited in more than one respect: novices in empirical research will have difficulties in handling the more or less unsystematic compilation of theoretical terms from various sociological and epistemological backgrounds. Researchers with broad theoretical background knowledge, and longstanding experience in the application of theoretical terms, will certainly not need such a list.

At first glance, Strauss and Corbin’s ‘coding paradigm’ represents a more user-friendly concept, since it describes the construction of a theoretical framework for the development of empirically grounded categories in an explicit manner. By drawing on this concept, researchers with limited experience in the application of theoretical knowledge can use grounded theory methodology without taking the risk of being flooded by the data. Researchers may feel too constrained by the specific theoretical perspective embedded in the coding paradigm which stems from a certain theoretical tradition: philosophical and sociological pragmatism originating from the works of Peirce, Dewey, and Mead. Glaser’s critique that the coding paradigm may lead to the forcing of categories on the data thus cannot be dismissed.

However, the conceptual design of the coding paradigm carries a broad and general understanding of action which is compatible with a wide variety of sociological theories (ranging from sociological phenomenology to Rational Choice Theory and even to functionalist role theory), and which is also entailed in several of Glaser’s coding families: viewed from a Glaserian perspective, the coding paradigm can be regarded as a combination of aspects which Glaser disperses among different coding families, since it blends parts of the first coding family (the six C’s, causes, contexts, contingencies, consequences, covariances, and conditions) with the ‘strategy family’ and the ‘means-goal family.’
The coding paradigm presented by Strauss and Corbin in 1990 is specifically linked to a micro-sociological perspective on social phenomena emphasizing the role of human action and agency in social life. Interestingly, this perspective is also shared by Glaser who rarely includes terms from macro-sociological approaches into his coding families. Furthermore, in Theoretical Sensitivity, he himself asserts that coding and coded incidents have to be related to actions of the actors in the empirical domain. Finally, the concept of ‘Basic Social Processes’ which was crucial for Glaserian grounded theory particularly at that time shows strong associations to sociological pragmatism. Thus it is at least doubtful whether theoretical coding sensu Glaser was originally developed to foster a highly pluralistic use of theoretical codes, including the use of concepts from macro-sociological approaches (although theoretical coding can be definitely expanded in that direction). All of the substantive (as compared to the formal) coding families presented by Glaser in Theoretical Sensitivity show a strong relation to a micro-sociological perspective which places actors and their actions in the focus of analysis (with categories like strategies, tactics, maneuverings, identity, goals, anticipated consequences, and others). There is no coding family referring to system theory (with terms like integration, differentiation, or emergent properties). However, an analyst familiar with such a theoretical perspective may easily develop such a coding family and apply it, for instance, in a grounded theory about a certain organization. However, researchers with a strong background in macro-sociology and system theory may feel uncomfortable with Strauss and Corbin’s coding paradigm, since such a micro-sociological and action-oriented approach goes contrary to their requirements.

Following the Straussian route by constructing one’s own coding paradigm connected to the theoretical tradition one prefers would be one possibility to stick with grounded theory methodology without adopting the (meta)theoretical orientation of its founding parents. The methodology of grounded theory is undoubtedly open enough to allow for such a stance. The other possibility to avoid unwillingly introducing unwanted theoretical tenets is to draw on Glaserian theoretical coding while choosing and developing suitable theoretical codes and coding families using a theoretical approach one finds suitable for the area under study. In a grounded theory study about the process of care giving for elderly and frail relatives at home, for instance, one may use theoretical codes derived from decision theory (focusing on the intentions of the actors, as well as their perceived ‘costs’ and ‘benefits’ of care giving) or codes based on role theory (focusing on the expectations of the social environment). Such a strategy is clearly more flexible in regard to the variety of theoretical perspectives which can be used. However, one has always to keep in mind that such theoretical codes must not be ‘forced’ on the data (which means that they may only be used if the data material itself suggests their use: one may only apply an approach based on utility theory, for instance, if the research subjects did refer to what can be considered ‘costs’ or ‘benefits’ of the care giving situation). This makes such
a strategy much more challenging for novices than the use of a readymade conceptual framework or coding paradigm.

As far as the role of previous theoretical knowledge is concerned, the crucial difference between Glaserian and Straussian category building lies in the fact that Strauss suggests the utilization of a specified theoretical framework based on a certain understanding of human action, whereas Glaser emphasizes that coding is a process of combining ‘the analyst’s scholarly knowledge and his research knowledge of the substantive field’ (Glaser, 1978: 70) and has to be realized in the ongoing coding process, which often means that it has to be conducted on the basis of a broad theoretical background knowledge which cannot be made fully explicit in the beginning of analysis.

Unfortunately, Glaser tends to exaggerate these differences and resorts to an inductivist rhetoric which produces a highly obscure image of empirical research. Therefore, he does not only expect a researcher working with grounded theory procedures to approach the research field without any precise research questions or research problems (‘He moves in with the abstract wonderment of what is going on that is an issue and how it is handled,’ Glaser, 1992: 22); he also burdens his method (and possible users) with unrealistic truth claims. Thus Glaser maintains: ‘In grounded theory (...) when the analyst sorts by theoretical codes everything fits, as the world is socially integrated and grounded theory simply catches this integration through emergence’ (Glaser, 1992: 84). ‘Grounded theory looks for what is, not what might be, and therefore needs no test’ (Glaser, 1992: 67). Such assertions display at least a gleam of epistemological fundamentalism (or ‘certism’; Lakatos, 1978) and perpetuate the outmoded and positivist idea that by using an inductive method empirical researchers may gain the ability to conceive ‘facts as they are.’ However, Glaser makes clear elsewhere that theoretical concepts do not simply arise from the data alone but require careful ‘theoretical coding’ (that means: the categorizing of empirical data on the basis of previous theoretical knowledge). Thus the suspicion arises that his ‘emergence talk’ does not describe a methodological strategy but simply offers a way to immunize theories from criticism with the help of a methodological rhetoric (claiming that researchers following the ‘true path’ of grounded theory can never go wrong since the categories have emerged from data).

Grounded theory was originally developed to provide an alternative to a strict hypothetico-deductive model of social research which restricts the function of empirical research to the testing of ready-made hypotheses. This experimental model of research represented a mainline approach towards methodology in social research at least in the 1950s and 1960s, although field studies conducted by famous ethnographers (like Malinowski or Margaret Mead) or outstanding sociological researchers (namely the members of the Chicago School) in the first decades of the twentieth century had demonstrated that field research in the social sciences cannot be pursued with the help of such a model. The concept of emergence which was meant to replace the
deductive concept of hypothesis testing drawn from experimental research can provoke misunderstandings (namely the idea that categories emerge from the data if a researcher avoids the use of any theoretical preconception whatsoever). Qualitative researchers applying grounded theory methodology must therefore use the emergence concept in an epistemologically informed way: they must take into account that although qualitative research does not start with ready-made and precise hypotheses, the development of categories from empirical data is dependent on the availability of adequate theoretical concepts. In the following it will be shown how the concept of empirical content can assist researchers in using their previous theoretical knowledge without forcing the data.

THE USE OF PREVIOUS THEORETICAL KNOWLEDGE IN CATEGORY BUILDING

If one sets aside Glaser’s inductivist rhetoric, his concepts of theoretical codes and coding families represent a way systematically to introduce theoretical knowledge into the coding process without ‘forcing’ preconceived categories on the data. The conception of coding families make clear that certain types of theoretical knowledge are clearly helpful in deriving grounded categories from the data. Furthermore it allows for more theoretical pluralism than Strauss and Corbin’s coding paradigm. However, as has been said before, the conception of coding families will be more difficult to employ for non-experienced researchers who may have problems in realizing the differences between theoretical knowledge, which forces the data, and theoretical concepts, which help with the emergence of suitable theoretical categories from the data. This situation may lead to the impression that theoretical sensitivity is merely an individual quality of researchers. However, if the use of previous theoretical knowledge is based on a sound methodological and epistemological basis novice researchers may easily learn to distinguish between theoretical notions that force the data and concepts that support the emergence of new categories. If that distinction can be made in a proper way the use of a predefined coding paradigm will also become possible: the researcher will then be able to select theoretical codes before the data are coded and to use predefined category schemes without abandoning basic principles of grounded theory (specifically the principle of discovering new patterns and relations in the data). But how can this task be achieved?

In the following text, it will be demonstrated that an understanding of classical methodological concepts like ‘falsifiability’ or ‘empirical content’ will be helpful for that purpose. A distinction between theoretical notions with high empirical content on the one hand and notions with low empirical content on the other hand can be an important tool for understanding the role of previous theoretical knowledge in developing grounded categories. Theoretical categories
with high empirical content or high falsifiability are an essential part of a hypothetico-deductive research strategy, but can easily lead to forcing of the data if a grounded theory approach is used. Take the following hypothesis as an example: ‘Young adults with a middle class background (defined in terms of income and educational status of their parents) have a better chance of attending university than young people with a working class background.’ This hypothesis can be empirically tested (and thus be falsified if counter evidence is discovered during such a test), since each category in this statement can in principle be operationalized and measured (e.g. by defining young adults as ‘men and women aged between 18 years and 25 years’ or defining educational status as ‘level of school leaving exam’). Categories and hypotheses of that type guide good quantitative social research, but the attempt to design a grounded theory project on such a basis would lead to methodological misuse: one would collect lots of material not suited to test the specific hypotheses and definitely disregard the richness of the data one can collect with the help of qualitative methods. The process of theory building in a grounded theory project can benefit from theoretical concepts with low empirical content which do not force the data but may serve as heuristic devices which will be discussed in the following two sections. In these sections I will present examples for two different types of heuristic concepts with limited empirical content which can be either utilized in open coding or as a means to define category schemes before data are collected: first, theoretical notions, definitions and categories drawn from ‘grand theories’ and second, common sense categories which relate to general topics of interest covered in the data.

**Theoretical categories with limited empirical content as heuristic devices**

A great number of theoretical notions, definitions, and categories drawn from grand theories (categories like identity, status, roles, systems, structure, values, and deviance), lack empirical content; they are so broad and abstract that it is difficult to directly deduce empirically contentful or falsifiable propositions from them which can be tested in a hypothetico-deductive framework. A proposition like ‘A social role defines the expected behavior connected to a given social position’ may serve as a good example of that. This statement has no (or very limited) empirical content which means that it is hard to imagine a direct empirical test which could disprove or falsify it: someone who tries to present a counter example (that would mean a social role not defining expected behavior connected to a social position) would even demonstrate thereby that he did not understand the meaning of such a statement: it is obviously not suited for being tested through empirical data. However, such theoretical categories can sensitize the researcher to identify theoretically relevant phenomena in their field. Herbert Blumer had proposed the term ‘sensitizing concepts’ to denote abstract notions which ‘lack precise reference and have no bench marks which allow a clean cut
identification of a specific instance' (Blumer, 1954: 7) and distinguished them from 'definitive concepts':

Whereas definitive concepts provide prescriptions of what to see, sensitizing concepts merely suggest directions along which to look. The hundreds of our concepts—like culture, institutions, social structure, mores and personality—are not definitive concepts but are sensitizing in nature (Blumer, 1954).

Sensitizing concepts can fulfill an important role in empirical research, since their lack of empirical content permits researchers to apply them to a wide array of phenomena. Regardless how empirically contentless and vague they are, they may serve as heuristic devices for the construction of empirically grounded categories. A category like 'identity' may serve as an example for that. To operationalize such a category is much more difficult than to operationalize terms like 'university attendance' or 'educational status of parents.' And to directly derive falsifiable and testable propositions from that concept alone (that means without using any further concepts or empirical information) could be a troublesome task; the assertion, for instance, that individuals develop a certain identity in a given social domain does not imply a lot of information or empirical content by itself. Such a concept could be, however, extremely useful if one wants to formulate empirical research questions for a given substantive field: Does identity formation play an important role in the empirical domain under study? What processes of identity formation take place? By which means do empirical actors develop and defend their identity? Is the identity or self-definition of certain actors challenged by others? Which strategies do actors employ to defend and to maintain their identity if challenged? How do others counteract such strategies? Therefore, a sensible way to use a heuristic concept like identity in grounded theory research is not to derive a 'hypothesis,' which can be 'empirically tested' (like the hypothesis about the relation between social background and university attendance presented above), but to employ it as a conceptual frame which helps to understand empirical phenomena found in the research field. By employing a term like identity in a study about juveniles living in institutional care one may identify strategies how young people preserve and develop their identity under difficult circumstances. Identifying such strategies may lead to the development of empirically contentful statements (maintaining, for instance, that young people in care homes develop different strategies of identity formation compared to children in foster families) which can be further examined in subsequent research. Thus one can apply abstract theoretical categories with a general scope (which refer to various kinds of phenomena) but with limited empirical content (like identity or social role) as heuristic devices to develop empirically grounded categories with a limited scope and high empirical content (like 'identity formation of juveniles taken away from their parents and living in care homes in the UK').

A variety of concepts coming from differing theoretical approaches in sociology and social psychology can be used in such a way. It is also possible to use
categories from schools of thought which are normally remote from qualitative research, like rational choice theory, as heuristic devices. A core assumption of rational choice theory is that human actors will choose the action which seems the most adequate for the achievement of a desired goal from a set of given action alternatives. However, without specifying which goals the actors pursue and which actions they consider to be adequate, the theory is like an ‘empty sack’ (cf. Simon, 1985). Thus, the core assumptions of utility theory contain almost no empirical content if they are not supplemented by auxiliary assumptions or ‘bridge hypotheses.’ Accordingly, rational choice theory may provide qualitative researchers with useful research questions, sensitizing concepts, and heuristic categories: one may, for instance, code incidents in the data which refer to the potential costs and benefits certain actions have for the actors, or one may code incidents which relate to the intentions and goals of the research subjects or to the means they use to reach their goals.

In this manner, a wide array of sensitizing categories from different theoretical traditions can be used to develop empirically grounded categories. Many researchers find it easier to let categories emerge if one stays with one particular theoretical tradition, however Glaser is certainly right with his frequent warnings that the utilization of a single pet theory will almost necessarily lead to the neglect of heuristic concepts better suited to the specific domain under scrutiny. There are heuristic concepts which capture a broad variety of different processes and events and nevertheless may exclude certain phenomena from being analyzed: thus the extended use of concepts from micro-sociological action theory (e.g. actors, goals, strategies) can preclude a system theory and macro-perspective on the research domain. A strategy of coding which uses different and even competing theoretical perspectives may often be superior to a strategy which remains restricted to a limited number of pet concepts. Furthermore, analysts should always ask themselves whether the chosen heuristic categories lead to the exclusion of certain processes and events from being analyzed and coded, since this would be an attribute of a category with high empirical content which refers to a circumscribed set of phenomena (akin to the definitive category ‘young adults’ excluding older persons).

Common sense categories

Another kind of heuristic categories which do not force data but allow for the discovery of empirically contentful categories are categories which refer to topics of interest contained in the data. Drawing on general commonsense knowledge can easily identify such ‘topic oriented categories.’ Categories like school, work, or family are simple examples of that, but topic oriented categories may be far more complex. They can also be related to specific local knowledge of the investigated field the researcher knows beforehand and thus mimic in vivo codes (although they are not discovered during open coding). However, as with heuristic theoretical concepts, one question remains of utmost importance: Does a certain
code serve for heuristic purposes and can it thereby be related to important phenomena in the field or does it exclude relevant processes or events from analysis?

Suitable heuristic categories, whether developed from grand theory or taken from commonsense knowledge, do not force the data since they lack empirical content. This makes them useless for a hypothetico-deductive strategy; their strength is in the context of exploratory, interpretive research. Such categories fit various kinds of social reality and it will not be necessary to know concrete facts about the investigated domain before data collection takes place. Heuristic categories play the role of a theoretical axis or a skeleton to which the flesh of empirically contentful information from the research domain is added. The goal is to develop empirically grounded categories and propositions about relations between these categories. Category building of that kind starts by using heuristic concepts and proceeds to the construction of categories and propositions with growing empirical content.

The use of categories with high empirical content in grounded theory

It may be also possible to fruitfully employ categories and assertions with high empirical content in category building under certain circumstances. A researcher investigating the process of caregiving to frail and elderly people, for instance, may discover that Arlie Hochschild's category 'emotional labor' (Hochschild, 1983) helps the researcher to understand social interactions in the research field. This category, initially developed to describe typical patterns of action and interaction between flight attendants and air passengers, can thus adopt a heuristic function in the sociological investigation of another research domain, although it comprises more empirical content than terms like 'identity' or 'intentions': not all kinds of social interactions demand emotional labor and, compared to other sensitizing categories or propositions, the assumption that certain professionals are providing emotional labor can be falsified more readily.

The concept can be helpful for the analysis of social phenomena in certain areas. Consequently, it can be prudent in grounded theory research to use distinct and well-defined categories and propositions which contain enough empirical content to be empirically tested. The use of such concepts and the examination of hypotheses represent an older, well-established tradition in qualitative research. In the 1930s, researchers and methodologists coming from the Chicago School had proposed a research strategy named 'analytic induction,' which was used thereafter in numerous well-known qualitative studies: initial hypotheses are examined and modified with the help of empirical evidence provided by so-called 'crucial cases' (cf. Cressey, 1953: 104f; Lindesmith, 1946). By applying such a research strategy, one always carries the risk that theoretical categories are introduced which are not suited for the specific research domain and which are forced on the empirical data. The danger that heuristic concepts may contain too much empirical content is even prevalent with Strauss and Corbin's coding paradigm. Although it represents an understanding of purposeful and intentional
human action useful for the description of a wide array of social phenomena, it may draw qualitative researchers towards a micro-sociological orientation which they do not necessarily share. The advice to use only categories with particular low empirical content may constrict inexperienced researchers, since not each heuristic concept can draw the researcher's attention to sociologically relevant phenomena. This danger relates to Glaser's coding families; novices in particular may be overstrained by the task to select the heuristic category most suited for their research field among a vast selection of theoretical schools of thought.

CONCLUDING REMARKS

The development of categories which are sufficiently grounded in the empirical data requires that researchers abstain from forcing theoretical concepts on the data. The process of category building is frequently described with the term 'emergence,' particularly in the earliest versions of grounded theory. This concept is obviously burdened with methodological problems while evoking empiricist ideas about researchers freeing themselves from any theoretical knowledge before approaching their empirical field. Glaser and Strauss did not overlook the fact, however, that researchers always have to draw on existing stocks of theoretical knowledge in order to understand, describe, and explain empirically observed phenomena. In The Discovery of Grounded Theory, they use the term 'theoretical sensitivity' to denote the ability to identify theoretical relevant phenomena in the data. In their further methodological writings, Glaser and Strauss elaborated this concept in quite differing directions. Previous theoretical knowledge may thus be used in different ways in the process of developing empirically grounded categories:

1. The strategy of 'theoretical coding' proposed by Glaser is especially interesting for experienced researchers with a broad background in sociological theory: if the researcher has a large stock of theoretical categories at hand and a deep reaching knowledge about their interrelations, they can easily start coding without any predefined category scheme in mind and may develop the relations between substantive codes by drawing on theoretical codes which in their opinion are suited best for the data. For that purpose, researchers can either resort to the coding families suggested by Glaser and combine them in meaningful ways or they may define their own coding families by using theoretical approaches not mentioned by Glaser.

2. Novice researchers may experience greater difficulties in selecting and combining the most appropriate coding families for their research domain. Therefore, they may benefit from using a predefined coding paradigm and thus avoid drowning in the data. This may include, but is not restricted to, the specific coding paradigm rooted in pragmatist theory of action recommended by Strauss and Corbin. As with Glaser's coding families, researchers may also construct their own coding paradigms. In doing so, one must be careful to draw on concepts which do not force the data but enhance theoretical sensitivity and can serve as heuristic devices.

A theoretically sensitive researcher employing an epistemologically informed concept of emergence would be able to differentiate between different types of theoretical categories (namely between definite and precise categories, and
broad and general heuristic concepts) and to reflect on their differing functions in the process of empirically grounded category building. Whereas definite and precise categories are suitable to form empirically contentful and falsifiable hypotheses (which in grounded theory should only be formulated at fairly advanced stages of the theory building process), broad and abstract theoretical notions are especially helpful for the empirically grounded generation of categories, since their lack of empirical content gives them flexibility to describe a great variety of empirical phenomena. Heuristic categories of that kind represent lenses through which researchers perceive facts and phenomena in their research field. If the difference between empirically contentful categories and categories as heuristic devices is observed, one need not refrain from inventing theoretical categories during open coding but may also develop a theoretical axis or coding scheme before data are coded. The decisive question which helps to make the distinction between empirically contentful and heuristic theoretical categories is: does the category exclude interesting phenomena in the empirical field from being analyzed?

REFERENCES


