

Processing Raw Data both the Qualitative and Quantitative Way

Dietmar Janetzko

Key words:

network
representation of
cognition,
probabilistic
model, change of
representation

Abstract: Representations and changes between them play a major role in education (e.g., HEWSON, BEETH & THORLEY 1998), problem solving (e.g., BAUER & REISER 1990), cognitive development (e.g., VOSNIADOU & BREWER 1992), processing of metaphors (e.g., INDURKHYA 1992) and the history of science (e.g., KUHN 1970). Change of representations (also called conceptual change) is amenable to both qualitative and quantitative analyses: either the degree of correspondence between some activities under study (quantitative aspect) and a hypothesized representation representation per se (qualitative aspect) can be the focus of investigations. This articles presents and discusses and knowledge tracking (KT), viz., an approach to analyze changes of representation on the basis of symbolic sequential data. KT allows the researcher to fully investigate both aspects of changes of representation. A web-site is described that provides free usage of the knowledge tracking engine (KTE), which is a tool for analyzing data according to knowledge tracking. The article concludes with a comparison between knowledge tracking and other approaches that rely on network representations of cognition.

Table of Contents

- [1. Introduction](#)
- [2. In Search of Mindsets](#)
 - [2.1 Talking about movies](#)
 - [2.2 Solving a problem](#)
- [3. Background Assumptions of Knowledge Tracking](#)
 - [3.1 KT analyzes sequences of symbolic data](#)
 - [3.2 KT is a theory-based method](#)
 - [3.3 KT can be employed in confirmative or in an inductive way](#)
 - [3.4 KT is amenable to both qualitative and quantitative analyses](#)
- [4. Data and Theories in Knowledge Tracking](#)
 - [4.1 The side of the theory](#)
 - [4.1.1 Elements of a Theory](#)
 - [4.1.2 Relations of a Theory](#)
 - [4.2 The side of the data](#)
- [5. Knowledge Tracking—A Condensed 5-Step Scheme](#)
 - [5.1 Step 1: Setting up relational structures](#)
 - [5.2 Step 2: Recording sequences of symbolic data](#)
 - [5.3 Step 3: Translating the relational structures into a probabilistic language](#)
 - [5.4 Step 4: Taking the relational structure to predict the data in the sequence of concepts](#)
 - [5.5 Step 5: Selecting the best fitting relational structure](#)
 - [5.6 Step 6: Synthesizing a new relational structure](#)
- [6. Related Work](#)

[7. The Knowledge Tracking Engine: Using Knowledge Tracking via the WWW](#)

[7.1 Trace](#)

[7.2 Structure](#)

[7.3 Track](#)

[8. Discussion](#)

[References](#)

[Author](#)

[Citation](#)

1. Introduction

Almost by definition, every debate has its perspectives that entail simplifications, recurrent patterns, epistemological grids or issues accepted as commonplaces while others are deeply doubted. The debate on qualitative and quantitative methods in the social sciences is no exception from this general observation. It seems to me that there are three prevailing complexities reducing and sometimes simplifying perspectives in any discussion about quantitative or qualitative methods in the social sciences:

- The *epistemological perspective* stresses that analysis of each domain or problem requires one or many appropriate methods that allow the researcher to investigate phenomena of interest. Combining various methods, especially when coming from seemingly diverse directions like qualitative and quantitative research traditions, may be fruitful since they are likely to bring various aspects of the phenomenon being researched to the fore.
- The *critical perspective* emphasizes that the quantitative method forces the objects under investigation into a narrow framework that puts the researcher into a more powerful position. These methods do not allow the persons under study to give voice to their concerns and—to make matters worse—they unjustifiably claim objectivity.
- The *historical perspective* seeks to identify the roots of a particular method into particular historical constellations (e.g., the parallel between the success of Taylorism in industrial countries and the rise of quantitative methods). [1]

One of the striking features of the debate on qualitative and quantitative methods sketched above is that various perspectives may be used to describe it. Viewed from a structural point of view, there is nothing special about this feature of the debate on qualitative and quantitative methods. In fact, change of perspectives is characteristic of many debates, discourses or forms of cognition. In this paper we will present and discuss a method that can be used to analyze such changes of perspectives. We will see that this method is open for both a quantitative and qualitative usage when examining changes of perspective. Clearly, change of perspectives is a slippery topic since it can be found both on the side of the subject under study and on the side of the researcher and his or her methods. Focusing on the subjects under study, this issue has become well-known as "change of representation", which means that individuals or groups tend to use

particular views, mindsets, or cognitive representations depending on their age, level of expertise, or according to the particular problem to be solved (BAUER & REISER 1990). Focusing on the researcher and the methods she or he used, this issue has become well-known under the heading of triangulation (e.g., CAMPBELL & FISKE 1959; DENZIN 1978). [2]

To make matters more complicated, there are good reasons to bring a quantitative and a qualitative approach to bear when one deals with both subject's and the researcher's change of representations. Driven by a qualitative research interest, the researcher may be interested in the peculiarities of a perspective or mindset of a subject or group, while seen from a quantitative point of view, the researcher wishes to know to what degree the subject under study adheres to a particular perspective. The goal of this contribution is to delineate *knowledge tracking*, which is a method that can be used to describe and assess cognitive representations of subjects and researchers both from a qualitative and quantitative point of view (JANETZKO 1996). [3]

The paper is organized as follows: Firstly, I will describe the notion of change of perspectives in cognition, viz. the phenomena that knowledge tracking has been developed to analyze. Secondly, the assumptions of knowledge tracking are outlined. Thirdly, I will focus on the distinction between the side of theory and the side of the data and on the role of each side for investigations that make use of knowledge tracking. Fourthly, the gist of knowledge tracking is presented in five steps along with some examples that should make some of the formal aspects of knowledge tracking more concrete. There is of course other work on network representations of cognition that bear a resemblance to knowledge tracking. Sixthly, I will take a look at a few of those approaches and discuss some of the parallels and differences between them and knowledge tracking. To actually use knowledge tracking for investigations it would be useful to have a tool the researcher can use without bothering all the time about the nitty gritty details of knowledge tracking. Luckily enough there is a tool like this available: The knowledge tracking engine (KTE). It is freely accessible via a web-site such that remote analyzes via KT become feasible: <http://www.knowledge-tracking.com> [Broken link, FQS, August 2005]. I will present KTE in the seventh section. Finally there is a discussion of knowledge tracking that focuses on its usage as either a quantitative or qualitative tool. [4]

2. In Search of Mindsets

I will now try to give a more concrete description of the phenomena that knowledge tracking is meant to capture. To do so, I will focus on changing mindsets, cognitive representations, or perspectives and forget for the moment the qualitative and quantitative mode of data processing. [5]

Each of the two examples ("Talking about Movies", "Solving a Problem") introduced below is meant to exemplify two things: First, the idea that the particular cognitive representation (of individuals or groups) that is dominant shapes their point of view. Without an appropriate cognitive representation,

orientation and acting in a complex environment would be fraught with severe problems. Second, cognitive representations of individuals or groups may change, and each cognitive representation will bring particular issues into the focus of these cognitive systems. Without evidence to the contrary, however, one cannot rule out that the rhythm of alternations in cognitive representations may vary considerably ranging from very quick and perhaps hectic changes to more stable alternations and even no changes indicating rigid cognitive representations. [6]

2.1 Talking about movies

Talking about movies provides a good example of the flexible but sometimes also rigid usage of cognitive representations, the assessment of which is the objective of KT. Like many other domains movies are open to different perspectives or cognitive representations. When talking of a movie you may conceive of it, say, in terms of the plot, i.e., the story-line. Alternatively, you could compare the role of a particular actress to the ones she has in other movies. Or thinking may follow the web of human relations, the historical time portrayed, the stereotypes, or many other lines your thoughts may follow. Clearly, talking about movies may involve many more cognitive structures than the ones mentioned above. Quite easily, cognitive structures may change or switch meaning on the basis of the different perspectives on a movie which may be taken. But there is of course the possibility of bringing just one cognitive structure to bear. Examples are the well-known—and sometimes notorious—attempts to re-tell a movie just in terms of the story-line. More often, however, quite a number of cognitive structures seem to establish the pattern our thinking is pursuing. [7]

2.2 Solving a problem

I will now turn to the second example that shows how different cognitive representations relate to different points of view. Let us imagine that I want to find a fault in a technical system. If I think that a short-circuit in the system has caused the fault, I will most probably think of the technical systems in terms of the conducting elements. If, however, the nature of the problem changes, a different cognitive representation has a better chance to shape my point of view. If, for instance, one component of the same technical system as the one mentioned above has to be replaced by another, more efficient one, I will tend to think of the system in terms of functionally equivalent components. [8]

Clearly, the close relationship between change of representation, point of view and successful actions is not tied to an individual level. As a matter of fact, groups may also change their dominant representations. The only difference to the individual level is that selection of a representation in groups has a greater chance to become de-automatized and public once opposing views have a chance to challenge a perspective or a within group-consensus. With these considerations in mind we can now turn to the examination of knowledge tracking, an approach to analyze changes of representation on the basis of symbolic sequential data. [9]

3. Background Assumptions of Knowledge Tracking

Knowledge tracking (KT), i.e., the method I will describe in this contribution, rests on a number of assumptions that need to be made explicit. [10]

3.1 KT analyzes sequences of symbolic data

The data fed into KT are sequences of symbols. Examples of this sort of data are abundant in almost all fields of social research: thinking aloud protocols and other sequences of symbols forming texts, eye tracking data (if concepts or symbols provide the units of analysis), cards sorted in card sorting tasks, or click-streams recorded by subjects involved in human computer interaction. The analysis of sequential symbolic data is done by referring to possible and competing interpretations as specified by several theories or relational structures (cf. 3.2 and 4.1). [11]

3.2 KT is a theory-based method

It might strike readers as somewhat puzzling—but KT needs theories to assess theories! Theories in KT are simply relational structures, which in turn are made up of concepts and one or many relations that connect these concepts in a meaningful way. The theories are usually descriptive in nature. For instance, if I want to set up a theory on the generation and transfer of knowledge in an organization, I might come up with a set of concepts like "experience", "practice", "course" and relations like "x_generates_y", "x_is_a_requirement_to_y". I might combine both concepts and relations in a meaningful way to spell out one or several (competing) theories about the domain in focus. [12]

Even if I have collected a number of relations that describe the field under study, data processing via KT is much easier, once theories (relational structures) are set up that rely on just one relation. In this way, a number of theories or relational structures is obtained. The overall objective of knowledge tracking is simply to find out which of many possible theories gives the best account of the sequence of symbolic data. This is nothing but a more technical description for the search for a cognitive representation that provides the best explanation of the data. [13]

3.3 KT can be employed in confirmative or in an inductive way

So far, I have described a confirmative usage of theories. This proceeds by setting up some competing theories in advance and finding out how well each of them is supported by the data. It is a confirmative or theory-driven approach since the basic question is how well given or pre-specified theories are confirmed by the data. Quite easily, however, the methodological relationship of data and theory may be turned upside-down: One may then ask which theories or relational structures can be induced on the basis of the data given. In this contribution, I wish to concentrate on the confirmative usage of knowledge tracking. A presentation of both the confirmative and inductive usage is given in JANETZKO (1996). [14]

3.4 KT is amenable to both qualitative and quantitative analyses

Note that the confirmative usage of theories does not by itself make KT either a qualitative or a quantitative method: (i) Clearly, the degree to which a theory is supported by data can be expressed quite well by a quantitative measure, viz., a number. (ii) Alternatively, one may also ask which parts of a theory under study are supported by the data and which are not. The result of the analysis is not a number but a relational structure, viz., a theory. Note that both ways of using KT are confirmative since each of them starts with a theory and finds out how well it is confirmed by the data. While both ways of bringing theory and data together in a meaningful way are theory driven, the first one leads us on a quantitative track, while the second one is qualitative in nature. [15]

4. Data and Theories in Knowledge Tracking

Knowledge Tracking uses data to decide how well one theory explains the data. If there are several competing theories, data decide which theories give the best account of them. I will now provide a more detailed explanation of the role that data and theories play in knowledge tracking. [16]

4.1 The side of the theory

Theories are required to run an analysis via KT. In fact, KT evaluates one or many theories on the basis of sequential symbolic data (see next section). This means that KT brings together theory and data in a meaningful way so that the data determine whether or not the theory under study is worth considering as the best explanation of the data. Note that KT does not claim to identify a "true" theory. KT only specifies two issues: First, it calculates the degree to which a theory explains or—which is the same thing—fits the data. In this way, KT supports selection of a theory among a number of competing theories. This aspect points to the quantitative usage of KT simply because the quantification of the degree to which theories fit the sequential data is done by using a numerical measure. Second, when starting with a theory set up to explain the data, KT finds out the most parsimonious subtheory that was really needed to explain the data. In this way, given a larger theory, KT supports a data-driven reduction of this initial theory. This second aspect shows the qualitative usage of KT, since in this case the result is not a numerical measure but a qualitative entity, i.e., a "slim" theory needed to explain the data. Theories in knowledge tracking are spelled out by using a simple but powerful formalism called relational structures. This formalism makes use of only two ingredients: elements (concepts) and relations:

*A relational structure*¹ is a set together with one or more relations on that structure (KRANTZ, LUCE, SUPPES & TVERSKY 1971, p.8). [17]

1 Following TARSKI's (1954) terminology, SUPPES and ZINNES (1963, p.5) use the concept of a *relational system* to designate what has later been termed *relational structure*. Especially in the German literature on measurement a relational structure is often dubbed a *Relativ* (e.g., ORTH 1974).

By selecting these two ingredients and putting them together to build theories, the researcher sets the scope of the theory KT is evaluating. [18]

4.1.1 Elements of a Theory

The elements (concepts) in a theory used by knowledge tracking are the basic entities under study. Theory building in a domain like the "Internet" will have to make use of concepts like, e.g., "traffic", "web_site", "home_page", "HTML", "content", "graphics", "e_commerce", "money", "credit_card", "career", "data", "future", "vision", "learning", "education", "art", "experiment", "communication", "culture", or "identity". [19]

When building a theory within the framework of KT, two things should be taken into consideration about the elements (concepts) of a theory. First, the theory should cover at least partially concepts or elements that occur in the data to be analyzed. Otherwise, KT has no possibility of inter-relating data and theory. In fact, data analyzed via knowledge tracking forms simply a sequence of elements of the same kind as those integrated into a theory. Second, by choosing a particular level of description, the level of granularity of the theory is specified. For instance: A theory on the accumulation of knowledge in organizations will most probably cover a (perhaps partially) different set of elements than a theory on learning in humans. [20]

Addressing the first issue mentioned in a proper way is clearly essential. Its fulfillment is required, otherwise KT simply can not be applied. The second issue relates to the domain to which KT is applied. In KT, there is no restriction concerning the number of elements in a theory. [21]

4.1.2 Relations of a Theory

Relations are the "glue" needed to put together elements (concepts) so that complex propositions can be expressed. In other words: Elements (concepts) and relations are the basic building blocks to put together or to spell out a theory in a bottom up way. To give an account of a domain like the "Internet" relations like, e.g., "x_is_a y", "x_is_physically_connected_to y", "x_is_a_requirement_to y", "x_is_needed_to_achieve y" have a good chance to be chosen. In contrast to the first issue mentioned with respect to elements, there is no requirement concerning an overlap between data and theory. The only overlap needed is with regard to elements, but not with respect to relations. In fact, the relations that most plausibly exist in the data are inferred by KT! [22]

The second issue mentioned in the preceding section has an equivalent in this section. This means, by choosing particular relations the theory with which knowledge tracking deals is selected and specified. For instance, in a technical theory of the Internet the relation "x_is_physically_connected_to y" may play a crucial role, while in a theory of the educational usage of the Internet the relation "x_is_communicated_to y" may be of greater importance. In KT, there is no restriction concerning the number of relations in a theory. In other words: the

researcher is free to integrate one or many relations in a theory. Analyses conducted on the basis of relational structures that use only one relation are easier to interpret, however. [23]

4.2 The side of the data

Having introduced the notion of theory in knowledge tracking, it is now quite easy to describe the type of data that KT uses to evaluate the theory. In short, the elements in the data are of the same kind as the elements in the theory. Ideally, the elements in the data form a superset of the elements in the theory. Clearly, this is not an exotic requirement, since every theory in a domain "X" should somehow be related to the data that are analyzed (or vice versa). If this requirement is fulfilled, then any type of symbolic sequential data can be analyzed via knowledge tracking:

- thinking aloud protocols (provided they are reduced to central concepts)
- sequences of symbols forming texts (provided they are reduced to central concepts)
- eye-movement data (provided eye movements fix symbolic entities)
- click-streams collected in Internet usage
- mouse-pointing or key-stroke data in human computer interaction
- cards selected in a card sorting task [24]

5. Knowledge Tracking—A Condensed 5-Step Scheme

Analyzing cognitive structures via knowledge tracking can be carried out by working through 5 steps of collecting and analyzing data. Note that step 4 and 5 represent the quantitative side of KT since these steps of analysis are required to express numerically how well a sequence of symbolic data can be explained by a theory (viz., a relational structure). Step 6 represents the qualitative side of KT since the output of the step is again a relational structure. Next, I will briefly list all five steps. This will convey a bird's eye view on KT. Later, I will give a more detailed account of each of the five steps underlying the usage of knowledge tracking. [25]

We may briefly list the steps as follows:

- *Step 1:* Eliciting concepts and relations in the domain under investigation and setting up theories, or using a more technical terminology, relational structures, about this domain.
- *Step 2:* Recording data (sequences of concepts).
- *Step 3:* Translating the relational structures from a graphical language into a probabilistic language.
- *Step 4:* Taking the relational structure to predict the data in the sequence of concepts. A numerical (and thus quantitative) score is used that expresses

how well a relational structure explains the data. This score is called a goodness of fit score or simply "gamma".

- *Step 5:* On the basis of the goodness of fit scores of various competing relational structures, the structure that produced the highest goodness of fit score is taken to be the best explanation of the data under study
- *Step 6:* The bridging inferences that have been carried out to calculate the goodness of fit scores are now collected and used to synthesize a new relational structure. This is the qualitative side of KT. [26]

In what follows, I will present these five steps in a more detailed way. [27]

5.1 Step 1: Setting up relational structures

To run an analysis of KT, the researcher has to set up a detailed description of the domain and the phenomena under study. This description needs to be given in terms of elements (concepts) and relations, both of which have been introduced in the preceding section. The outcome of this description is called a relational structure. "Relational structure" is a key notion in knowledge tracking since it can be viewed both as a qualitative or a quantitative hypothetical construct. Relational structures can be very well presented graphically as network models—this is the qualitative view of relational structures. If for instance, you want to analyze the role of the WWW as seen by different groups like, e.g., entrepreneurs, artists, or educators, set up a relational structure by first collecting concepts and relations that might be of importance to this question. Usually, the researcher will set up several relational structures each of which uses one relation, e.g., "x_is technically_supported_by_y". The researcher may then build a relational structure like

```
(x_is technically_supported_by
  (Web_Site Web_Server)
  (Web_Server Operating_System)
  (Web_Server Computer)
  (Operating_System Computer)) [28]
```

Clearly, more relational structures are needed, e.g., the relational structure built by the relation "x_is_a y". This leads to a relational structure like this one:

```
(x is_a y
  (Web_Site Software)
  (Web_Server Software)
  (Browser Software)
  (Computer Hardware)) [29]
```

Suppose, the researcher wants to test the hypothesis that a particular group speaks and thinks of the domain under study in terms of flow of (human)

communication rather than using other epistemological grids, e.g., in terms of streams of bits and bytes. To test this hypothesis, the researcher has to express all competing hypotheses in terms of relational structures. For instance:

(x communicates_to y
(teacher pupil)
(pupil class)
(pupil pupil) ...). [30]

Note that KT can only process two-place relations like the one used so far. If the researcher wishes to use a three-place relation (x communicates via y to z), she or he has to recode this relation in terms of a two-place relation:

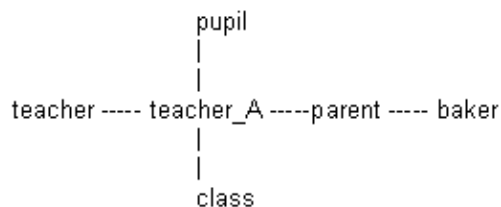
(x communicates_to_y_via z
(teacher email)
(pupil email)
(teacher chat)
(pupil chat) ...) . [31]

5.2 Step 2: Recording sequences of symbolic data

Sequences of symbolic data can be acquired in a variety of ways: Thinking aloud protocols, texts, eye-tracking data etc. The formal requirements to be met have already been spelled out in one of the preceding sections (cf. Section 4.2). [32]

5.3 Step 3: Translating the relational structures into a probabilistic language

Relational structures can be expressed as graphical networks. Take for instance the simple relational structure that describes the major communication partners of a particular teacher "A" in a school setting. Let us assume that the teacher communicates mainly with other teachers, parents, pupils, and classes. We have also introduced a person who has no established role in an educational setting (baker) to indicate that communication is not confined to this surrounding. [33]



The qualitative model sketched above can be easily transformed into a probabilistic description. Then, if the researcher focuses on the concept "teacher_A", she or he has to make sure that weights (probabilities) are attached to the links leading to the neighbor concepts all of which sum up to one.

Accordingly, in the example sketched above the researcher obtains a probability distribution like

```
teacher_A -- .25 .. parent
teacher_A -- .25 .. class
teacher_A -- .25 .. teacher
teacher_A -- .25 .. pupil [34]
```

The (qualitative) network or relational structure I have just re-described in terms of (quantitative) transition probabilities is clearly a very simple one. The maximum path length in this network is 1. As a matter of fact, this approach needs to be generalized so that large networks (usually with path lengths greater than one) can also be recast in probabilistic terms. This can be easily achieved by the knowledge tracking engine (KTE), which is a computer tool that is accessible via the World Wide Web: <http://www.knowledge-tracking.com> [Broken link, FQS, August 2005] (see section 7). [35]

In cases where there are path lengths that are greater than one (i.e., in all relational structures that attempt to capture realistic phenomena or theories) the researcher has to specify the way the probabilities decreases with larger path lengths. I call this the decay rate. Usually, an exponential decay rate is assumed (JANETZKO 1996). [36]

5.4 Step 4: Taking the relational structure to predict the data in the sequence of concepts

This is the step that brings together theories (relational structures) and data. Let us assume that the researcher uses the relational structure introduced in the preceding section to explain raw data, viz. a sequence of concepts. For instance, the researcher has collected thinking aloud protocols as raw data that are reduced to a sequence of concepts. If the sequence is made up of concepts, like

```
teacher_A—teacher—class_room—teacher—pupil—class,
```

there are 5 transitions of concepts each of which is actually compared to the relational structure under study. This will produce the following results:

1. transition teacher_A-> teacher .25
2. transition teacher -> class-room 0
3. transition class- -> teacher .25
room
4. transition teacher -> pupil .25
5. transition pupil -> class .25 [37]

The researcher may simply calculate the average of all transition probabilities ($.25 + .25 + .25 + .25 / 5 = .2$), which is no longer a probability score but a

goodness of fit score, viz., a score that shows how well the theory (relational structure) which is being tested can explain the data. A poor theory will yield low scores, while good theories will produce high scores. Again the introduction given is a simplified version of the actual usage of this method. Usually, the researcher will apply different theories that are "competing" in order to determine whether they can explain the data. As a consequence, there are sometimes various alternative ways to explain a particular transition in the data that differ with respect to the path lengths needed to explain the transition. There is a mechanism built into the computer program that supports KT, which makes sure that the shortest path through a structure under study is selected². In other words: if a theory can offer various alternative explanations of a phenomenon in the data, the most simple explanation is chosen. [38]

5.5 Step 5: Selecting the best fitting relational structure

Usually, the researcher sets up several relational structures to analyze one sequence of relational data. This actually means that several competing theories are tested on the basis of one set of data. The researcher then gets a goodness of fit score for each relational structure and takes the relational structure that produces the best score as the theory that (in comparison to other competing theories) provides the best explanation of the data. [39]

5.6 Step 6: Synthesizing a new relational structure

Hitherto, I have only described the quantitative side of KT, where a score is calculated for each of one or many theories (relational structures) that expresses how well this theory explains a sequence of symbolic data (cf. step 4 and step 5). [40]

Now, let us take a close look at the way this score is calculated. We focus on a transition of two concepts in the data. We then turn to the theory (relational structure) and ask which specific paths have to be traversed in a structure so that a path between the two concepts in a transition is found. In the simplest case, we have a transition of two concepts like "teacher—classroom" and find a path in the relational structures that is identical to the transition upon which our work is focused. Sometimes, however, we have to traverse more than one path in a relational structure to find a connection between the two concepts in a transition under study. Let us take a look at two concepts, say "formal_occupational_socialization" and "informal_occupational_socialization" and a relational structure:

2 The repeated search for the shortest path in a relational structure together with the collection of all paths traversed (or the transition probabilities linked to each of the paths traversed) is in fact a standard method in machine learning called construction of a decision tree (e.g., QUINLAN 1986).

(is a
(informal_occupational_socialization
occupational_socialization)
(formal_occupational_socialization
occupational_socialization)) [41]

To find a path connecting "formal_occupational_socialization" and "informal_occupational_socialization" in the relational structure sketched above, we have to traverse not only one, but two paths:

informal_occupational_socialization—
occupational_socialization

and

occupational_socialization—
formal_occupational_socialization [42]

The path traversed to find a connection between two concepts in a transition is called the bridging inference. Now, to see the whole picture, we only have to generalize our observations slightly. Usually, when analyzing a trace we do not just have one single transition, but many of them. For each transition, we obtain a bridging inference. We may then collect and synthesize all bridging inferences and form a new relational structure! A relational structure like this represents a subset of the relational structure that has been applied initially to explain the data. This new structure is nothing but a newly generated theory on the data. In section 4.1, I have referred to this theory as the most parsimonious sub-theory of a prespecified theory needed to explain the data given. [43]

6. Related Work

A number of approaches both in psychology (e.g., BAKEMAN & GOTTMAN 1997; COOKE, NEVILLE & ROWE 1996) or sociology (e.g., ABELL 1988) have been inspired by the idea that actions or thoughts (that are evidenced by actions) follow patterns, which can be fruitfully analyzed by network models of cognition. The type of model resulting from these approaches is rooted in the work of QUILLIAN (1968), COLLINS and QUILLIAN (1969) and COLLINS and LOFTUS (1975). Their investigations provide evidence that concept networks are in fact good models for representing knowledge since they seem to reflect the organizing principles of the semantic memory. [44]

Many different examples of network models of cognition have been derived from the early work on network representations mentioned above. Network models of cognition are used in fields and disciplines as diverse as Artificial Intelligence, cognitive science, psychology, sociology, and pedagogy. The broad acceptance of network representations indicates that their usage is not confined to models of

cognition, which is the focus of this contribution. In what follows, I use the notion of network representations of cognition to refer to models of cognitive processes of both individuals or groups that employ symbolic or conceptual nodes and links between them as their basic building blocs. The semantics of models of network representations are usually derived from their nodes and links. Both of them carry a meaning that can be expressed by concepts, which in turn may be taken from ordinary language or from a particular theory. Some models, however, do not specify the links semantically (e.g., Pathfinder: SCHVANEFELD 1990). Almost by definition, network models of cognition are examples of qualitative models since their major building blocks are concepts.³ However, when procedural elements become part of the model or the model itself is subject to calculations (e.g., calculating the similarity between models or nodes, i.e., concepts, within a model) they also become quantitative in nature. [45]

Models of network representations of cognition can be compared with regard to quite different aspects, including the function primarily addressed, the kind of data that feed into the method, the algorithm used (if an algorithm is used at all), or the kind of the network representation generated. In what follows, I will concentrate on the function the network representation is meant to fulfill. Along with the classification grid based upon functions presented below I will mention a few network models of cognition. The collection of models discussed is far from exhaustive. The models considered here have been selected because each of them highlights a particular aspect typical of the class in this grid. [46]

There are a number of reasons why network representations of cognition have a strong appeal to many researchers in sociology and psychology. Firstly, network representations of cognition are often used as a kind of *communication tool* that conveys the meaning of processes, relationships, activities, or theories in an intuitive way. Examples are mapping techniques, which are often used as methods to structure discussions in groups or concept maps that are employed in pedagogical settings (e.g., NOVAK 1998). [47]

Secondly, artificial neural networks are examples of network representations of cognition that promise to be "*close*" to the *neural basis of cognition*. In general, however, many neurobiological details known to influence processing of real neurons (e.g., activities of neurotransmitters) are left out. In artificial neural networks, there is usually a distinction between distributed and local ("localist") representations. The former is meant to represent information (e.g., of a concept) across a number of nodes, while the latter represents information about processes or a concept "locally" on a single node. Viewed as a model of neural processing this representation is often called "grandmother cell representation" since it suggests that we have dedicated neurons for symbols like grandmother. [48]

Thirdly, network representations of cognition provide a *description of cognitive processes*. Clearly, description is always an important reason for setting up and using network representations (e.g., concept maps). However, there are great

³ Note however that some network models (e.g., Bayesian networks, see below) can easily be extended such that numerical variables may become nodes in the network representation.

differences as to whether the approach provides mechanisms supporting descriptive accounts of cognitive processes in terms of networks. A typical example of a method that provides this kind of support is the structural modeling technique called Pathfinder (SCHVANEFELD 1990). The Pathfinder algorithm allows one to build network models on the basis of proximity data (e.g., similarity judgments or relatedness ratings between two entities). Comparative studies of different groups (e.g., experts and novices) can be carried out by using Pathfinder. Based upon Pathfinder is the PRONET method (COOKE, NEVILLE & ROWE 1996). This method is suited to analyze sequential and thus behavioral data, which is the type of data KT analyzes, too. The output of both Pathfinder and PRONET is clearly a network representation. However, none of these methods provides a semantic specification of the links in the generated networks as KT does. [49]

Fourthly, some network representations of cognition are *procedural models*, which basically means that they can be employed for calculations such that a number of activities in model formation and testing (e.g., selection and shrinking of models, calculating the degree to which a theory is supported by empirical data) may be conducted by computer-driven data processing. The place of KT within the classification grid is in this class, since KT is a procedural model, but the procedural elements are not taken to generate some aspects of cognition. [50]

Fifthly, network representations of cognition are often models that not only describe but also *simulate (and thus generate) some aspects of cognition*. The reason to distinguish between the fourth and the fifth group of functions is that some network representations are procedural so that calculations can be carried out (group four), while others use this feature to simulate or generate some aspects of cognition by a computer model (group five). A case in point is Bayesian networks, which are also called Bayesian belief networks (BBN), or causal nets (PEARL 1988; 2000). BBNs are directed acyclic graphs (DAG), the nodes of which represent probability distributions of variables that can be used both for descriptive model construction and inference. An example of using BBNs is the work of REHDER (1999) who uses this type of network representation of cognition in the field of processing categories. In particular, he takes BBNs to simulate classification of examples into categories. [51]

On the basis of the issues discussed so far KT may be qualified as a procedural model, which also provides descriptive account of cognition. Similarly to PRONET (COOKE, NEVILLE & ROWE 1996) it also uses sequential data. This type of data is discussed at length in RITTER and LARKIN (1994), SANDERSON and FISHER (1994), and SANDERSON et al. (1994). [52]

7. The Knowledge Tracking Engine: Using Knowledge Tracking via the WWW

The web-site <http://www.knowledge-tracking.com> provides a number of materials on the method described in this contribution. Apart from a collection of literature and a tutorial on knowledge tracking this web-site allows visitors to carry out remote analyses of data on the basis of the knowledge tracking engine (KTE): <http://cogweb.iig.uni-freiburg.de/KT/kt.run/index.html>. This section concentrates on the usage of the knowledge tracking engine as a tool for analyzing sequential symbolic data. A bird's eye view on the usage of the interface when conducting an analysis KTE is presented in Fig. 1. A more detailed description of the web-site and its usage is given below. [53]

Each user who wants to carry out analyses of sequential data via <http://www.knowledge-tracking.com> [Broken link, FQS, August 2005] is requested to log on to the system. When visiting <http://www.knowledge-tracking.com> [Broken link, FQS, August 2005] for the first time, the user has to choose a password. This allows him or her to enter the system at a later point in time and continue working on the data (traces of symbolic data) and theories (structures) or conducting analyses. Having successfully logged into the system, the user will enter the control panel. Here the user will find the main menu where she or he may choose between 3 groups of commands, which are summarized by the concepts of "trace", "structure", and "track". Apart from the 3 major groups of commands in the main menu of KTE there is also a device that lists frequently asked questions (FAQ) about knowledge tracking. While working with the system there is an indication of the user (more precisely: the password), the particular trace selected and the particular structure selected. Next is an introduction of each of the three major groups of commands of the main menu.

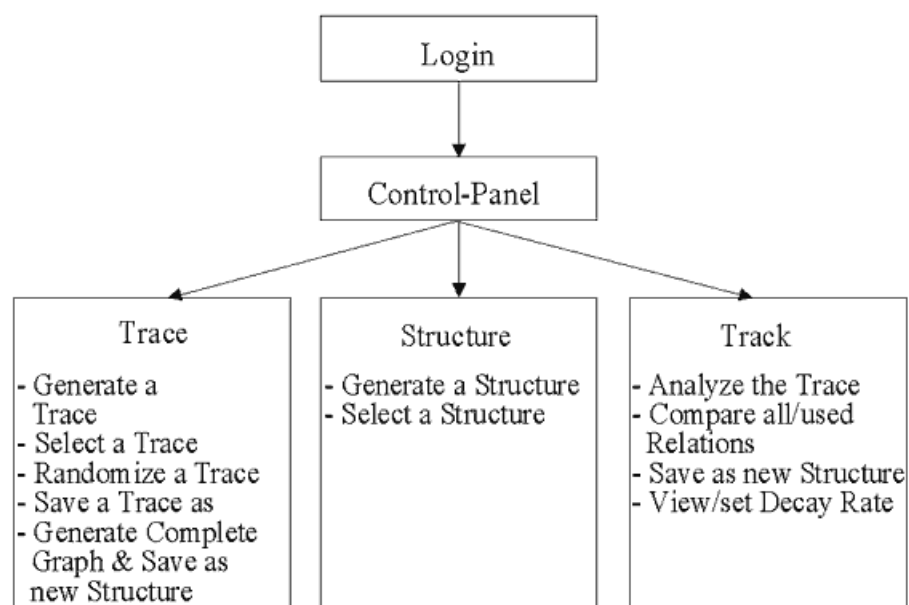


Fig. 1: A Bird's Eye on the Knowledge Trace Engine (KTE) [54]

7.1 Trace

This group of commands summarizes all operations on sequences of symbolic data, i.e., traces, that enter knowledge-tracking as empirical data (cf. 4.2. The side of the data). The user may for instance enter a trace just by copying the data into a window of the interface and save it under a name chosen by the user (cf. Fig. 2). All traces saved are listed, and each of them can be freely chosen to run an analysis. Each trace may be *randomized*.

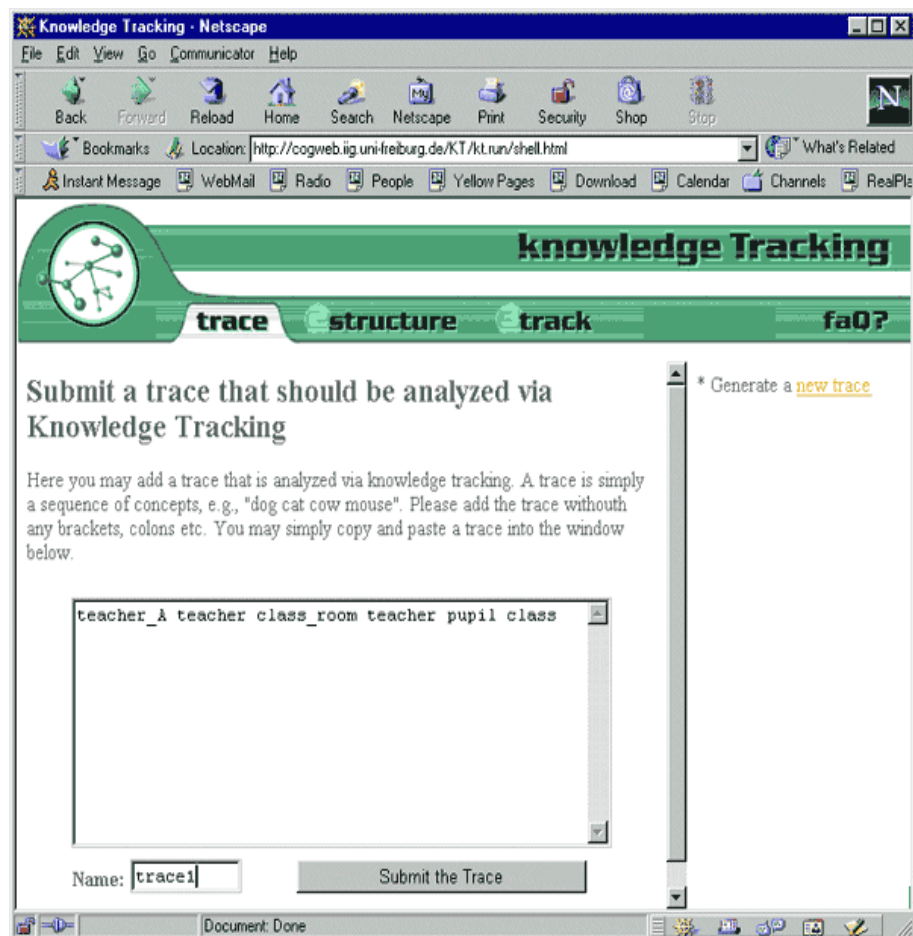


Fig. 2: Entering Sequential Data into KTE [55]

Moreover, each trace can be taken to generate the *complete graph* over the concepts of a trace, i.e., the *random structure*. The former (*randomized trace*) represents the principle of chance on the side of the data. The latter (*random structure*) represents the principle of chance on the side of the theory. Both are needed as reference scores of knowledge tracking:

- Applying the random trace to all relational structures under study is a method that indicates whether or not there are biases. This method is an example of a *Monte Carlo Study*. A random trace leading to high goodness of fit scores (high gamma-scores) when applied to some relational structures while other

relational structures produced low scores, would be a clear sign of a distortion or a bias.

- Applying the random structure to all traces under study will lead to the result that is to be expected, if we take the principle of randomness to explain the data. Clearly, any true theory-based relational structure can only then claim to give an account of the data if it produces better results than the random structure. Non-parametric tests for matched pairs may be employed to find out whether or not Gamma-scores differ in a significant way (e.g., Wilcoxon matched-pairs signed rank test). [56]

In the program interface of knowledge tracking, saving of the random structure is done via the menu option "Trace" and not "Structure". This is simply due to the fact that this structure is based on a trace. In the resulting random structure each concept is linked to each other concept (complete graph). [57]

7.2 Structure

"Structure" is the name of the second group of commands the user finds in the main menu of KTE. It provides an editor to the user that is tailored to enter a relational structure that may be saved using a unique name. When entering a trace (sequence of concepts) no particular format is required. However, when entering a relational structure it is necessary to stick to an appropriate format. For this reason, an editor is provided where the user has to specify the name of the relational structure as a whole and also the name of the particular relation(s) of a structure (cf. Fig. 3). If the relational structure covers only one relation, both names may be the same.

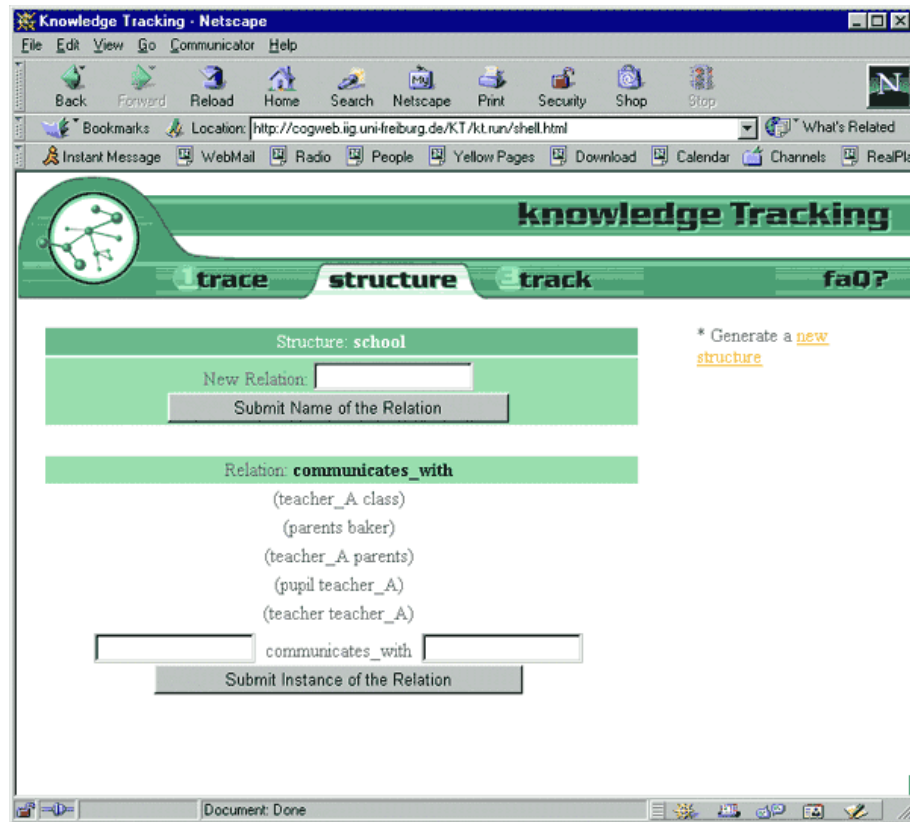


Fig. 3: Entering Relational Structures into KTE [58]

By entering the names and the instances of one relation the relational structure is completely specified and saved. Note that each relational structure can be enlarged at any point in time. In the current implementation of knowledge tracking, modification of a relational structure (i.e., deletion of instances or of a relation) is not possible, however. [59]

7.3 Track

The name of the third group of commands provides a clear indication of the commands to be found here. Since knowledge tracking rests on theories (relational structures) and data (traces), examples of both of them must be entered before the user may analyze a trace. Note that this step provides a quantitative analysis of a trace and becomes evident by two outputs. This first output is a figure that shows graphically how well the relational structure which has been applied can explain transitions of concepts in the trace under study. The second output uses numbers to show the same (cf. Fig. 4). Usually, several relational structures will be taken to conduct an analysis of a set of sequential symbolic data. The structure that leads to the best gamma score will be taken to give the best account of the data under study. This is basically a *selection of a relational structure* among a set of candidate structures.

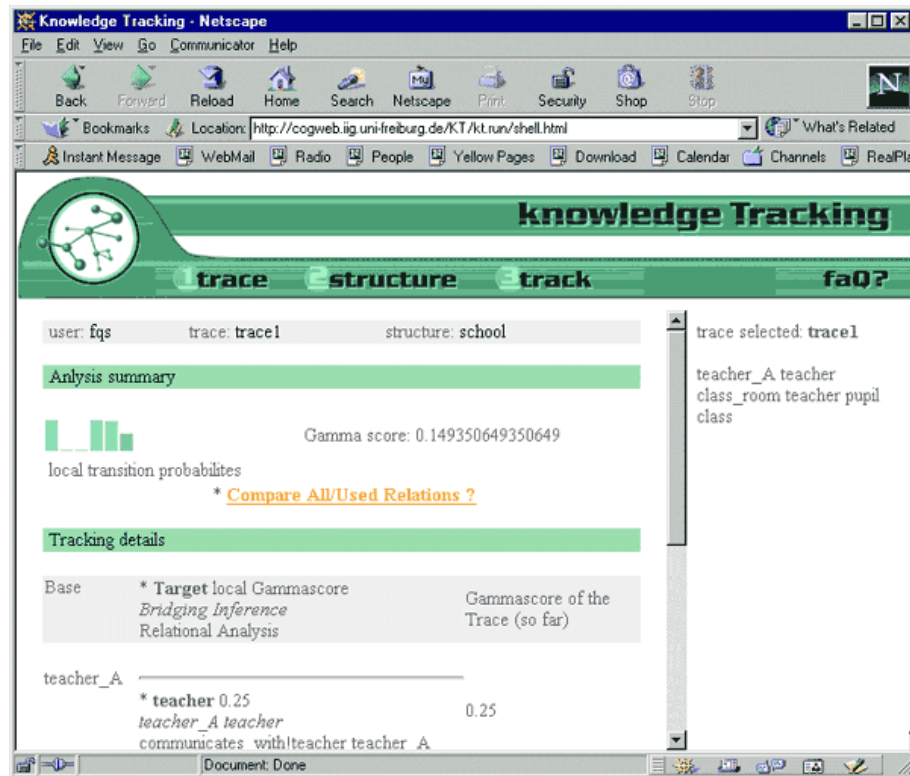


Fig. 4: Graphical output (top) and numerical output of KTE (bottom) [60]

Having analyzed a trace via the procedure described so far, the user may make use of the command "compare all/used relations" (cf. Fig.5). Note that choosing this command is equivalent to using knowledge tracking in a qualitative way, since this step will provide the bridging inferences actually deployed by the system to conduct an analysis of the trace. The output is a subset of the structure initially applied to analyze the trace. This is basically a *shrinking of a relational structure*. Since this subset is again a relational structure, it may be saved under a new name. Thus, it can be used for subsequent analyses.



Fig. 5: Comparison of the prespecified theory (normal font) with the most parsimonious subtheory need to explain the data (bold font) [61]

In addition to the commands discussed so far, the option "track" of the main menu provides the user with the possibility of fixing the decay rate in the relational structure under study. The decay rate is a parameter of the representation of each relational structure. In knowledge tracking, the transition probabilities between two concepts of a relational structure become lower the greater the path lengths between them (spreading activation). The decay rate specifies precisely the degree of this diminishment. [62]

8. Discussion

Bringing together theories and data in a meaningful way can take many forms. Quantitative and qualitative methods are two major classes of the way this endeavor is accomplished. Usually, methods fall either into the one or the other class. KT is an exception from this rule since it can be used in a qualitative or quantitative mode. Thus, under the roof of one method, qualitative and quantitative methodological approaches can be fruitfully used in conjunction since each type of method brings to the fore a particular array of phenomena. Used in a quantitative mode, KT conducts a data-driven selection between many competing theories—used in a qualitative mode, KT carries out a data-driven reduction of one theory. This is clearly a very high-level view of relating theory of data in a meaningful way. Taken down to earth this means that the researcher can determine which is the dominant mindset in a group or individual. In addition, she

or he may compare two groups or individuals concerning their representations or mind-sets by specifying possible theories in advance and letting KT do the task of a data-driven reduction within a set of theories or within one theory. [63]

Knowledge Tracking provides a coherent framework for processing symbolic data. Within this framework data, processing may take up possibly raw data and (qualitative) relational structures) and proceed to more condensed entities in data processing like goodness of fit scores. In fact, one of the most attractive features of knowledge tracking is that very often it can be directly applied to many kinds of raw data, e.g., those collected in human-computer interaction. However, a lot of preparation efforts have to be put into both the collecting method (since it has to take up only sequences of symbols) and also into the process of setting up theories (relational structures). [64]

To properly apply KT on a set of sequential symbolic data, in both basic and applied research, it is necessary to recast the theory in terms of relational structures. This formalism is a very simple, but very powerful, mechanism that allows us to re-describe a wide array of phenomena in sociology and psychology. The very act of viewing something like, say, communication streams in an organization or the cognitive representations that various groups might entertain of a particular topic in terms of a formalism may seem farfetched and difficult to accomplish. But once this initial formalization is mastered, the initial investment pays off in a powerful method that permits us to view phenomena both in a qualitative and quantitative mode. [65]

References

- Abell, Peter (1988). The 'structuration' of action. In [Nigel G. Fielding](#) (Ed.), *Action and structure. Research methods and social theory*. London: Sage Publications.
- Roger Bakeman & John M. Gottman (1997). *Observing interaction: An introduction to sequential analysis*. (2nd edition). New York: Cambridge University Press.
- Bauer, Malcolm I. & Reiser, Bryan (1990). Incremental envisioning: The flexible use of multiple representations in complex problem solving. In *Proceedings of the 12 th Annual Conference of the Cognitive Science Society* (pp.317-324).
- Campbell, Donald T. & Fiske, Donald W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Collins, Allan M. & Loftus, Elizabeth F. (1975). A spreading activation theory of semantic processing. *Psychological Review*, 82, 407-428.
- Collins, Allan M. & Quillian, M. Ross (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8, 240-247.
- Cooke, Nancy J.; Neville, Kelly J. & Rowe, Anna L. (1996). Procedural network representations of sequential data. *Human-Computer Interaction*, 11, 29-68.
- [Denzin, Norman](#) (1978). *The research act: A theoretical introduction to sociological methods*. New York: McGraw-Hill.
- Hewson, Peter W.; Beeth, Michael E. & Thorley, Nancy R. (1998). Teaching for conceptual change. In Kenneth G. Tobin & Barry J. Fraser (Eds.), *International handbook of science education* (pp.199-218). Dordrecht, Netherlands: Kluwer Academic Publishers.
- Indurkha, Bipin (1992). *Cognition and metaphor*. Dordrecht: Netherlands: Kluwer.
- Janetzko, Dietmar (1996). Knowledge tracking—A method to analyze cognitive structures. *IIG-Berichte*, 2.

- Krantz, David H.; Luce, R. Duncan; Suppes, Patrick & Tversky, Amos (1971). *Foundations of measurement* (Vol I). San Diego: Academic Press.
- Kuhn, Thomas (1970) *The structure of scientific revolutions* (2nd edition). Chicago: The University of Chicago Press.
- Novak, Joseph D. (1998). *Learning, creating, and using knowledge: Concept maps as facilitative tools in schools and corporations*. Mahwah, NJ: Erlbaum.
- Orth, Bernhard (1974). *Einführung in die Theorie des Messens*. Stuttgart: Kohlhammer.
- Pearl, Judea (1988). *Probabilistic reasoning in intelligent systems. Networks of plausible inference*. San Mateo, CA: Morgan Kaufmann.
- Pearl, Judea (2000). *Causality*. Cambridge, UK: Cambridge University Press.
- Quillian, M. Ross (1968). Semantic memory. In Marvin Minsky (Ed.), *Semantic information processing* (pp.227-270). Cambridge, MA: MIT Press.
- Quinlan, John Ross (1986). Induction of decision trees. *Machine Learning*, 1(1), 81-106.
- Rehder, Bob A.(1999). Causal-model theory of categorization. In Martin Hahn & Scott C. Stones (Eds.), *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society* (pp.595-600). Mahwah, NJ: Erlbaum.
- Ritter, Frank E. & Larkin, Jill E. (1994). Developing process models as summaries of HCI action sequences. *Human-Computer Interaction*, 9, 345-383.
- Sanderson, Pamela M. & Fisher, Carolanne (1994). Exploratory sequential data analysis: foundations. *Human-Computer Interaction*, 9, 251-317. [100-68].
- Sanderson, Pamela M.; Scott, Jay; Johnston, Tom; Mainzer, John; Watanabe, Larry & James, Jeff (1994). MacShapa and the enterprise of exploratory sequential data analysis (ESDA). *International Journal of Human-Computer Studies*, 41, 633-681.
- Schvaneveld, Roger W. (1990). *Pathfinder associative networks: Studies in knowledge organization*. Norwood, NJ: Ablex.
- Suppes, Patrick & Zinnes, Joseph L. (1963). Basic measurement theory. In R. Duncan Luce; Robert R. Bush, & Eugene Galanter (Eds.), *Handbook of mathematical psychology* (Vol 1, pp.1-76). New York: Wiley.
- Tarski, Alfred (1954) Contributions to the theory of models I, *Indagationes Mathematicae*, 16, 572-581.
- Vosniadou, Stella & Brewer, William F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, 24(4), 535-585.

Author

Dietmar JANETZKO, Dr. phil., Dipl. Psych., Study of Psychology, Philosophy and Theology in Bochum/Germany. Works at the Department of Cognitive Science at the Institute of Computer Science and Social Research of the University of Freiburg/Germany. His main fields of research are psychology of concepts, online research, research methods, and human computer interaction. Major publications: *Statistische Anwendungen im Internet*. München: Addison-Wesley.

Contact:

Dietmar Janetzko

Phone: +49 (0) 761 203 4948

E-mail: dietmar@cognition.iig.uni-freiburg.de

Citation

Janetzko, Dietmar (2001). Processing Raw Data both the Qualitative and Quantitative Way [65 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 2(1), Art. 11, <http://nbn-resolving.de/urn:nbn:de:0114-fqs0101111>.

Revised 7/2008